UNIVERSITEIT GENT
2005
FACULTEIT ECONOMIE EN BEDRIJFSKUNDE
VAKGROEP MARKETING

Using Predictive Modeling for Targeted Marketing in a Non-Contractual Retail Setting

Wouter Buckinx

2005

Dissertation submitted to the Faculty of Economics and Business Administration, Ghent University, in fulfillment of the requirements for the degree of Doctor in Applied Economic Sciences

Promotor: Prof. dr. Dirk Van den Poel

Doctoral jury:

Prof. Dr. Dirk Van den Poel
(Ghent University)

Prof. Dr. Patrick Van Kenhove
(Ghent University)

Prof. Dr. Douglas MacLachlan
(University of Washington, Seattle)

Prof. Dr. Tammo Bijmolt
(University of Groningen)

Decaan Prof. Dr. Roland Paemeleire
(Ghent University)

Prof. Dr. Eddy Omey
(Ghent University)

ACKNOWLEDGEMENTS

This is the end. Or is it more like another start? For many relatives and friends, the finishing of my doctoral thesis will definitely be considered as the end of my student days. It's hard to remember how many times people made the well-meant comment during these last four years, that I was 'still' studying at the University. These days have passed and a new age is near. Anyway, it was a great time during which I learned a lot, and it was not only about economics, marketing or modeling. To put it briefly: if I would have the chance, I would do it the same way.

Without the support of different hands, this work would have not been realized. Therefore, I want to take the opportunity to show my gratitude to all those people who helped me whenever, wherever.

First, I am grateful to my advisor Dirk Van den Poel. Thanks to him I was introduced to the 'modeling world' during the exciting years of the postgraduate course Master of Marketing Analysis and Planning. Only a word was needed to engage in this doctoral research process. I also greatly appreciated his lobbying efforts at the collaborating companies, which assured the continuation of my dissertation. Finally, I am grateful for all helpful contributions in this work and his passed dedication to translate these results in submissions to precious international journals.

Furthermore, I want to thank Makro Belgium for their financial support during the last three years of my stay at Ghent University. I especially express my thanks to Philippe Nowé, Chris De Weerdt, Kathleen Geys, Els Gorrebeeck, and Rik Deruytter of their Marketing Department, for the interesting discussions and the many opportunities we had to test our theory in real-life. Finally, I may not forget Kaat Geerts and Ludo Peeters for their technical support.

I am grateful to all my colleagues at the Department of Marketing for the pleasant and inspiring working environment, the welcome coffee breaks at moments of research blues, and our gastronomic expeditions in Ghent. Geert, it was a great time at university, both during our joint studies as during our ‘after-school’ activities. I hope that the end of this PhD process won’t mean that these things will belong to the past. Thanks also to Delphine, Tine and Bart who engaged in this PhD process at the same moment, and to Marie, Bernd, Jonathan, Isabel and Nele. I wish all my colleagues the best for the finishing of their own dissertations and I’m looking forward to be invited at their future defences. Karin and Carole, thanks for the practical help and support.

Next, it would be ungrateful when I would forget all my friends who indirectly lent a helping hand during these academic years. I especially remember the great times I had with my soccer team Maptiko and all people in Ghent and Leuven, and, off course, the wonderful moments I experienced with my friends in Wimmertingen, where I was always welcome to have some fun.

A special word of thanks goes to my mum and dad for their love and their continuous support during my stay at many academic institutions, at different places throughout the country. You both were the cornerstone of my realizations and gave me the chance to make the most of my opportunities. Thanks also Katleen, Eva and Frédéric for your ongoing moral support and interest.

The last but most important gratitude goes to Greet. Thanks for your love and absolute support during all these years. You were my help and stay at moments when I needed to give life variety and think about other things than my research (or when trains were delayed). In the near future we’ll get the chance to express our everlasting love so we can stay together for many many years. I’m very lucky and proud to be at your side. I wish you the best during the last miles of your own PhD and I’m sure you’ll finish it in a few months with splendid results.

TABLE OF CONTENTS

SAMENVATTING

Marketing management ondergaat sedert enige tijd een evolutie van een productgerichte naar een meer klantgerichte aanpak. In die ontwikkeling gaat er veel aandacht naar hoe marketingacties individueel kunnen worden aangepast aan de noden van de klant, vermits zowel de klant als het bedrijf hier wel bij varen. Dit doctoraal proefschrift onderzoekt op welke manier huidige toepassingen in het domein van klantgerichte marketing kunnen worden verbeterd door gebruik te maken van analytische, predictieve technieken. De focus ligt daarbij op bedrijven uit de 'retail' sector waar klant en bedrijf niet gebonden worden door middel van een contract.

Eerst onderzoeken we aanpassingen aan huidige methodes die gebruikt worden voor het uittekenen van direkt marketing strategieën, zowel voor traditionele ondernemingen als voor ondernemingen actief via het Internet. Het succes van een direkt mailing wordt onder meer bepaald door de accuraatheid waarmee de toekomstige opbrengst van elke klant kan worden ingeschat. Wij stellen een geavenceerde methode voor die enkel rekening houdt met de netto impact van een marketingactie en gebruiken predictieve modellen om elk van de elementen, die deel uitmaken van deze berekening, te modeleren. Hierdoor bekomen we een verbeterde segmentatie van de klanten en worden marketingkosten sterk gereduceerd. De implementatie van onze methode in het mailingproces van een Europese retailer toonde aan dat het aantal te versturen mailings met vijfenzestig procent kon worden verminderd, terwijl de totale winst van het bedrijf toch met vijf procent steeg.

Hoewel in vele studies het gebruik van waardebonnen wordt ondersteund om producten te promoten, is het uiteindelijk gebruikspercentage van dit medium erg laag. Wij onderzochten in welke mate analytische modellen een oplossing kunnen bieden om het gebruik van dit promotiemiddel individueel te voorspellen en zo de verdeling ervan doelgerichter te organiseren. Bovendien zorgen afzonderlijke modellen voor waardebonnen uitgegeven door toeleveranciers en eindverdelers ervoor dat ze elk hun eigen klantensegment beter kunnen detecteren om zo onderlinge concurrentie te vermijden.

De CRM-mogelijkheden voor e-commerce zien er veelbelovend uit. Deze ondernemingen beschikken over veel meer individueel klantengedrag dankzij de registratie van het

surfgedrag op het Internet. Het huidig koopgedrag via Internet is echter nog erg beperkt. Daarom onderzochten we welk van het geregistreerd klantengedrag het aankoopgedrag van klanten bepaalt. Onze bevindingen tonen aan dat zowel algemeen als gedetailleerd klikgedrag van belang zijn om koopgedrag tijdens een toekomstig bezoek aan de website te voorspellen, waardoor ook e-commerce managers zinvolle veranderingen kunnen aanbrengen aan hun individueel gerichte marketingstrategieën.

Trouwe klanten beschikken over een aantal voordelen die belangrijk zijn voor de groei, de winstgevendheid en de toekomst van een onderneming. Huidige marketingplannen kunnen echter moeilijk rekening houden met het getrouwheidsniveau van de klant omdat deze informatie niet beschikbaar is voor een onderneming. Wij stellen twee methodes voor om trouwe klanten in een klantenbestand op te sporen. De ene methode bepaalt trouwe klanten louter op basis van twee gedragskenmerken uit de interne database. De andere methode maakt gebruik van een enquête, uitgestuurd naar een beperkt aantal klanten van een onderneming, om via predictieve modellen en gegevens uit de database, het getrouwheidsniveau van alle klanten in te schatten.
Daarenboven tonen we aan hoe deze informatie nuttig kan worden aangewend om doelgerichte marketingacties uit te stippelen. We ontwikkelen een methode voor retailers om trouwe klanten te detecteren die in de nabije toekomst hun aankopen volledig of gedeeltelijk bij de concurrentie zullen maken. Ten tweede stellen we de momenteel toegepaste beloningsprogramma's in vraag vermits zij vooral herhalingsaankopen belonen en stimuleren. Onze alternatieve aanpak beloont klanten voor meerdere van hun voordelen tegelijk, door beloningen te verdelen volgens hun echte of voorspelde trouw.

We maken gebruik van verschillende analytische technieken in elk van de behandelde topics: multivariate regressiemodellen, logistische regressiemodellen, beslissingsbomen, Random Forests en neurale netwerken. In elke studie werd een uitgebreide set van klantenvariabelen in rekening genomen om de performantie van de modellen te vergroten en om de relevantie van de verschillende datatypes te kunnen evalueren. Tenslotte, om de predictieve kracht van de modellen te verhogen, of om de ideale combinatie van variabelen te bepalen die de hoogste performantie verzekert, pasten we uiteenlopende variabelenselectietechnieken toe.

SUMMARY

Contemporary marketing management has experienced an evolution from a product-oriented to a customer-oriented policy. In that development, targeted marketing has gained a lot of attention, since both customers and companies are benefited by customized marketing actions. This doctoral dissertation examines in what way current targeted marketing activities, in a non-contractual retail setting, can be enhanced by making use of predictive modeling techniques.

First, we present alterations to current direct marketing policies for traditional and online retailers. The success of a direct mailing campaign is dependent on the accuracy by which customers' future contributions can be estimated. An advanced method is provided which accounts only for the net effect of a targeting action and predictive models are developed to estimate each element of the profit function. An improved ranking of the customers in the segmentation list and a reduction of the optimal mailing depth bring about increased company profits. The implementation into the direct mailing system of a European retailer showed a reduction of the number of mailings by sixty-five per cent while profits augmented by five per cent.

Though several studies support the use of coupons to advertize products, redemption rates are low. We examined to what extent predictive models can be employed to define customer proneness for manufacturer and retailer coupons, so both parties are able to identify target segments and a competitive battle can be moderated. As a result, the entire customer base can be divided into four segments.

For e-commerce, CRM opportunities look promising. Much more customer data are available thanks to the registration of customer behavior on the Internet and client relations can be outlined in a dynamic way. However, current online purchase behavior is rather limited. Therefore we examine the features that control site visitors' decision whether or not to make purchases. Our findings indicate that general and detailed clickstream behavior are useful for modeling future purchase intentions which provides a powerful tool for managers to fine-tune targeting strategies.

Loyal customers exhibit beneficial behavior which is important for the growth, the profitability and the continuity of a company. However, the planning of targeted marketing actions towards these clients is not straightforward since, typically, no information concerning customers' loyalty is available. We specify two methods to track loyal customers. First, we define loyals based on two behavioural attributes derived from the internal database. Next, we enrich this information with survey data from a limited number of clients in order to build predictive loyalty models for the entire customer base. The results point to the ability of marketing management to detect loyal customers to an acceptable degree.

Besides, we examine how this additional information can be usefully applied for targeted marketing purposes. We present a feasible method for companies to detect which of their loyal customers have the intention to switch their purchases towards competitors. We introduce the aspect of partial defection in order to signal disadvantageous intentions as early as possible. Secondly, we question the effectiveness of current loyalty programs. Whereas these methods reward and stimulate especially repeat-purchase behavior, we suggest to compensate customers in proportion to their true or predicted loyalty since these criteria consider different loyalty benefits at the same time.

Several different analytical techniques were used to resolve each of the targeting problems: multiple linear regressions, logistic regressions, decision trees, Random Forests and neural networks. In each study, we employed an elaborate list of customer attributes to explain as much as possible of the model variance and to evaluate the relevance of different variable types. To increase predictive power or define the optimal combination of inputs, we made use of several feature-selection techniques.

LIST OF TABLES

Chapter V

Chapter VI

Appendices

LIST OF FIGURES

Appendices

BACKGROUND AND OBJECTIVES

1. INTRODUCTION

This work presents contributions with respect to the optimisation of targeted marketing for retailers in a non-contractual setting. It examines how companies can improve their marketing strategies by using their internal database information and analytical models.

1.1 Targeted marketing

Since a few decades, it is a well-supported fact that customization of marketing activities carries high potential (Rossi et al., 1996). Customers are, by definition, not homogenous and differ with regard to their characteristics and preferences. As a result, they require customized treatment by which companies try to define different customer segments in order to approach each of them with adapted marketing actions. The evolution in marketing, from a product-oriented to a customer-oriented view, is generally acknowledged as “the paradigm shift in marketing” (Brodie et al., 1997). In that discourse, the concept of customer relationship management (CRM) was introduced, which is “an enterprise approach to understanding and influencing customer behavior through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability” (Swift, 2001). It is based on the principle that focusing on individual customers is the best way to win, retain and increase company business. Firms try to learn what their customers want and tailor their marketing strategy accordingly (Brown, 2000).

The proceeds of customer targeting go to both customers and companies. In general, mistargeted communication wastes environmental resources (Gönül et al., 2000). First, clients identify inappropriate actions as interfering, which might endanger a good relationship and undermine efforts to build loyalty and trust (Malthouse, 1999). Moreover, Campbell et al. (2001) expect an increasing satisfaction and retention rate if customers receive a more suitable treatment. This has an immediate effect on companies' performance considering the theorem that acquiring new customers is several times more expensive than retaining existing ones (Rosenberg and Czepiel, 1984).
Second, it is beneficial for a company if it is able to define whom to address with specific marketing activities. Considering a company with a response rate of 30% when targeting the entire customer base, marketing costs can be diminished by 70% if it would be able to target without failure (Nash, 1994).

1.2 Database marketing

The cost effectiveness of such targeted marketing can be increased by making marketing decisions based on internal database information (Roberts and Berger, 1989, p147). Considering the definition according to Roberts (1997), database marketing is the management of marketing activities by using individually-stored customer information in combination with analytical capabilities and information technology.

Typically, internal database information consists of socio-demographic characteristics, purchase behavior, information concerning marketing actions, satisfaction data and any kind of interaction information (Verhoef et al., 2002). That way database marketing attempts to provide the ultimate - individual - customer segmentation. The availability of these data and the employment of analytical models enable companies to retrieve individual information about the future behaviour of their clients. Thereby, modeling is a commonly used technique (Desarbo and Ramaswamy, 1994) and is proved to be a profitable tool in fine-tuning direct marketing strategies (Elsner et al., 2004).

In summary, database marketing can be evaluated as the application of data mining for marketing management, while data mining is the "discovery stage" of the KDD (knowledge discovery in databases) process, which can be described as "the non-trivial extraction of implicit, previously unknown, and potentially useful knowledge from data" (Adriaans and Zantinge, 1996, p5).

This way of marketing management has experienced acceleration during the last decade(s) and is among the fastest growing channels of marketing (Zahavi and Levin, 1997). The constant reduction of storage costs, the ever-increasing computer power and the rising number of software packages are the main sources of this development (Bult and Wansbeek, 1995; Rossi et al., 1996; Bult 1993). So more individual data can be collected and more available data involves better segmentation opportunities and more profit (Rossi et al., 1996). Besides, in accordance with the findings of the Direct Marketing Association, companies are coming to realize that the use of quantitative techniques improve customer relationships and, therefore, have a positive effect on their profits (The Direct Marketing Association, 2004).

1.3 Research topics

This work deals with different topics about the enhancement of customer targeting strategies for marketing purposes by using predictive models. It can be divided into two main parts. The first part discusses direct marketing strategies in traditional and online store environments. The second part examines the value of detecting loyal customers for targeted marketing and how marketing management can track this specific type of customers.

1.3.1 Direct marketing

Whereas database marketing covers different kinds of analyses to create strategies for marketing, direct marketing is one of these research topics which define individual communication and distribution strategies to increase customer response (Tapp, 2005).
Roberts and Berger (1989) define direct marketing as:

> "Direct marketing is an interactive system of marketing which uses one or more advertising media to effect a measurable response and/or transaction at any location."

At its infancy direct marketing was mainly applied in nonstore settings where particularly mail-order companies made use of database marketing to customize mailing activities (Spiller and Baier, 2004). But, also traditional retailers started to apply these techniques in order to take care of their promotional actions. For many years, the Direct Marketing Association reports an increase of the direct marketing industry, which is growing faster than all other sectors (Spiller and Baier, 2004).

Direct marketing has gained a lot of attention in customer relationship management and literature has already tackled different aspects to optimize mailing strategies. The advertising literature confirms two particular types of advertising media to be of major importance: the distribution of mailings (The Direct Marketing Association, 2004) and the issue of coupons (Bawa, Srinivasan and Srivastava, 1997).

1.3.1.1 Direct mailing

Consequently, the first study in this PhD thesis discusses the optimization of methods that are currently used in direct mailing. The most important element, which defines the success of a direct marketing campaign, is the definition of the mailing list (Bult and Wansbeek, 1995). Many studies discussed this crucial topic but still, to our knowledge, several contributions can be made concerning the exploitation of the profit function, which is necessary to define individuals' value. Besides, most literature makes no expectations about customers' behavior in case no marketing action is undertaken. We consider both shortcomings by building four different predictive models to estimate each element of the profit function. The presented method is empirically tested with real-life data of a European retailer in fast moving consumer goods (FMCG) and durables.

1.3.1.2 Coupon dispensing

Manufacturers and retailers are investing heavily in the distribution of coupons in order to convince as many customers as possible to buy their products. Literature shows that the supply of coupons has a positive effect on customer behavior: it increases and accelerates product usage (Taylor, 2001) and convinces customers to switch brands (Bell, Jeongwen and Padmanabahn, 1999). In contrast, the redemption rate of coupons is very low. As a consequence, manufacturers and retailers are plunged in a battle. However, literature suggests that both distributors possibly can avoid each other since their product assortment appeals to different customer segments (Ailawadi, Neslin and Gedenk, 2001). We examined the composition of separate response models, for the use of each coupon type, in order to advance both parties' targeting strategy. Results are validated on data of a worldwide retailer in FMCG.

1.3.1.3 Online direct marketing

The advent of the Internet has changed the distribution possibilities substantially since firms are able to offer products in an online virtual store. Typically, online stores have much more customer data at their disposal than traditional retailers (Moe and Fader, 2002) and Internet choice behavior seems different from the explored behavior in conventional store-retail settings (Bucklin et al., 2002). Besides, companies can maintain customer relationship through their website which means they can better outline client relations (Bauer et al., 2002). Therefore, we examined to what extent targeted marketing and customer selection are also possible in an online environment. In this respect, it is interesting to get insight into which available customer data are important for online marketing management and whether or not online stores have an advantage in modeling customer responses because they can track more customer information.

1.3.2 Loyal customers

An elaborate list of marketing literature supports the value of loyal customers. They increase their spending over their lifetime (Reynolds and Arnold, 2000), they make positive recommendations to their relatives (Reichheld, 2003), they can be served at diminished costs (Dowling and Uncles, 1997), they exhibit a lower responsiveness to competitive pull (Stum and Thiry, 1991), they become price insensitive and have a positive impact on company's employees (Reichheld and Sasser, 1990). Consequently, it is important to keep these clients into a company's customer base. So, specific targeted marketing actions towards this segment seem appropriate. This thought is confirmed by the development of loyalty or reward programs, which are conceived to reward and to stimulate such desirable customer behavior (Kivetz and Simonson, 2003; Dowling and Uncles, 1997). Indeed, Reichheld (1996) argues that strategies should be in line with the relationship potential of customers.

So, for companies it is valuable to obtain knowledge about customers' loyalty in order to incorporate this information into their targeting strategy. However, these data are not readily available in transactional databases since typically firms have no information about customers' behavior at competitive stores. Consequently, we investigated how firms can define the loyalty level of their clients. A first approach splits the database into a loyal and a nonloyal segment based on the internal customer data. A second approach makes use of data enrichment and predictive modeling to determine individual loyalty scores.

Companies can take advantage of this loyalty information for their targeted marketing. First, we already justified why it is valuable to keep loyal customers in the customer base. In contrast, relationships are transitory and competition is fierce. In a non-contractual setting, clients do not signal when they are switching their purchases to competitors or when they tend to totally abandon their relationship. So, one of our studies investigates to what extent predictive models can be built to track loyal customers who will (partially) defect in the near future.

Secondly, reward or loyalty programs tend to stimulate and recompense beneficial behavior of loyal clients. However, in practice, companies are rewarding mainly repeat-purchase behavior since their remuneration criteria are customer spending and/or length-of-relationship. Our study examines whether companies are able to compensate for the other loyalty benefits by making use of another reward criterion. Therefore we employ a predictive model, which is built based on information from the internal transactional database and a survey that was sent to a small sample of customers. The results are validated in two different settings by collaboration of a European retailer that offers FMCG and durables.

1.4 Non-contractual setting

All our studies were validated with data of retailers active in a non-contractual setting. In this environment customers can easily change their purchase behaviour to competitors without being confronted with high switching costs and without informing the company about it. This enhances competition: AC Nielsen reported in 2001 that in FMCG retailing, more than seventy per cent of all customers shop around in different supermarkets.

2. OBJECTIVES

The main objective of this work is to present improved methods for targeted marketing in a non-contractual retail setting. Several studies examine different topics within the domain of direct marketing and loyalty management, each supported with specific literature. The respective main objectives of these studies are considered in the subsequent paragraphs.
For direct mailing purposes, we examine to what extent it is possible to compose an advanced profit function, which calculates customers' value based on the outcomes of different predictive models. Simultaneously, we want to evaluate the value of only taking into account the net effect of a direct marketing action instead of its total effect on company's profit.
In order to increase coupon redemption rates, we investigate whether it makes sense to build predictive models to compose companies' targeting list for coupon distribution. In addition, the research studies the possibility to use separate models for manufacturers and retailers to avoid that both parties get stuck into a battle for the same set of customers, which possibly enforces unnecessary competition.
Additionally, we research whether targeting is feasible also for online retailers and if they are able to infer the future goal of their site visitors in order to adapt their targeting strategy. Moreover the study aims to conclude whether online retailers are in an advantage compared to traditional retailers since the former have much more data available for predictive purposes.
Since loyal customers are valuable for the growth, continuity and profitability of a company, a next study aims to predict which loyal customers will partially leave a company so that companies can anticipate such defective behavior. Further, we study how to include customer loyalty into the targeting scheme by making predictive models based on internal company data and a survey that is administered to a limited set of customers. Lastly, loyalty programs are evaluated by investigating to what extent the use of predicted loyalty is able to remunerate loyal clients for the benefits they deliver compared to currently applied reward programs.

Each of these studies aspires to maximize the performance of the analytical models. Therefore, in most studies, we benchmarked several analytical techniques in order to select the one that delivers the best predictive power. We examine multiple linear regressions, logistic regressions, C4.5 decision trees, Random Forests and Automatic Relevance Determination neural networks.

Furthermore, in each of the cases, we incorporated as much explanatory variables as possible, from different variable types. However, if many predictors are included into a model, the estimation set can suffer from overfitting problems, which results in a decrease of the outcome when validated. As a consequence, we examine to what extent different feature selection techniques can increase predictive power and avoid overtraining. We made use of procedures like Forward and Backward selection, the global score algorithm of Furnival and Wilson (1974) and Relief-F.

Finally, for each of the topics, it is valuable to define which of the variables are relevant for the problem at hand. An evaluation of the variable selections and an interpretation of variables' importance are made in the last chapter of this work.

3. DATA COLLECTION

The realization of this work required the examination of different databases. All information was provided by diverse retailers having Belgian, European and worldwide outlets, which transferred their entire database to one of our servers at Ghent University. It all concerned unprocessed data at the individual customer/orderline level so a lot of effort was put in preparing it for analysis. Further, we collected additional information by means of a self-administered survey for data enrichment purposes. In total 3000 questionnaires were distributed in two different retail settings.

4. REFERENCES

Adriaans, P. and Zantinge, D., 1996. *Data Mining*. Harlow, U.K.: Addison-Wesley.

Ailawadi, K.L., Neslin, S.A. and Gedenk, K., 2001. Pursuing the value-conscious consumer: Store brands versus national brand promotions, *Journal of Marketing*, 65(1), 71-89.

Bauer, H.H., Grether, M. and Leach, M., 2002. Customer Relations through the Internet, *Journal of Relationship Marketing*, 1(2), 39-56.

Bawa, K., Srinivasan, S.S. and Srivastava, R.K., 1997. Coupon Attractiveness and Coupon Proneness: A Framework for Modeling Coupon Redemption, *Journal of Marketing Research*, 34(4), 517-525.

Bell, D.R., Jeongwen, C. and Padmanabahn, V., 1999. The Decomposition of Promotional Response: An Empirical Generalization, *Marketing Science*, 18(4), 504-526.

Brodie, R.J., Coviello, N.E., Brookes, R.W., and Little, V., 1997. Towards a Paradigm Shift in Marketing? An Examination of Current Marketing Practices, *Journal of Marketing Management*, 13(5), 383–406.

Brown, S.A., 2000. *Customer Relationship Management*. John Wiley & Sons, Canada.

Bucklin, R.E., Lattin, J.M., Ansari, A., Gupta, S., Bell, D., Coupey, E., Little, J.D.C., Mela, C., Montgomery, A. and Steckel, J., 2002. Choice and the internet: From clickstream to research stream, *Marketing Letters*, 13(3), 245-258.

Bult, J.R., 1993. Semiparametric Versus Parametric Classification Models: An Application to Direct Marketing, *Journal of Marketing Research*, 30(August), 380-390.

Bult, J. R. and Wansbeek, T., 1995. Optimal Selection For Direct Mail, *Marketing Science*. 14(4), 378-394.

Campbell, D., Erdahl, R., Johnson, D., Bibelnieks, E., Haydock, M., Bullock, M. and Crowder, H., 2001. Optimizing Customer Mail streams at Fingerhut. *Interfaces,* 31(1), 77- 90.

Desarbo, W.S. and Ramaswamy, V., 1994. CRISP: Customer Response Based Iterative Segmentation Procedures for Response Modeling in Direct Marketing, *Journal of Direct Marketing*, 8(3), 7-20.

Dowling, G.R. and Uncles, M., 1997. Do Customer Loyalty Programs Really Work? *Sloan Management Review*, 38 (Summer), 71–82.

Elsner, R., Krafft, M. and Huchzermeier, A., 2004. Optimizing Rhenania's Direct Marketing Business Through Dynamic Multilevel Modeling (DMLM) in a Multicatalog-Brand Environment, *Marketing Science*, 23(2), 192-206.

Furnival, G.M. and Wilson, R.W., 1974. Regressions by Leaps and Bounds, *Technometrics*. 16, 499-511.

Gönül, F.F., Kim, B.-D. and Shi, M., 2000. Mailing smarter to catalog customers, *Journal of Interactive Marketing*, 14(2), 2-16.

Kivetz, R. and Simonson, I., 2003. The Idiosyncratic Fit Heuristic: Effort Advantage as a Determinant of Consumer Response to Loyalty Programs, *Journal of Marketing Research*, 40(4), 454–467.

Malthouse, E.C., 1999. Ridge Regression and Direct Marketing Scoring Models, *Journal of Interactive Marketing*, 13(4), 10-23.

Moe, W., Fader, P.S., 2002. Dynamic Conversion Behaviour at e-Commerce Sites, *Wharton Marketing Department Working Paper.*

Nash, E.L., 1994. *Direkt Marketing: strategy, planning, execution*. McGraw-Hill, 3rd edition, US.

Reichheld, F.F., 1996. Learning from customer defections, *Harvard Business Review,* 74(2), 56-69.

Reichheld, F.F., 2003. The One Number You Need to Grow, *Harvard Business Review*, December, 46–54.

Reichheld, F.F. and Sasser, E.W., 1990. Zero Defections, Quality Comes to Services, *Harvard Business Review*, 68 (September-October), 105–11.

Reynolds, K.E. and Arnold, M.J., 2000. Customer Loyalty to the Salesperson and the Store: Examining Relationship Customers in an Upscale Retail Context, *Journal of Personal Selling and Sales Management*, 20 (2), 89–98.

Roberts, M.L., 1997. Expanding the role of the direct marketing database, *Journal of Direct Marketing*, 11(4), 26-35.

Roberts, M.L and Berger, P.D., 1989. *Direct marketing management*. US: New Jersey: Prentice Hall inc. .

Rosenberg, L.J. and Czepiel, J.A., 1984. A Marketing Approach to Customer Retention, *Journal of Consumer Marketing*, 1, 45–51.

Rossi, P.E., McCulloch, R.E. and Allenby, G.M., 1996. The Value of Purchase History Data in Target Marketing, *Marketing Science*, 15(4), 321-340.

Spiller, L.D. and Baier M., 2005. *Contemporary direct marketing*. US, New Jersey: Pearson Prentice Hall.

Stum, D. and Thiry, A., 1991. Building customer loyalty, *Training and Development Journal,* 45(4), 34-36.

Swift, R.S., 2001. Accelerating Customer Relationships, using CRM and Relationship Technologies. U.K., London: Prentice Hall.

Tapp, A., 2005. *Principles of direct and database marketing*. Pearson Education Limited, 3rd edition, U.K. .

Taylor, G.A., 2001. Coupon Response in Services, *Journal of Retailing*, 77(1), 139-151.

The Direct Marketing association, 2004. *The DMA quarterly Business Review*, 4th quarter.

Verhoef, P.C., Spring, P.N., Hoekstra, J.C. and Leeflang, P.S.H., 2002. The Commercial use of segmentation and predicitve modeling techniques for database marketing in the Netherlands, *Decision Support Systems*, 34, 471-481.

Zahavi, J. and Levin, N., 1997. Applying Neural Computing to Target Marketing, *Journal of Direct Marketing,* 11(4), 76- 93.

CHAPTER I

ASSESSING AND EXPLOITING THE PROFIT FUNCTION BY MODELING THE NET IMPACT OF TARGETED MARKETING[1]

[1] This chapter is based on the following reference: Wouter Buckinx, Dirk Van den Poel. Assessing and Exploiting the Profit Function by Modeling the Net Impact of Targeted Marketing, to be submitted to *Management Science*.

CHAPTER I:

ASSESSING AND EXPLOITING THE PROFIT FUNCTION BY MODELING THE NET IMPACT OF TARGETED MARKETING

1. INTRODUCTION

For many years marketers have recognized direct marketing as an effective and efficient way of communicating with customers. However, it seems that it has not yet reached the height of its power and is still coming to full bloom. Since the foundation of their Quarterly Business Review in 2002, the Direct Marketing Association (DMA) reported for the sixth consecutive quarter a positive expansion of the direct marketing industry (The Direct Marketing association 2004). The last reported figures of 2004 show a record growth index and direct marketers are expecting this trend to continue in 2005. Moreover, currently, more than 50 per cent of all advertisement expenditures are made on direct marketing.

Several reasons can be found to support this continuing development. Most authors ascribe the progress to the constant reduction of data storage costs, the available amount of computing power and the rising number of software packages (Bult and Wansbeek 1995, Rossi et al. 1996, Bult 1993). These trends enable companies to collect more and more individual (detailed) customer data, so more well-founded decisions can be taken. Besides, and maybe even more important, companies are realizing that the employment of these facilities and the implementation of innovative modeling techniques to improve customer

relationships, have a positive effect on profitability and sales (The Direct Marketing Association 2004). So, the research for better procedures and techniques goes on.
Not coincidentally, direct marketing has gained so much attention in customer relationship management (CRM) literature during the last decades. Several studies already tackled different aspects of direct marketing in order to optimize mailing strategies. Response modeling is a well known and commonly used technique by direct marketing analysts (Desarbo and Ramaswamy 1994). It has proven to be a profitable tool in fine-tuning direct marketing strategies (Elsner et al. 2004) since even small improvements attributed to modeling can create great financial gains (Malthouse 1999).

Different elements define the success of a direct marketing campaign. Bult and Wansbeek (1995) consider the most important one to be the composition of the mailing list. Many authors confirm this theorem (Levin and Zahavi 2001, Bitran and Mondschein 1996, Bhattacharyya 1999). Bitran and Mondschein (1996), for example, put it this way:

> ”One of the most important decisions that a manager must make in the catalog sales industry is defining the mailing policy, i.e., which rental lists to employ and the fraction of the people in those lists that should receive a catalog”.

So, basically, such selection boils down to two major steps: first, for each customer, one has to define how useful it is to send him or her a mailing and, secondly, a meaningful cut off point needs to be set to determine the number of customers to be targeted (mailing depth). Evidently, all these steps have to be taken while keeping in mind the maximization of company profits (Bhattacharyya 1999).
A good many studies discussed one or both of the above mentioned steps. However, to our knowledge, the currently proposed procedures are still open to improvement. Nearly all of the examined studies recognize the importance of profit functions to resolve their targeting challenge (step 1 and step 2). A profit function is applied to balance revenues and costs of a direct mailing to determine valuable targets (see next section). However, none of the studies is employing the possibilities of predictive modeling to substitute all of the elements in these functions. As a consequence, the solutions provided concerning steps 1 and 2, can still be optimized.

First, the majority of the studies concerning direct mailing are disregarding the determination of the optimal mailing depth. Secondly, most researches are only making use of purchase propensity and are neglecting the level of expenditures to determine customer value. Finally, retailers are generating traffic by distributing catalogs to a subset of their customers. However, in this setting it is common practice that also customers who were not targeted do make purchases. If a company wants to be efficient in its targeting, such customer behavior should be integrated into the profit function in order to optimize the justification of outgoing mailings: only the net effect of a marketing action on company profit needs to be considered. Throughout this paper, this last phenomenon is referenced as the ‘clearance’ of customer profit. All these shortcomings are considered crucial when companies aim to maximize profit. To the best of our knowledge, no such a study exists which exploits the full potential of modeling each item of individual expected profit functions when defining a customer list and the optimal mailing depth for direct marketing purposes. Certainly not when the profit function only accounts for the ‘net’ effect of sending a mailing.

This article is organized as follows: Section 2 reviews the existing literature concerning list segmentation, the determination of mailing depth and cleared profits. We point to the existing gaps in direct marketing literature from which the contributions of this paper arise. Section 3 explains the methodology we applied and gives mathematical details of our models. Our real-life application is explained in Section 4. Section 5 considers the results and Section 6 ends this paper with conclusions, a discussion and issues for further research.

2. BACKGROUND AND LITERATURE REVIEW

2.1 Profit function

The existing literature concerning direct marketing has shown a tremendous growth during the last decades. Many authors recognize the traditional procedure of composing a mailing selection: score and rank customers in accordance with their usefulness and choose the ideal depth of the target list. Regardless of the scoring technique used, the mathematical computation of the customers’ value involves the consideration of an expected profit function. An early article of Magidson (1988) about direct marketing already stated that, when one needs to define the depth of a mailing and profits are the purpose, a financial analysis should be performed by making use of the outputs of the scoring models. Bult (1993) makes this idea more concrete and poses that only the people should be mailed who's expected contribution margin is higher than the cost of the mailing. These thoughts result in the following general acknowledged individual profit function:

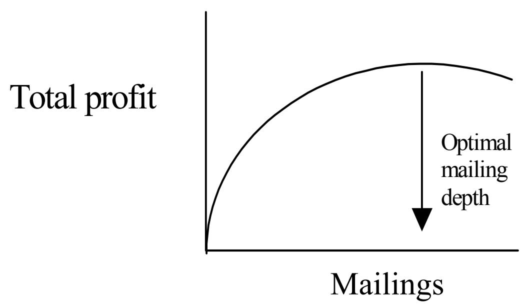
$$\pi_i = R_i \cdot M - C \qquad (1)$$

Where 'π_i' is the profit or the contribution of customer 'i', 'R_i' equals the individual revenue, 'M' is the general margin of the company and 'C' is the cost of sending the mailing. Customers' revenue can be subdivided (equation (2)):

$$\pi_i = (E_i \cdot P_i) \cdot M - C \qquad (2)$$

In this profit function, 'P_i' is the customer's probability of purchasing and 'E_i' represents the customer's individual expenditures when a visit is made. If the profit is positive it is wise to put the particular customer in the mailing list. Consequently, if customers are ranked in accordance with the individual profit functions, management should invest in sending mailings up to the point of diminishing overall returns (Campbell et al. 2001)(see Figure 1).

Figure 1.1: Optimal mailing depth curve.



The better the expected probabilities and expenditures reflect customer's 'real' behavior, the better customers can be ranked according to their contribution and the better the optimal mailing depth point can be defined. Most of the researches, however, only made use of predictive models to define the propensity of purchasing (P_i) (Gönül et al. 2000, Hansotia and Rukstales 2002, Gönül and Shi 1998, Bult and Wansbeek 1995, Muus et al. 1996, Bult 1993, Bauer 1988, Magidson 1988). Whereas the assessment of individual customer expenditures (E_i) is just as crucial to get a more accurate expectation of customers' profit. More specifically, some studies totally ignore the expenses (E_i) in the profit function so no meaningful evaluation can be made concerning expected revenues and the cut off point in the target list must be set arbitrary or is defined by budget constraints (Gönül et al. 2000, Bhattacharyya 1999, Bult 1993, Bauer 1988, Magidson 1988, Prinzie and Van den Poel 2005). Other studies do include an average expenditure that is calculated across all customers (Elsner et al. 2004, Gönül and Shi 1998, Bult and Wansbeek 1995, Muus et al. 1996). Still, an average does not reflect the variance of the purchase levels across customers. Furthermore, Bult and Wansbeek (1995), underline the inclusion of heterogeneity in customer returns in their issues for future research. A few studies do make predictions of customers' expenses. But only the study of Campbell et al. (2001) uses this information to complete all parts of the profit function and to define the depth of their mailing. Bhattacharyya (1999) who uses genetic algorithms to model profit is restricted to budget constraints while Malthouse (1999) who applies ridge regression does not use this information to finish step 2, the determination of the mailing depth.

2.2 Cleared profits

Besides, we want to stress that the most prevalent objective of direct marketing procedures is to increase cost efficiency by precluding superfluousness of mailings being sent (Elsner et al. 2004). Certainly in retail settings, customers are able to make purchases even if they did not receive a mailing or catalog. So targeting such customers is a waste and more profits can be made when these customers can be left out of the target list. This study proposes to extend profit function (2) to take such behavior into account by including the purchase probability and the expected expenditure in case an individual does not get a mailing. That way, the expected profit is discounted in accordance with the propensity of purchasing and the related expenditures of each individual when (s)he is not being mailed. This addition is valuable to the extent that customers are able to make purchases without being targeted.

Only a few recent direct marketing studies did cover compensations for such kind of customer behavior. Gönül, Do Kim and Shi (2000) use a ratio of two hazard function models in order to decrease similar wasteful mailings. However, they do not consider heterogeneity of expenditures across customers. Besides, they make no difference between the spending level of mailed and not mailed customers, whereas we expect the spending of mailed customers to differ from the spending of customers who did not receive a catalog. Hansotia and Rukstales (2002) calculated individual net incremental expected profits but also focused on purchase propensity only and did not take into account expected revenue. Finally, Campbell et al. (2001) made use of a saturation matrix to compensate for the impact of earlier catalogs for future purchases. This matrix, which is the outcome of a timing matrix and a similarity matrix, results in a discount factor which is not individualized but equal for all customers. All of these studies point to the importance of compensation for customer behavior when no treatment is performed. However, none of them fully exploit the elements of the profit function.

A summary of the literature shows that none of the present studies makes use of a profit function where: a) both purchase propensity and expected revenue are substituted by means of individual prediction models; b) customers' contribution is discounted for their behavior in case no treatment would occur. In contrast, we are convinced that these shortcomings have a serious impact on the customer ranking (step 1) and on the optimal depth of mailing (step 2), whereas both steps are considered to be among the most important in direct mailing strategies. These gaps can be checked in Table 1, which gives an overview of the cited studies. It is not our intention to give an exhaustive overview of all previous work in the area of direct marketing. To reduce the number of references, this table focuses only on studies that explicitly considered a procedure to define optimal mailing depths, modelled customer expenditures or considered some kind of profit clearance. It shows which techniques were applied for each of the predictive models and how the results were evaluated. The table highlights the contributions of this paper.

Our study has several extensions for the existing literature. We propose a profit function in which individuals are evaluated depending on the 'net' effect of a mailing. Besides, we are the first to substitute each item of such an advanced profit function, which implies that we use four different predictive models. The contributions of using individual predictions instead of substituting average expenses are shown. Three different predictive techniques are analyzed: multiple regression, logistic regression and Random Forests. A variable selection technique is used to overcome overfitting problems. In addition, for each of the response models we detect the most important predictors in order to define which customer behavior is essential when making purchase predictions with and without sending a mail. Finally, to evaluate the results, we implemented our findings in a real-life experiment where we were able to manipulate an entire mailing stream of the collaborating company.

Table 1.1: Literature review.

Author	Techniques				Evaluation level
	Prediction probability (treatment)	Prediction expenses (treatment)	Prediction probability (no treatment)	Prediction expenses (no treatment)	estimation/ validation/ field test
Bhattacharyya (1999)	-	genetic algorithm	-	-	estimation/validation
Bult and Wansbeek (1995)	CHAID, logit, profit max., semiparam. method	-	-	-	estimation/validation
Campbell, Erdahl, et al. (2001)	regression and linear programming	regression	-	-	field test
Elsner, Kraft and Huchzermeier (2004)	markov-chain, CHAID	-	-	-	field test
Gönul, Kim and Shi (2000)	proportional hazard function	-	ratio of hazards	-	estimation/validation
Gönül and Shi (1998)	algorithm of probits	-	-	-	estimation
Hansotia and Rukstales (2002)	logit, CHAID	-	profit function	-	tain/validation
Malthouse (1999)	-	ridge regression	-	-	estimation/validation
Muus, Van der Scheer and Wansbeek (1996)	bayes (normal posterior, laplace, MCMC)	-	-	-	estimation/validation
This study	logit, random forests	regression, random forests	logit, random forests	regression, random forests	estimation/validation/field test

3. METHODOLOGY

3.1 Profit function

The proposed optimization of direct mailing campaigns comes down to an adaptation of customers' expected profit function (2) by taking into account purchase probabilities and expected expenditures with and without treatment. So, we need for each customer two different probabilities and two different expenditures. P_i^m, being the purchase propensity after receiving a catalog; P_i^n, being the purchase propensity if no catalog is received; E_i^m, the expenditures when a mailing is sent and purchases are made and E_i^n, the expenditures when the individual receives no mailing but does make purchases. Such a decomposition of company revenue can also be found in a recent study by van Heerde and Bijmolt (2005). Since we want to maximize the profitability of our entire customer base, the mathematical representation of our decision problem becomes:

$$Max\sum_{i=1}^{n}\left[((E_i^m \cdot P_i^m)\cdot M)\cdot x_i-(C\cdot x_i)+((E_i^n \cdot P_i^n)\cdot M)\cdot(1-x_i)\right] \quad (3)$$

$$\sum_{i=1}^{n} x_i \le T \quad (4)$$

where:

n represents the number of customers in the database

x_i represents the decision whether or not to mail to customer i

$E_i^{m/n} \cdot P_i^{m/n}$ represents the expected revenues of customer i given mail (m) or no mail (n)

T represents the total number of customers to be mailed

M is the general margin of the company

C is the cost of sending one mailing

Equation (4) represents the budget constraint. Rewriting equation (3) of this maximization problem indicates that we need to consider the difference between customers' contribution generated if treatment occurs and their contribution in case no treatment takes place:

$$Max\sum_{i=1}^{n}\left[(((E_i^m \cdot P_i^m)-(E_i^n \cdot P_i^n))\cdot M)\cdot x_i-(C\cdot x_i)+((E_i^n \cdot P_i^n)\cdot M)\right] \quad (5)$$

The first part of this equation represents the net contribution by sending the mailing. The last part accounts for the regular purchase behavior of customers in case no action occurs. So, the individual profit function becomes:

$$\pi_i = ((E_i^m \cdot P_i^m) - (E_i^n \cdot P_i^n)) \cdot M - C \qquad (6)$$

We emphasize the importance of estimating all the items of the profit function. So four different predictive models are required to get accurate individual expectations about customers' profits.

3.2 Model techniques

For the execution of the different predictions in equations (5) and (6) we need binary classification models to predict the individual purchase probabilities and regression models to estimate the expenditures. Next two paragraphs describe the techniques we applied.

3.2.1 Discrete choice models

Several studies support the use of logistic regression to analyse the probability of an event. It is a commonly used nonlinear technique, which has shown to perform very well in database marketing (Bult 1993, Zahavi and Levin 1997, Magidson 1988) and is used to explain discrete customer choice behavior (purchase or no-purchase). Other studies have pointed to the dominant position which logistic regression has compared to other techniques. Finally, the output of a logistic regression can easily be transformed into a probability between 0 and 1, which is a requirement for incorporation in our advanced profit function. We refer to other work for more details about logistic regression (Anderson 1982).

Next to this uncomplicated classification technique we also made use of Random Forests for means of performance comparison and validation of the results. This recent technique, proposed by Breiman (2001) has the advantages of traditional decision trees (ease of use and interpretation) and creates an ensemble of trees in order to overcome robustness problems and suboptimal performance. In this case, we made use of a random subspace method to compose the ensembles, which randomly selects a subset of variables to grow a tree. Besides, this technique automatically selects the relevant variables, which avoids overfitting between estimation and validation performance (see next paragraph). Random Forests has been discussed in other literature. We will not review it here once more.

3.2.2 Continuous prediction models

For the estimation of customers' expenses we make use of another commonly employed technique: multivariate regression. This has been discussed widely in many studies and therefore won't be handled in detail in this paper (Cohen and Cohen 2003). Also for our continuous predictions we compare the most frequently used technique from the statistical regression literature with Random Forests. The ensemble of trees is said to generate improved accuracy for models with many input variables in proportion to the number of observations. Considering the limited number of observations to model the expected expenditures in case no treatment occurs, Random Forests are an interesting benchmark. Namely, Random Forests do not require a test set because an out-of-bag set is automatically selected from the estimation set. Consequently, we only need to split the data into an estimation and a validation set so more observations can be used to build the model.

3.3 Variable selection and performances

Many studies show the relevance of using a variable selection technique and determine the selection of input variables as a critical step in response modeling (Ha et al. 2004). Overfitting to the estimation data is a well-known problem in predictive modeling (Bhattacharyya 1999) and is our main reason to apply feature selection. Certainly in case a large number of predictors is being used so the model becomes more complex (Ha et al. 2004), the performance on the estimation data can be misleading and performance may decrease dramatically on the validation data.

Backward- and forward selection procedures are probably the most well-known selection techniques. However, it has been proven that these techniques often fail to select the best performing model due to their linear selection procedure. Therefore, we make use of the global score algorithm proposed by Furnival and Wilson (1974). This technique selects the best predictors in accordance with the score chi-square statistic. The branch and bound algorithm avoids performing a complete search of the variable space, being the set of all possible variable combinations, so the computation time is reduced.

The performance of the binary models is evaluated by the area under the receivor operating characteristic curve (AUC), which is a widely accepted criterion since it evaluates the ranking for different thresholds (Ha et al. 2004). The continuous models are evaluated by the R^2, the adjusted R^2 and the RMSE.

4. EMPIRICAL STUDY AND REAL-LIFE TEST

4.1 Data

For our empirical study we got collaboration of a European retailer selling both products that are offered in grocery shops (food, beverages, cosmetics,…), as well as general merchandise products (electronics, apparel, do-it-yourself,…). In the remainder of this study, the first category of products is called the 'food' category and the second category is called the 'non food' category. In order to motivate their customers to visit one of their stores, biweekly catalogs are sent to a part of their customers. Since the use of a member card is mandatory to purchase at the store, we are ensured to have this information for all customers of the company (more than 1 million). The data delivered were very elaborate and contained customer demographics, ticket-line purchase information and information concerning past mailing actions. It was tracked at the individual customer level during more than five years: from July 1999 till March 2005 and concerned all of their Belgian outlets.

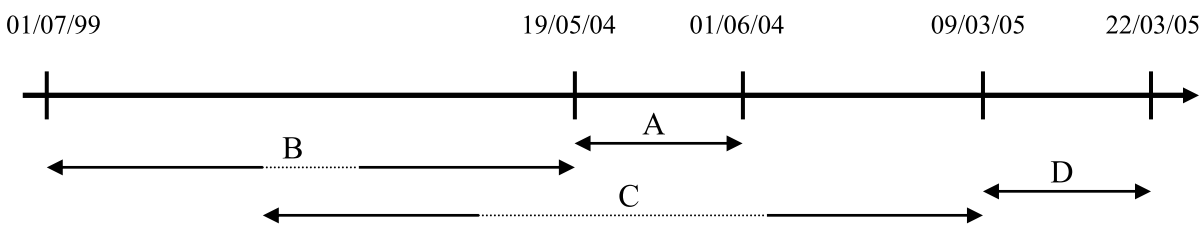
4.2 Real-life test for the usefulness of cleared profits

The encouraging results of the models (see results in Section 5), convinced our collaborating company to perform a real-life test, so we could validate our results in a subsequent mailing period. The purpose of this test was to find out whether or not the inclusion of cleared profits into the profit function leads to a reduction of the optimal number of mails and higher profits can be achieved by saving catalog costs. So, during one mailing period two sets of randomly chosen clients (two times 9898 clients) were put at our disposal for which we could manipulate the entire mailing list. One set of customers was treated by using a profit function that does not take into account the cleared profits (profit function (2)), while for the other set customers' expected purchase probability and expenses were compensated for their behavior when no leaflet would be sent (profit function (6)). For both samples, the optimal mailing depth was defined and the resulting number of customers was sent a catalog. Traditionally, the performance of models used for direct marketing purposes is evaluated by comparing the response rate of the customers being mailed (Haughton and Oulabi 1993) or by the percentage of observations that are correctly classified (Bult 1993). In our case, however, the goal is to eliminate customers from the target list who would shop even without getting a mailing. So, it is important to include the response of the customers who are not mailed. The evaluation of the real-life test is done by considering the response rate and total profit generated by all 9898 customers, in each of the manipulated sets.

4.3 Random Samples

As our profit function indicates, we need four different models in order to substitute each of the function's parameters. Typically, these models have to be estimated based on randomly drawn data from the complete customer base (Bult 1993). So, to build our models, the company mailed a random selection of customers in order to model behavior after treatment. And, to model customer behavior without treatment, the retailer left out of her mailing list, by design, a randomly chosen set of 15 540 customers. In all our models, fifty per cent of the available data were used for estimation and twenty-five per cent was used in the test and the validation sets. Figure 2 shows which data are used to build the four models and to test them in real-life.

Figure 1.2: Periods of observation for independent and dependent variables



Part A represents the company's mailing period of two weeks that was used to compute the dependent variables (buying or not buying and the expenses customers made) for the estimation of the models. Thereby, we aggregated all expenses made during these two weeks. Part B covers the time period used to compute the independent variables for all model estimations. Next, all transactional data during time period C were used to compute independent variables for our real-life test and, finally, customers' behavior in period D is used to compute the real-life results.

Table 1.2: Customer behavior with or without treatment on the estimation set.

Model		**Case**	
		With treatment	*Without treatment*
Purchase probability	Number of customers in estimation set	370 616	7 770
	Response rate	30,82%	18,40%
Expenditures	Number of customers in estimation set	114 236	1 430
	Average Spending during visit	162 €	144 €

Table 2 shows an overview about customer behavior in each of the four estimation sets that were used in each of the different models. It reports the size of the data sets and indicates the response rate and the average spending levels for the probability and the expenditure models respectively. As we expected, the response rate and the spending of customers that received a catalog exceed the one of customers without treatment. These data make it possible to decompose the effect of a promotional action, by analogy with van Heerde and Bijmolt (2005). Namely, the change in total revenue can be attributed to an increased customer spending and an enhanced number of customers that visited one of the stores (response rate).

4.4 Variables

The quantity of data delivered by the retailer is extensive. So, we could calculate an elaborate set of predictors, which are used in both the models that explain purchase propensities as well as the models that predict customers' expenses. In total 68 explanatory variables were computed. Appendices 1.A and 1.B summarize these inputs, together with a brief description of how they are calculated, based on demographic data, individual purchase history and mailing information. Rossi et al. (1996) pointed to the enormous potential of making use of household purchase histories for direct marketing models. The estimate results are also reported in this table but will be discussed in a next section of this paper. The variable set can be subdivided into different types.

The first type of variables are RFM related predictors. There exists virtually no study dealing with direct marketing strategies that does not include one or more of these widely known variables. Recency, frequency and the amount purchased are all considered to be effective predictors for future purchase behavior. Bauer (1988) made clear assumptions about the signs of the estimates of these variables. Both the frequency of purchasing and the amount of money spent will increase the likelihood of future purchasing while a higher recency might be the indication of lower purchase chances. However, this last assumption might only be true in case of fast moving consumer goods. Other studies indicate that for durables, for example, the response rate might increase with the recency (Bitran and Mondschein 1996). Therefore we included different operationalizations of these variables. First, all RFM variables are calculated using the entire purchasing data. Besides, the same variables, except for one, are calculated by considering purchases done in the food category and the non food category separately. Besides, since no agreement exists on how these predictors have to be measured (Bauer 1988) and studies stress the importance of choosing the right amount of data that needs to be incorporated (Heilman et al. 2003), we used several measures for these predictors. The spending and frequency variables are measured by using the entire purchase history, the last two years, the last year, the last six months, the last month and the last two weeks of data. Next to the typical recency variables concerning all purchases (Recency), purchases in the food category (Frecency) and purchases in the non food category (Nfrecency), we also included the average number of days between their purchases, being the interpurchase time (Ipt, F_ipt and NF_ipt). Since the time window of the estimated models had to be observed, for some customers no information was available to compute recency related variables. The dummies FRec_dum and NFRec_dum compensated for these cases. Finally, we also included some relative figures: the average spending (rSpend_freq, rFSpend_freq and rNFSpend_freq) and the amount spent relative to the length of customer's relationship (rSpend_lor, rFspend_lor and rNFSpend_lor).

Bhattacharyya (1999) indicated that the response to previous mailings might contain interesting information for future purchasing behavior. Consequently, we included the percentage of times someone went to the shop when he or she received a mailing (PercResp_Leaf). We also add the percentage of times a customer made a visit when he was not in the target list (PercResp_Noleaf). Besides, we measured how many times an individual came more than once to the store during one and the same mailing period, since we expect that customers who are very likely to come to the store without having received a mailing will come regardless of the existing mailing periods (Morethanonce). Finally, a relative measure of this last variable (Perc_morethanonce) and a dummy to indicate whether or not sufficient data were available to compute the mailing-related variables for a customer (Respdum), were added in the models.

Several studies considered the use of returned goods to express the strength of a relationship (Reinartz and Kumar 2002, Buckinx and Van den Poel 2005). So, the total value of returned goods and the total value of returned empty bottles were worked out (Retour, Amount_deposit).

Finally, we included several demographics. The availability of most of this information was dependent on the voluntariness of the customer at his or her registration. We assume that customers who provide more demographic information, have a more positive attitude towards the company and therefore have a higher purchase propensity. So we added whether or not customers did provide their fax number, phone number or e-mail (Fax_dum, Phone_dum, Email_dum). Besides, some customers do have more than one customer card, which might indicate a more intense relationship (Cardholders_dum). We also include the distance between customer's place of living and the nearest store as a predictor in the model (Distance) and we include whether customers are living in a house or an apartment (Box_dum). Further, customers who also purchase products for a company, might have different purchase intentions or quantities (VAT_dum) and in order to incorporate geodemographics we include the native language of a customer (Language_dum). Magidson (Magidson 1988), finally, points to the importance of the length of customer's relationship with the firm (Lor).

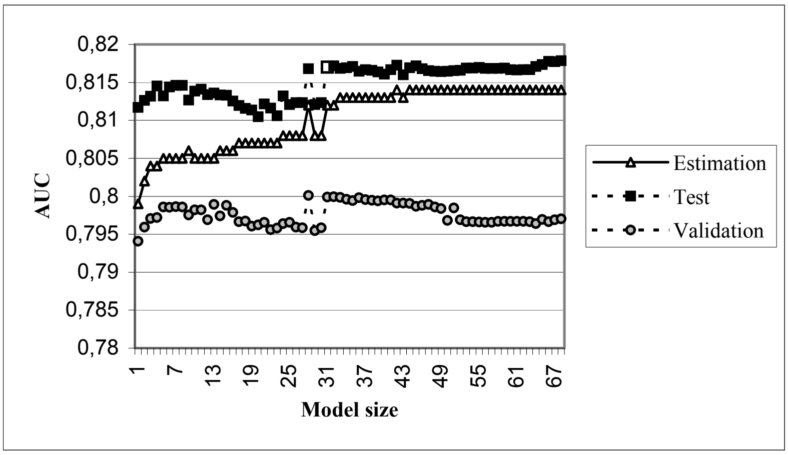
5. RESULTS

5.1 Model Performances

5.1.1 Variable selection

For the multiple linear regressions and the logistic regressions, we applied a variable selection procedure to avoid overfitting and to ensure an optimal predictive performance. Our dataset was split in three parts: an estimation set was used to estimate the models, a hold-out test set was used to make an appropriate model choice with the feature selection procedure and a hold-out validation set was kept to check for the resulting predictive performance. The optimal model size was defined by selecting the smallest model size whose performance did not significantly differ from the performance of the model with the best performance. We illustrate this selection procedure for one of the four models. Figure 3 shows the performance on the estimation, test and validation set for the prediction of purchase probability without treatment. The model with the best performance on the test set (highest AUC) was the model with size 68. However, all models with a model size larger than 30 show a performance which is not significantly different from the one with 68 variables, so model '31' was chosen as the optimal model since it is the one with the lowest number of predictors (see white coloured square within the test performances). Such a subset selection was done for all models.

Figure 1.3: Feature selection, purchase probability without treatment



The optimal model size for the prediction of purchase propensity after receiving a leaflet is twenty-two. For the prediction of expenses with treatment the most favorable size is one variable and for the determination of expected expenses without treatment the best number of variables to use is two. Appendices 1.A and 1.B give an overview of these final models together with the standardized parameter estimates of the variables. The tables also present the univariate standardized parameter estimates of all the variables. These results can be used for the interpretation of the relevance of the different predictors whereas the multivariate results show which variable set presents the best predictive performance.

5.1.2 Predictive performances

This section describes the predictive power of the different models. Table 3 is divided in four subparts and demonstrates for each model the performances of the Random Forests and either the multiple regression or the logistic regression, dependent on the type of the model. We compare the results of the *full* model – being the model that incorporates all 68 predictors – with the results of the *final* model – being the model that remains after the subset-selection procedure. Since Random Forests are not sensitive to overfitting, no feature-selection procedure was necessary and no full model information is available.

Table 1.3: Model performances

Table A: Purchase probability with leaflet

	Logistic regression		Random Forests
	Full model	Final model (v=22)	Final model
AUC	0,7368	0,7367	0,712

Table B: Expected expenses with leaflet

	Multiple linear regression		Random Forests
	Full model	Final model (v=1)	Final model
R^2	0,0046	0,2026	0,4084
Adj R^2	0,0035	0,2026	0,4077
RMSE	579,0103	297,9283	244,85

Table C: Purchase probability without leaflet

	Logistic regression		Random Forests
	Full model	Final model (v=31)	Final model
AUC	0,7970	0,7999	0,7759

Table D: Expected expenses without leaflet

	Multiple linear regression		Random Forests
	Full model	Final model (v=2)	Final model
R^2	0,2759	0,3769	0,3338
Adj R^2	0,2027	0,3752	0,2664
RMSE	202,3815	179,8794	193,8897

The evaluation shows that we got acceptable results for all of the models: all of them exhibited a significance level below 0.0001. Concerning the prediction of the purchase probabilities, the logit models did not show an overfitting problem. The performances on the full model are very comparable to the ones on the final model. Apparently, it is easier to predict whether someone will visit a store in case he did not receive a leaflet: the AUC is remarkably better than the one of the model that predicts the visiting behavior when someone did receive a catalog. In both cases, the power of the models exceeds the 0.5 benchmark of the null model and the results of the Random Forests are inferior to the ones of the logit models.

In contrast, the models that predict customers' expenses do signal overfitting difficulties. For both models, the adjusted R^2 of the full models are considerably lower than the ones of the final models. In other words, the predictive performance of our models increases by selecting the relevant predictors, which supports the application of our model selection technique. And again, considering the results, it is more convenient to predict the expenses when no leaflet was sent. Random Forests only outperform multiple regression when estimating the expenses after treatment.

As mentioned in the literature section, in previous studies it was a rare practice to model customers' expenditures as an input for the profit function. Additionally, the prediction of expenditures when no leaflet was sent was never done before. Instead, in previous studies, the expected expenses in the profit functions were mostly substituted with the average past expenses across all customers and sometimes by the average spending of a customer. Table 4 and 5 show the R^2 and the adjusted R^2 in case one would use the average past expenditures per customer to approximate expected expenditures. The performance of the averages is less good than the ones of our models (see Table 3, B and D), which supports the necessity of modeling all aspects of the profit function.

Table 1.4: Expected expenses with treatment, model fit of past individual average expenses.

	Model fit
R^2	0,1768
Adjusted R^2	0,1768
RMSE	286,272

Table 1.5: Expected expenses without treatment, model fit of past individual average expenses.

	Model fit
R^2	0,3689
Adjusted R^2	0,3681
RMSE	186,35

5.1.3 Variable Importance

The univariate standardized parameter estimates indicate which variables are most important for each of the predictions. To model purchase propensities, virtually all variable types are relevant. More specific, demographical variables have the lowest standardized estimates whereas variables related to the return of goods and recency related variables have the highest estimates for the prediction of purchase chances with treatment, and purchase propensity without treatment respectively. In contrast, more distinctions can be made between the predictors when explaining the purchase amounts. Here, variables related to customers' overall spending, spending in the food category and relative spending variables have the most notable standardized estimates. Remarkably, frequency-related variables, recency-related variables and variables concerning past mailings have lower estimates compared to the predictions of purchase probabilities. Again, demographics are among the ones with the slightest relevance. These results confirm the findings of Gupta (1988). His study showed that most of the variation in the purchase quantity is accounted for by customers' average past purchase quantity. Besides, again similar to our conclusions, interpurchase time did not show up to be an important predictor in the model.

5.2 Real-life test

5.2.1 Expected results

We could implement our proposed procedure during one of the mailing periods of the European retailer. In this real-life test, the proposed profit function (6) was used to define the optimal mailing depth and the resulting target list. As a benchmark, the traditional profit function (2) was used for another (similar) set of customers. In both cases customers were ranked based on the result of their individual profit function (step 1). The components of these functions were substituted by the outcomes of the multiple linear regressions and the logistic regressions for reasons of consistency. Figures 3 and 4 show the optimal mailing depth (step 2) - being the maximum of the accumulated outcomes of the profit functions - for our proposed case and the benchmark case respectively. Interestingly, as we expected, the probabilities of purchasing after being targeted are higher than the purchase probabilities of customers that are not mailed (Mean=0.2512, median=0.1825 versus Mean =0.1982, median=0.1322).

Figure 1.4: Optimal mailing depth, profit function (2).

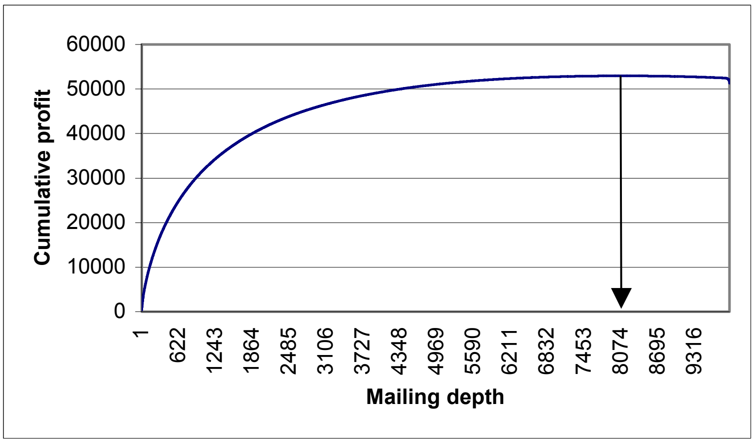
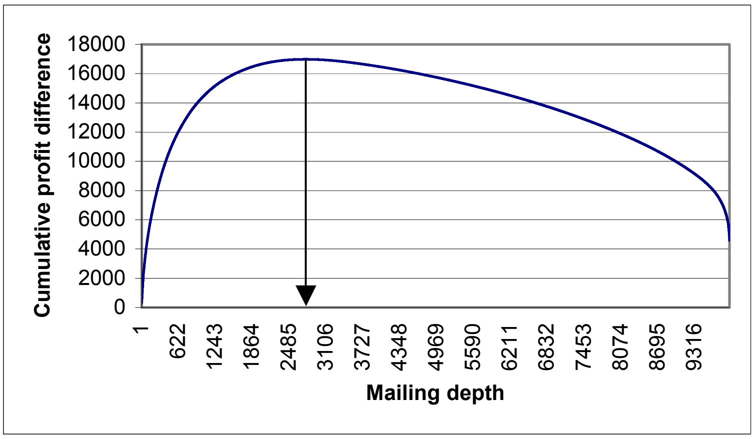


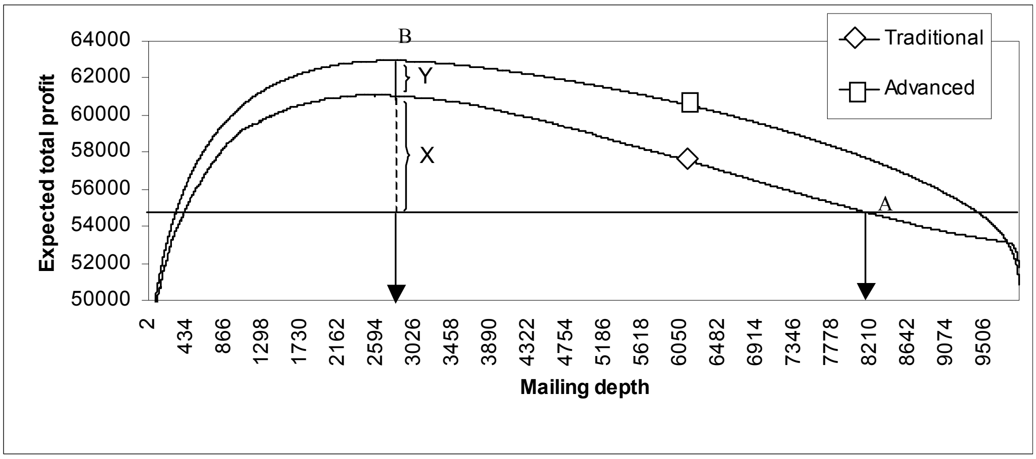
Figure 1.5: Optimal mailing depth, profit function (6).



These results show that indeed much less customers need to be mailed when we incorporate cleared profits. The optimal number of clients that had to be mailed – on a total of 9898 clients in each test case - was 2761 (Figure 5) for the advanced profit function and 8094 for the traditional approach (Figure 4).

Besides, we can define the expected profit difference between each of the customer bases. Therefore, it is not sufficient to compare the resulting profits in Figure 4 and Figure 5 since the first one reports the total profit and the second figure reports the net impact on the profit (cleared profits). Recall that in our case it is not adequate to consider the revenue of the customers being mailed. We need the expected profits of the entire customer base since our intention is to consciously leave certain customers out of the target list. So, the total profit is the profit generated by all mailed and all not mailed customers. For each customer we can calculate his/her expected individual profit contribution given that (s)he is mailed and given that (s)he is not mailed, which results, after accumulation in the total expected profits. Further, instead of reporting the expected profits for both selected mailing depths (8094 and 2761 customers), we show the expected profits for all mailing depths in each of the two cases (see Figure 6). Both curves do not start at the origin. This can be attributed to the profit that all clients are expected to generate in case none of them receives a leaflet. Appendix 1.C shows the total graph.

Figure 1.6: Attribution of profit difference to mailing depth and ranking changes.



Secondly, the figure shows that defining target lists based on the advanced profit function is beneficial at each mailing depth. Besides, when considering that the test involves less than 1 per cent of the total customer database, the expected profit difference between the advanced and the traditional method is substantial: 62939 euro (point B) versus 54840 euro (point A). Moreover, the curves show that the optimal mailing depth of the advanced method indeed guarantees the optimal profit level. Whereas this is not the case for the traditional procedure.

Besides, it is clear that the profit difference between the two approaches can be attributed to a) the savings made by reducing the mailing cost, and b) the alternative ranking of the customers in the segmentation list. This is shown in Figure 6 where the total profit difference between point A and B can be split in part X (attribution a) and Y (attribution b) respectively.

5.2.2 After implementation

To check whether these expectations hold in a real-life environment, the optimal number of mailings, according to each method, were distributed to the respective customer sets. Table 6 shows the results of both systems.

The results of our real-life test confirm the expectations: the figures prove that our advanced method indeed generates more profit than the traditional method. The customers in the set of the traditional procedure, generate more revenue in total, but, since their total mailing cost is significantly higher, the remaining profit, after considering margin and mailing costs, is 2151 euro lower. An extrapolation to the total customer base yields more than 200 000 euro per mailing, being an increase of the total company profitability of five per cent.

Table 1.6: Results of real-life test, traditional and advanced profit function methods.

	Traditional	Advanced
Number of customers	9898	9898
Number of mailings sent	8094	2761
Total Revenue by all clients	332 997 euro	317 117 euro
Mailing costs	6 880 euro	2 347 euro
Profit[2]	43 070 euro	45 221 euro
Response rate mailed (%)	2043 (25,24%)	1042 (37,74%)
Response rate not mailed (%)	94 (5,21%)	1017 (14,25 %)

6. DISCUSSION AND EXTENSIONS

The success of a direct marketing campaign depends on how a company is able to define customers' value and to what extent it can determine the optimal size of its target list. Both these decisions are considered to be the most important steps for direct marketing management and are driven by the profit function applied.

We propose a new direct mailing method, which makes use of a more advanced profit function that values customers based on the net effect of companies' targeting action. This seems appropriate since in retail settings customers are able to make purchases even if they do not receive a mailing. In addition, we are the first to use individual predictive models to substitute each item of this elaborate function. The degree to which expected purchase

[2] Considering a profit margin of 15 per cent.

probabilities and expected expenses correspond to real behavior has a direct impact on the significance of the profit function and therefore on the success of the selection method. By accounting for customers' cleared profits and providing a more reliable approximation of probabilities and expenses, we present an improved mailing method that selects customers who need a stimulus to make purchases and disregards customers who will buy anyhow.

We used logistic regression, multivariate regression and Random Forests to estimate the purchase probability and the expenses in case a customer is treated and in case a customer is not being treated. Sixty-eight predictors of different types were used as explanatory variables. All the models show valid prediction performances. The individual prediction of expected expenses has a better fit with customers' real expenses compared to the use of past average expenses. Besides, the amount expended with treatment differs from the expenses without treatment. This demonstrates the contribution of applying modeling techniques for all items of the profit function. A feature-selection procedure, based on the algorithm of Furnival and Wilson (1974), choose the optimal number of inputs for each of the models. The results show that mainly the predictions of future expenses experience overfitting problems, for which variable selection demonstrates its usefulness. Random Forests, however, could only outperform the other techniques for the prediction of future expenditures after a customer did receive a mailing which highlights the strength of logistic regressions for binary classification problems. For the prediction of purchase propensities almost every variable type is of relevance. In contrast, the modeling of customers' expenses is especially explained by spending-related variables.

Most interestingly, in collaboration with a European retailer, we implemented the method presented in this paper in a real-life environment. The results show that companies, whose customers have the possibility to make purchases without being treated, are sending too much mailings when applying traditional profit functions for customer evaluation. The use of our advanced profit function causes a substantial reduction in the number of mailings that need to be sent, while the total profit significantly increases. This can be attributed to the elimination of customers from the mailing list, who make purchases regardless of whether they receive a leaflet. Moreover, our results show that the profit difference can be credited to both the reduction of the number of mailings and to the changing order of the customers in the segmentation list. Besides, the expected profit curves across all mailing depths indicate that this profit difference exists at each mailing size. Consequently, even if the optimal number of customers cannot be targeted, for example, due to budgetary constraints, or if the company wants to mail more customers than the optimal mailing depth suggests, it is more profitable to use the advanced profit function to compose the customer ranking. To conclude, these findings are particularly interesting for marketing management since with our method, higher profits can be generated with lower marketing expenditures. Besides, applying the advanced profit function causes substantial changes into the profile of the customers being targeted. Whereas traditional approaches typically target the ‘best’ customers, our method focuses on customers who need to be stimulated the most. That way, less ‘promising’ clients are also in the target list which means they are reactivated and shrinkage of the active customer base over time might be avoided (Elsner et al. 2004).

This study is not without limitations. In our case, customers are able to shop regardless of the treatment they received, which is common practice for traditional store retailers. The inclusion of cleared profits in the profit function gains importance to the extent that customers who are not targeted generate sales. In a mail-order setting where catalogs are distributed with a constantly changing catalog content, for example, it is rather impossible to make purchases if no mailing is received. Further research needs to investigate the contributions of our advanced profit function in other settings, which use direct marketing to stimulate purchase behavior.

The power of the models has a direct influence on the predictive performance of the profit function and is therefore crucial for the entire mailing strategy. So, the use of modeling techniques with a predictive ability that outperforms the ones presented in this study will result in increased accuracy, better customer ranking and higher profits. Besides, the inclusion of other relevant explanatory variables might increase the performance of the models as well.

In our method, customer value is evaluated based on their contribution during a single mailing. Some studies, however, suggest that customers’ profits need to be maximized over a longer period, including more than one mailing (Piersma and Jonker 2004, Bitran and Mondschein 1996). So, the inclusion of the advanced profit function into such methodologies is worth the investigation.

REFERENCES

Anderson, J.A. 1982. *Logistic discrimination*. In: Krishnaiah, P.R., Kanal L.N. (Eds.), Handbook of Statistics, **2** 169-191, Amsterdam, The Netherlands.

Bauer,s C.L. 1988. A Direct Mail Customer Purchase Model. *Journal of Direct Marketing.* **2**(3) 16-24.

Bhattacharyya, S. 1999. Direct Marketing Performance Modeling Using Genetic Algorithms. *INFORMS Journal on Computing*. **11**(3) 248-257.

Bitran, G.R., S.V. Mondschein. 1996. Mailing Decisions in the Catalog Sales Industry. *Management Science*. **42**(9) 1364-1381.

Breiman, L. 2001. Random Forests. *Machine Learning*. **45**(1) 5-32.

Buckinx, W., D. Van den Poel. 2005. Customer Base Analysis: Partial Defection of Behaviorally-Loyal Clients in a Non-Contractual FMCG Retail Setting. *European Journal of Operational Research*. **164(**1) 252-268.

Bult, J.R. 1993. Semiparametric Versus Parametric Classification Models: An Application to Direct Marketing. *Journal of Marketing Research*. **30**(August) 380-390.

Bult, J. R., T. Wansbeek. 1995. Optimal Selection For Direct Mail. *Marketing Science*. **14**(4) 378-394.

Campbell, D., R. Erdahl, D. Johnson, E. Bibelnieks, M. Haydock, M. Bullock, H. Crowder. 2001. Optimizing Customer Mail streams at Fingerhut. *Interfaces.* **31**(1) 77-90.

Cohen, J., P. Cohen (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd ed.)*. Mahwah, NJ: Erlbaum.

Desarbo, W.S., V. Ramaswamy. 1994. CRISP: Customer Response Based Iterative Segmentation Procedures for Response Modeling in Direct Marketing. *Journal of Direct Marketing*. **8**(3) 7-20.

Elsner, R., M. Krafft, A. Huchzermeier. 2004. Optimizing Rhenania's Direct Marketing Business Through Dynamic Multilevel Modeling (DMLM) in a Multicatalog-Brand Environment. *Marketing Science*. **23**(2) 192-206.

Furnival, G.M., R.W. Wilson. 1974. Regressions by Leaps and Bounds. *Technometrics*. **16** 499-511.

Gönül, F.F., M. Shi. 1998. Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models. *Management Science.* **44**(9) 1249-1262.

Gönül, F.F., B.-D. Kim, M. Shi. 2000. Mailing smarter to catalog customers. *Journal of Interactive Marketing*. **14**(2) 2-16.

Gupta, S. 1988. Impact of Sales Promotions on When, What, and How Much to Buy. *Journal of Marketing Research*, **25**(November) 342-355.

Ha, K., S. Cho, D. Maclachlan. 2004. Response Models Based on Bagging Neural Networks. *Journal of Interactive Marketing*. **19**(1) 17-30.

Hansotia, B., B. Rukstales. 2002. Incremental Value Modeling. *Journal of Interactive Marketing*. **16**(3) 35-46.

Haughton, D., S. Oulabi. 1993. Direct Marketing Modeling with CART and CHAID. *Journal of Direct Marketing*. **7**(3) 16-26.

Heilman, C.M., F. Kaefer, S.D. Ramenofsky. 2003. Determining the Appropriate Amount of Data for Classifying Consumers for Direct Marketing Purposes. *Journal of Interactive Marketing*. **17**(3) 5-28.

Levin, N., J. Zahavi. 2001. The Economics of Selection of Mail orders. *Journal of Interactive Marketing.* **15**(3) 53-71.

Magidson, J. 1988. Improved Statistical Techniques for Response Modeling. *Journal of Direct Marketing*. **2**(4) 6-18.

Malthouse, E.C. 1999. Ridge Regression and Direct Marketing Scoring Models. *Journal of Interactive Marketing*. **13**(4) 10-23.

Muus, L., H. Van der Scheer, T. Wansbeek. 1996/2002. A decision Theoretic Framework for Profit Maximization in Direct Marketing. *Econometric Models in Marketing.* **16** 119-140.

Reinartz, W.J., V. Kumar. 2002. The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration. *Journal of Marketing*. **67**(January) 77-99.

Piersma, N., J.-J. Jonker. 2004. Determining the Optimal Direct Mailing Frequency. *European Journal of Operational Research.* **158** 173-182.

Prinzie, A., D. Van den Poel. 2005. Constrained optimization of data mining problems to improve model performance: A direct-marketing application. *Expert System with Applications*. **29** (3).

Rossi, P.E., R.E. McCulloch, G.M. Allenby. 1996. The Value of Purchase History Data in Target Marketing. *Marketing Science*. 15(4) 321-340.

The Direct Marketing association 2004. *The DMA quarterly Business Review*. 4th quarter.

Van Heerde, H.J., T. Bijmolt. 2005. Decomposing the Promotional Revenue Bump for Loyalty Program Members versus Non-Members. *Journal of Marketing Research*, forthcoming.

Zahavi, J., N. Levin. 1997. Applying Neural Computing to Target Marketing. *Journal of Direct Marketing*. **11**(4) 76- 93.

APPENDICES

Appendix 1.A: Description and standardized parameter estimates for multivariate and univariate models of purchase probabilities.

Variable	Description	Models			
		Purchase with treatment		Purchase without treatment	
		Multivariate	Univariate	Multivariate	Univariate
Frequency	Number of purchases in total history.	0,046 ***	0,484 ***		0,509 ***
Frequency_2Y	Number of purchases during last two years.	-0,258 ***	0,560 ***	-0,134	0,584 ***
Frequency_1Y	Number of purchases during last year.	0,211 ***	0,580 ***	0,255 ***	0,608 ***
Frequency_6M	Number of purchases during last six months.	0,040 ***	0,550 ***		0,578 ***
Frequency_1M	Number of purchases during last month.	0,029 ***	0,345 ***		0,415 ***
Frequency_2W	Number of purchases during last two weeks.		0,223 ***		0,261 ***
FFrequency	Number of purchases in total history in food category.	-0,028 ***	0,440 ***	0,069	0,449 ***
FFrequency_2Y	Number of purchases during last two years in food category.		0,514 ***		0,516 ***
FFrequency_1Y	Number of purchases during last year in food category.		0,536 ***	-0,126 **	0,536 ***
FFrequency_6M	Number of purchases during last six months in food category.	0,052 ***	0,512 ***		0,517 ***
FFrequency_1M	Number of purchases during last month in food category.		0,316 ***		0,371 ***
FFrequency_2W	Number of purchases during last two weeks in food category.		0,198 ***	0,032	0,231 ***
NFFrequency	Number of purchases in total history in non food category.		0,434 ***		0,493 ***
NFFrequency_2Y	Number of purchases during last two years in non food category.	0,081 ***	0,484 ***	0,031	0,547 ***
NFFrequency_1Y	Number of purchases during last year in non food category.		0,487 ***		0,555 ***
NFFrequency_6M	Number of purchases during last six months in non food category.		0,452 ***		0,523 ***
NFFrequency_1M	Number of purchases during last month in non food category.		0,269 ***	0,058 **	0,352 ***
NFFrequency_2W	Number of purchases during last two weeks in non food category.		0,176 ***	-0,042 *	0,215 ***
Spending	Spending in total history.	-0,049 ***	0,594 ***		0,431 ***
Spending_2Y	Spending in last 2 years.		0,706 ***	10,936 **	0,512 ***
Spending_1Y	Spending in last year.		0,667 ***		0,510 ***
Spending_6M	Spending in last 6 months.	0,098 ***	0,603 ***		0,500 ***
Spending_1M	Spending in last month.		0,292 ***		0,316 ***
Spending_2W	Spending in last 2 weeks.	0,023 ***	0,195 ***		0,215 ***
FSpending	Spending in total history in food category.		0,630 ***		0,357 ***
FSpending_2Y	Spending in last 2 years in food category.		0,781 ***	-7,129 **	0,440 ***
FSpending_1Y	Spending in last year in food category.		0,767 ***	0,185 ***	0,434 ***
FSpending_6M	Spending in last 6 months in food category.		0,740 ***		0,427 ***
FSpending_1M	Spending in last month in food category.		0,417 ***	0,020	0,308 ***
FSpending_2W	Spending in last 2 weeks in food category.		0,268 ***		0,218 ***
NFSpending	Spending in total history in non food category.		0,304 ***		0,376 ***
NFSpending_2Y	Spending in last 2 years in non food category.		0,328 ***	-6,076 **	0,424 ***
NFSpending_1Y	Spending in last year in non food category.		0,306 ***		0,423 ***
NFSpending_6M	Spending in last 6 months in non food category.		0,264 ***	0,063 **	0,420 ***
NFSpending_1M	Spending in last month in non food category.		0,132 ***		0,212 ***
NFSpending_2W	Spending in last 2 weeks in non food category.		0,091 ***		0,138 ***
Recency	Number of days since last purchase.	-0,078 ***	-0,424 ***	-0,434 ***	-1,227 ***
FRecency	Number of days since last purchase in food category.		-0,228 ***		-0,668 ***
NFRecency	Number of days since last purchase in non food category.		-0,302 ***	0,091	-0,954 ***
Ipt	Average number of days between store visits.		-0,600 ***		-1,811 ***
F_ipt	Average number of days between store visits in food category.		-0,216 ***		-0,562 ***
NF_ipt	Average number of days between store visits in non food category.		-0,501 ***		-1,296 ***
FRecdum	Dummy to indicate absence of data to compute Frecency		-0,059 ***		-0,195 ***
NFRecdum	Dummy to indicate absence of data to compute NFrecency	0,008 ***	-0,015 ***		-0,066 **
rSpend_freq	Average Spending in a visit.		-0,081 ***	0,036	-0,157 ***
rFSpend_freq	Average Spending in a visit in the food category.		0,029 ***	-0,104 ***	0,019 *
rNFSpend_freq	Average Spending in a visit in the non food category.		-0,161 ***		-0,253 ***
rSpend_lor	Relative Spending to the length of customer's relationship		0,548 ***	-2,568 *	0,424 ***
rFSpend_lor	Relative Spending in the food category to the length of customer's relationship.		0,579 ***	1,723 *	0,359 ***
rNFSpend_lor	Relative Spending in the non food category to the length of customer's relationship.		0,298 ***	1,320 *	0,358 ***
PercResp_Leaf	Percentage of times a purchase is made in case a promotion leaflet was received.	0,285 ***	0,497 ***	0,290 ***	0,604 ***
PercResp_Noleaf	Percentage of times a purchase is made in case no promotion leaflet was received.	0,015 ***	-0,319 ***	-0,061 *	-0,418 ***
Morethanonce	Number of times that a customer visits more than once in one and the same promotion period.	0,129 ***	0,464 ***	0,043	0,526 ***
Perc_morethanonce	MoreThanOnce divided by the number of times a customer bought at least once in a promotion period.	-0,050 ***	0,204 ***	-0,028	0,291 ***
Respdum	Dummy to control for missing data concerning mailing information	-0,038 ***	0,079 ***	0,030	0,508 ***
Retour	Total value of returned goods.		0,967 ***		0,202 ***
Amount_deposit	Total value of empty botles returned.		0,676 ***		0,145 ***
Language_dum	Customer's language (1=Dutch, 0 = French)	-0,015 ***	-0,015 ***		0,096 ***
Vat_dum	Customer has VAT number or not (1/0)	-0,009 ***	-0,027 ***		-0,065 ***
Fax_dum	Fax number in database (1= yes, 0= no)		0,006 ***	-0,030	-0,042 **
Phone_dum	Phone number in database (1= yes, 0= no)		-0,001		0,042 **
Remark_dum	Remark in database (1= yes, 0= no)		-0,011 ***	-0,022	-0,022
Email_dum	E-mail address in database (1= yes, 0= no)		-0,004 *		-0,023
Box_dum	Living in flat (1= yes, 0= no)		-0,005 **	-0,026	0,003
Cardholders_dum	2 cardholders (1= yes, 0= no)		0,039 ***		0,205 ***
Relation_dum	Relation indication in database (1= yes, 0= no)		-0,012 ***	0,058 ***	0,076 ***
Distance	Distance to the store	-0,031 ***	-0,131 ***	-0,041 **	-0,190 ***
Lor	Length of customer's relationship.	0,007 ***	0,081 ***		0,112 ***

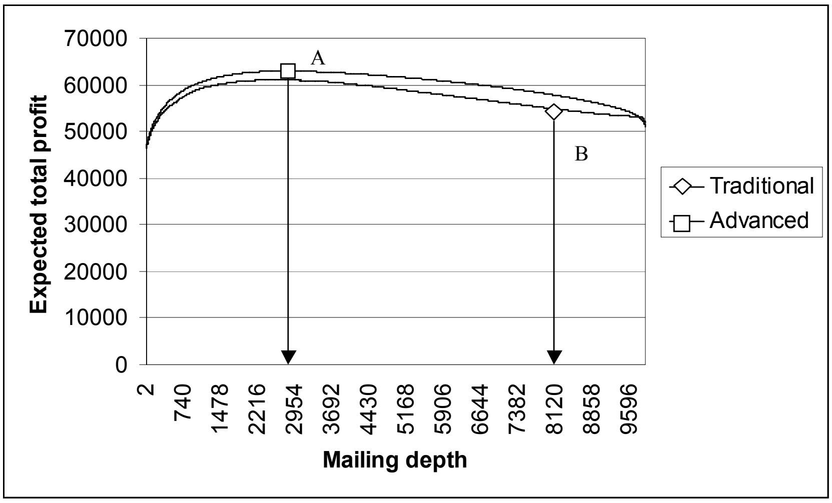
* p <.10
** p <.05
* **p <.01

Appendix 1.B: Description and standardized parameter estimates for multivariate and univariate models of expenses.

Variable	Description	Models			
		Expenses with treatment		Expenses without treatment	
		Multivariate	Univariate	Multivariate	Univariate
Frequency	Number of purchases in total history.		0,099 ***		-0,018
Frequency_2Y	Number of purchases during last two years.		0,120 ***		-0,001
Frequency_1Y	Number of purchases during last year.		0,120 ***		0,020
Frequency_6M	Number of purchases during last six months.		0,115 ***		0,031
Frequency_1M	Number of purchases during last month.		0,092 ***		0,051 *
Frequency_2W	Number of purchases during last two weeks.		0,066 ***		-0,005
FFrequency	Number of purchases in total history in food category.		0,111 ***		-0,007
FFrequency_2Y	Number of purchases during last two years in food category.		0,128 ***		0,016
FFrequency_1Y	Number of purchases during last year in food category.		0,127 ***		0,042
FFrequency_6M	Number of purchases during last six months in food category.		0,124 ***		0,055 **
FFrequency_1M	Number of purchases during last month in food category.		0,104 ***		0,066 **
FFrequency_2W	Number of purchases during last two weeks in food category.		0,075 ***		0,011
NFFrequency	Number of purchases in total history in non food category.		0,045 ***		-0,004
NFFrequency_2Y	Number of purchases during last two years in non food category.		0,059 ***		0,009
NFFrequency_1Y	Number of purchases during last year in non food category.		0,065 ***		0,032
NFFrequency_6M	Number of purchases during last six months in non food category.		0,065 ***		0,038
NFFrequency_1M	Number of purchases during last month in non food category.		0,048 ***		0,047 *
NFFrequency_2W	Number of purchases during last two weeks in non food category.		0,033 ***		-0,019
Spending	Spending in total history.		0,679 ***		0,186 ***
Spending_2Y	Spending in last 2 years.		0,669 ***		0,210 ***
Spending_1Y	Spending in last year.		0,608 ***	0,184 ***	0,257 ***
Spending_6M	Spending in last 6 months.		0,631 ***		0,231 ***
Spending_1M	Spending in last month.		0,428 ***		0,223 ***
Spending_2W	Spending in last 2 weeks.		0,323 ***		0,124 ***
FSpending	Spending in total history in food category.	0,703 ***	0,703 ***		0,163 ***
FSpending_2Y	Spending in last 2 years in food category.		0,689 ***		0,187 ***
FSpending_1Y	Spending in last year in food category.		0,625 ***		0,246 ***
FSpending_6M	Spending in last 6 months in food category.		0,658 ***		0,232 ***
FSpending_1M	Spending in last month in food category.		0,462 ***		0,210 ***
FSpending_2W	Spending in last 2 weeks in food category.		0,357 ***		0,170 ***
NFSpending	Spending in total history in non food category.		0,158 ***		0,143 ***
NFSpending_2Y	Spending in last 2 years in non food category.		0,183 ***		0,140 ***
NFSpending_1Y	Spending in last year in non food category.		0,184 ***		0,153 ***
NFSpending_6M	Spending in last 6 months in non food category.		0,169 ***		0,132 ***
NFSpending_1M	Spending in last month in non food category.		0,120 ***		0,136 ***
NFSpending_2W	Spending in last 2 weeks in non food category.		0,100 ***		0,019
Recency	Number of days since last purchase.		-0,006 *		0,070 ***
FRecency	Number of days since last purchase in food category.		-0,016 ***		-0,007
NFRecency	Number of days since last purchase in non food category.		-0,004		0,029
Ipt	Average number of days between store visits.		-0,012 ***		0,024
F_ipt	Average number of days between store visits in food category.		-0,017 ***		-0,023
NF_ipt	Average number of days between store visits in non food category.		-0,005		0,010
FRecdum	Dummy to indicate absence of data to compute Frecency		-0,005 *		-0,016
NFRecdum	Dummy to indicate absence of data to compute NFrecency		-0,001		0,012
rSpend_freq	Average Spending in a visit.		0,336 ***	0,229 ***	0,288 ***
rFSpend_freq	Average Spending in a visit in the food category.		0,356 ***		0,281 ***
rNFSpend_freq	Average Spending in a visit in the non food category.		0,058 ***		0,116 ***
rSpend_lor	Relative Spending to the length of customer's relationship		0,673 ***		0,198 ***
rFSpend_lor	Relative Spending in the food category to the length of customer's relationship.		0,696 ***		0,169 ***
rNFSpend_lor	Relative Spending in the non food category to the length of customer's relationship.		0,155 ***		0,155 ***
PercResp_Leaf	Percentage of times a purchase is made in case a promotion leaflet was received.		0,037 ***		-0,029
PercResp_Noleaf	Percentage of times a purchase is made in case no promotion leaflet was received.		-0,007 **		-0,003
Morethanonce	Number of times that a customer visits more than once in one and the same promotion period.		0,073 ***		0,012
Perc_morethanonce	MoreThanOnce divided by the number of times a customer bought at least once in a promotion period.		0,083 ***		0,093 ***
Respdum	Dummy to control for missing data concerning mailing information		-0,001		-0,032
Retour	Total value of returned goods.		0,287 ***		0,072 ***
Amount_deposit	Total value of empty botles returned.		0,259 ***		0,071 ***
Language_dum	Customer's language (1=Dutch, 0 = French)		0,019 ***		0,059 **
Vat_dum	Customer has VAT number or not (1/0)		0,059 ***		0,087 ***
Fax_dum	Fax number in database (1= yes, 0= no)		0,050 ***		0,019
Phone_dum	Phone number in database (1= yes, 0= no)		0,018 ***		0,048 *
Remark_dum	Remark in database (1= yes, 0= no)		-0,004		-0,004
Email_dum	E-mail address in database (1= yes, 0= no)		0,046 ***		0,061 **
Box_dum	Living in flat (1= yes, 0= no)		-0,017 ***		-0,027
Cardholders_dum	2 cardholders (1= yes, 0= no)		-0,001		0,017
Relation_dum	Relation indication in database (1= yes, 0= no)		-0,034 ***		-0,028
Distance	Distance to the store		0,028 ***		0,101 ***
Lor	Length of customer's relationship.		-0,002		-0,012

* p <.10
** p <.05
* **p <.01

Appendix 1.C: Expected profit for all mailing depths, traditional versus advanced profit function



CHAPTER II

CUSTOMER-ADAPTED COUPON TARGETING USING FEATURE SELECTION[3]

---

[3] This chapter is based on the following reference: Wouter Buckinx, Elke Moons, Dirk Van den Poel, Geert Wets, 2004. Customer-Adapted Coupon Targeting Using Feature Selection, *Expert Systems with Applications*, Vol 26(4), 509-518.

CHAPTER II:

CUSTOMER-ADAPTED COUPON TARGETING USING FEATURE SELECTION

1. INTRODUCTION

Since the late nineteenth century companies bring into play coupons in their marketing strategy. Today, this type of promotion still is the most important promotion medium (Bawa, Srinivasan & Srivastava, 1997). For several popular product categories more than half of the sales volume is generated when products are offered with a price reduction (Blattberg and Neslin, 1990). This results from the fact that most products are in an advanced stage of their life cycle and product differentiation becomes a hard job (Papatla & Krishhnamurthi, 1996). As a result, the distribution of coupons is more than ever an important topic to be considered by marketing managers.

2. LITERATURE REVIEW

In previous literature a lot of studies can be found that support the use of coupons as a promotional tool. First, thanks to coupons consumers tend to increase their purchase volume of the specific product and even accelerate the purchase timing of the goods (Blattberg, Eppen, & Lieberman, 1981). Ailawadi and Neslin (1998) confirm this finding and attribute the increase to high inventories that entail fewer stockouts and an increase in usage rate. They argue, however, that it does not hold for all product categories.

Some studies found a positive effect of coupons on repeat purchases. Taylor (2001) concluded that customers who redeemed a coupon were about 7 times more in favor of making purchases in the period after the promotion. Without making a distinction between users and non-users, Lattin and Bucklin (1989) found a significant increase in customers' purchase probability after a promotional purchase while Ailawadi, Lehmann and Neslin (2001) only found a limited impact on customer retention for the product. Papatla and Krishnamurthi (1996) add: 'Positive effects could be a consequence of preference reinforcement for a purchased brand'. Several studies support that promotions result in brand-switching behavior (Bell, Jeongwen & Padmanabahn, 1999). Some, however, stress the short-term character of this behavior (Bonnici, Campbell, Fredenberger and Hunnicutt, 1996).

Though the supply of coupons is intense, their redemption rate typically is relatively low. Moreover, current coupon strategies are in nature unprofitable (Bawa, Srinivasan & Srivastava, 1997). Therefore, modelling the release of coupons in order to optimize the promotional strategy carries a lot of potential. First, making predictions concerning the proneness of customers for coupons is necessary in order to define target segments and for being able to make strategy evaluations (Bawa, Srinivasan & Srivastava, 1997). This makes it possible to limit marketing costs and define the exact budgets that need to be allocated. Bucklin and Gupta (1992) indicate that segments of customers exist that differ concerning their response to promotions. Being able to make predictions concerning the correct class each customer will belong to after the distribution of coupons would make it possible to accomplish the abovementioned advantages. Moreau, Krishna and Harlam (2001) support this by stating: "The effectiveness of any promotional strategy depends on how accurately channel members predict consumers' perceptions of their promotional activity".

However, little efforts have been made to investigate the ability to make predictions concerning coupon redemption (Bawa, Srinivasan & Srivastava, 1997). Most past research focuses on coupon proneness and the effects of coupons on product purchases (Papatla & Krishnamurthi, 1996). They hardly ever examine predictability nor question which information is valuable for the models in this respect. This may be due to the fact that no possibility existed to target coupons.

Besides making predictions concerning coupons in general, it might be recommended to build models for different kind of coupons since customers can be interested in specific types of promotions. Building models without taking this into account possibly leads to worse predictions. As mentioned above, Papatla and Krishnamurthi (1996) signal the existence of different segments in the population, some being more sensitive to promotions than others. Moreover, they confirmed that the difference could be in more than one dimension. This means that the price sensitivity could be determined by among other things coupon's brand or product category. Taylor (2001) mentioned that preferences for specific categories or brands determine the level of sensitivity for promotions within their respective categories. Bawa and Shoemaker (1987) add: "Studies have shown that regular purchasers of a brand have a higher likelihood of coupon redemption than purchasers with low prior purchase probability".

Blattberg and Neslin (1990) note that redemption rates vary widely across product categories and suggest managers to examine coupon use at a more detailed level in order to build strategies. Finally, Bawa, Srinivasan and Srivastava (1997) concluded that coupon attractiveness increases for consumers' usual preferences.

So one may hypothesize that segments of people can be detected that differ concerning their redemption behavior for specific types of coupons. As a result building separate models for different kind of promotions seems to be recommended.

3. METHODOLOGY

3.1 Retailers' versus manufacturers' coupons

As claimed above, promotion strategies might be improved by making predictions concerning coupon redemption and by building specific models for different types of coupons. An interesting level of distinction one can make in a company's product taxonomy is the one that separates national brands from store brands. Ailawadi, Neslin and Gedenk (2001) show that usage of store brands and national brands' promotions attract a different kind of people with respect to psychographics. Store brand users are driven by economic

benefits whereas redemption of national brand promotions is driven by hedonic benefits (shopping enjoyment, innovativeness, variety seeking, impulsiveness). Consequently, national brand promotions and store brand promotions satisfy different needs and different groups of customers. Being able to classify each customer into the right segment is very important to marketing managers. Manufacturers and retailers typically are involved in a battle where both parties are devising money-consuming promotional actions in order to convince as many customers as possible. However, the aforementioned availability of media which allow targeting of individual consumers at reasonable cost, can lead to a reduction of the conflict between manufacturers and retailers (Ailawadi, Lehmann & Neslin, 2001).
As a result, we will build two models: one making predictions concerning the redemption behavior of coupons for store brand products and a second model making predictions concerning the redemption of coupons for national brands.

3.2 Feature selection method: Relief-F

Feature selection strategies are often implied to explore the effect of irrelevant attributes on the performance of classifier systems. A feature selection method ranks all the attributes (features) in descending order of relevance. In this analysis, the Relief-F feature selection method is opted for since it can easily be combined with the C4.5 induction algorithm.
Feature selection strategies can be regarded as one way of coping with the correlation between attributes. This is relevant because the structure of trees is sensitive to the problem of multicollinearity. Multicollinearity means that some variables are simply redundant (given the presence of other variables). Redundant variables do not affect the impact of the remaining variables in the tree model, but it would simply be better if they were not used for splitting. Therefore, a good feature selection method for this analysis would search for a subset of relevant features that are highly correlated with the class variable that the tree-induction algorithm is trying to predict. In addition, the variables also have to be as uncorrelated with each other as possible.
Relief (Kira and Rendall, 1992), the predecessor of Relief-F, is a distance-based feature weighting algorithm. It orders attributes according to their importance. To each attribute it assigns the initial value of zero that will be adapted with each run through the instances of the dataset. The features with the highest values are considered to be the most relevant, while those with values close to zero or with negative values are judged irrelevant. Thus Relief

imposes a ranking on features by assigning a weight to every variable. The weight for a particular feature reflects its relevance in distinguishing the classes. In determining the weights, the central concepts are *near-hit* and *near-miss*. A *near-hit* of instance *i* is defined as the instance that is closest to *i* (based on Euclidean distance) and which is of the same class (concerning the output variable), while a *near-miss* of *i* is defined as the instance that is closest to *i* (based on Euclidean distance) and which is of a different class (concerning the output variable). The algorithm attempts to approximate the following difference of probabilities for the weight of a feature X:

W_X = P(different value of X | nearest instance of different class)
- P(different value of X | nearest instance of same class)

So, Relief works by randomly sampling an instance and locating its nearest neighbor from the same and opposite class. The nearest neighbor is defined in terms of the Euclidean distance, so in an n-dimensional space, the following distance measure will be used:

$d(\boldsymbol{x}, \boldsymbol{y}) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are two n-dimensional vectors.

By removing the context sensitivity provided by the "nearest instance" condition, attributes are treated as mutually independent, and the previous equation becomes:

Relief_X = P(different value of X | different class)
- P(different value of X | same class).

Relief-F (Kononenko, 1994) is an extension of Relief that can handle multiple classes and noise caused by missing values, outliers, etc. To increase the reliability of Relief's weight estimation, Relief-F finds the *k* nearest hits and misses for a given instance, where *k* is a parameter that can be specified by the user. For multiple class problems, Relief-F searches for nearest misses from each different class (with respect to the given instance) and averages their contribution. The average is weighted by the prior probability of each class.

3.3 C4.5

Decision tree induction can be best understood as being similar to parameter estimation methods in econometric models. The goal of tree induction is to find the set of Boolean rules that best represents the empirical data. In this study, the trees were induced using the C4.5 method (Quinlan, 1993), which works as follows. Let there be given a set of choice observations i taken from activity-travel diary data. Consider the n different attributes $X_{i1}, X_{i2}, ..., X_{in}$ and the choice variable $Y_i \in \{1,2,...,p\}$ for i = 1, … I. In general, a tree consists of different layers of nodes. It starts from the root node in the first layer or first parent node. This parent node will split into daughter nodes on the second layer. In turn, each of these daughter nodes can become a new parent node in the next split, and this process may continue with further splits. A leaf node is a node, which has no offspring nodes. Nodes in deeper layers become increasingly more homogeneous. An internal node is split by considering all allowable splits for all variables and the best split is the one with the most homogeneous daughter nodes. The C4.5 algorithm recursively splits the sample space on X into increasingly homogeneous partitions in terms of Y, until the leaf nodes contain only cases from a single class. Increase in homogeneity achieved by a candidate split is measured in terms of an information gain ratio. As stated in Quinlan (1993), the information theory on which the gain ratio criterion is based can be explained in the following statement:

Definition 1: Information of a message

The information conveyed by a message depends on its probability and can be measured in bits as minus the logarithm to base 2 of that probability.

For example, if there are four equally probable messages, the information conveyed by any of them is - $\log_2$ (1/4) = 2 bits.

Definition 2: Information of a message that a random case belongs to a certain class

$$-\log_2\left(\frac{freq(C_i,T)}{|T|}\right) \; bits$$

with T a training set of cases, C_i a class i and freq(C_i, T) the number of cases in T that belongs to class C_i.

Based on these definitions, the average amount of information needed to identify the class of a case in a training set (also called entropy) can be deduced as follows:

Definition 3: Entropy of a training set

$$\text{info}(T) = -\sum_{i=1}^{k} \frac{freq(C_i, T)}{|T|} \times \log_2\left(\frac{freq(C_i, T)}{|T|}\right) bits$$

with T a training set of cases, C_i a class i and freq(C_i, T) the number of cases in T that belongs to class C_i.

Entropy can also be measured after that T has been partitioned in n sets using the outcome of a test carried out on attribute X. This yields:

Definition 4: Entropy after the training set has been partitioned on a test X

$$\text{info}_X(T) = \sum_{i=1}^{n} \frac{|T_i|}{|T|} \times \text{info}(T_i)$$

Using these two measurements, the *gain criterion* can be defined as follows:

Definition 5: Gain criterion

gain(X) = info(T) - info$_X$(T)

The gain criterion measures the information gained by partitioning the training set using the test X. In ID3, the ancestor of C4.5, the test selected is the one which maximizes this information gain because one may expect the remaining subsets in the branches will be the most easy to partition. Note, however, that by no means this is certain because we have looked ahead only one level deep in the tree. The gain criterion has only proved to be a good heuristic.

Although the gain criterion performed quite well in practice, the criterion has one serious deficiency, i.e. it tends to favour tests with many outcomes. Therefore, in C4.5, a somewhat adapted form of the gain criterion is used. This criterion is called the *gain ratio criterion*. In this criterion, the gain attributable to tests with many outcomes is adjusted using some kind of normalization. In particular, the split info(X) measurement has to be defined.

Definition 6: Split info of a test X

$$\text{split info}(\mathrm{X}) = -\sum_{i=1}^{n} \frac{|T_i|}{|T|} \times \log_2\left(\frac{|T_i|}{|T|}\right)$$

This indicates the information generated by partitioning T into *n* subsets. Using this measure, the gain ratio is defined as follows:

Definition 7: Gain ratio

gain ratio(X) = gain(X) / split info(X)

This ratio represents how much of the gained information is useful for classification. In case of very small values of split info(X) (in case of trivial splits), the ratio will tend to infinity. Therefore, C4.5 will select the test which maximizes the gain ratio, but subject to the constraint that the information gain must be at least as large as the average information gain over all possible tests.

After building the tree, pruning strategies are adopted. This means that the decision tree is simplified by discarding one or more sub-branches and replacing them with leaves.

3.4 Evaluation Criteria

The overall aim of this study is to investigate whether there are different relevant variables necessary in predicting the use of manufacturers' versus retailers' coupons. For both coupons, a set of decision rules was extracted from the data. First, we will rank all the attributes and identify the most relevant ones for each coupon's strategy separately. Then, we will build the C4.5 trees incorporating only a subset of the most relevant attributes.

We used all cases of the training set (cf. infra) to build and optimize the decision trees for each decision step, while the data from the test set (an equally sized but out-of-sample test set) were used as validation set to compute accuracies (percentage of correctly classified instances).

To determine the selection of variables, the following procedure was adopted. Several decision trees were built, each time removing one more irrelevant attribute, as they appear lowest in the ranking that has been provided by the FS method. For each of these decision trees, its accuracy was calculated and compared to the accuracy of the full decision tree using all attributes and the decision tree yielding the highest performance. The smallest decision tree, which resulted in a maximum decrease of 2% in accuracy on the training set compared to the decision tree with the highest performance, was chosen as the final model. Based on this final model, predictions concerning coupon usage were made.

4. EMPIRICAL STUDY

4.1 Data

For the empirical study we made use of real-life customer data that were made available by a worldwide retailer in fast-moving consumer goods (FMCG). The data were collected thanks to the intensive use of the company's loyalty card. Around 85% of company's purchase incidences are registered and, consequently, attributable to individual clients. Besides, it is stored at the most detailed level. Two periods of observation were used so the data were split in a training set (September 2001) and an out-of-sample test set (October 2001). All company data stored between the purchase for which a redemption probability is built and 1st of January 2001, were used to compose the explanatory variables. For both models (store brand and national brand coupon use) two times 3 500 observations were used for the analysis.

Figure 2.1: Timing of data periods.

4.2 Predictors

The data that were made available consist of historical customer behavior and customer demographics at the individual level. Both types of data are repeatedly supported by prior research to be incorporated in predictive models (Baesens et al., 2002; Buckinx and Van den Poel, 2005). We incorporated as many predictors as possible in order to determine which type of information is the most important in predicting coupon redemption. Therefore, we compiled 98 explanatory variables based on all observed data. As a result five types of variables can be distinguished: variables capturing data concerning past coupon usage, variables containing information about promotional behavior, predictors capturing information about past purchase behavior, customer demographics and a class containing variables of different kinds. Table 1 gives an overview and a description of all derived predictors. The following paragraphs provide a motivation for the use of each of the variable types.

Table 2.1: Model predictors for manufacturers and retailers

Variable Type	*Variable Name*	*Description*	*Model*
			SB: Store brand model *NB: National brand model* *B: Both models*
Dependents	Store_use	store brand coupon use or not (0/1)	SB
	Nat_use	national brand coupon use or not (0/1)	NB
Coupon usage	TotNumCoup	total number of coupons used	B
	TotNumStore	total number of store brand coupons used	B
	TotNumNat	total number of national brand coupons used	B
	TotAmStore	total amount of store brand coupons used	B
	TotAmNat	total amount of national brand coupons used	B
	TotAmSpec	total amount ofspecial coupons used	B
	rSIAmCoup	total amount of coupons used relative to the frequency of shopping	B
	rSIAmStore	total amount of store brand coupons used relative to the frequency of shopping	B
	rSIAmNat	total amount of national brand coupons used relative to the frequency of shopping	B
	rSIFreqCoup	total frequency of coupons used relative to the frequency of shopping	B
	rSIFreqStore	total frequency of store brand coupons used relative to the frequency of shopping	B
	rSIFreqNat	total frequency of national brand coupons used relative to the frequency of shopping	B
	NumCoup	number of coupons during (last) shopping incidence	B
	MaxCoup	maximum number of coupons once used in a shopping incidence	B
	rLorFreqCoup	total frequency of coupons used relative to the lor	B
	rLorFreqStore	total frequency of store brand coupons used relative to the lor	B
	rLorFreqNat	total frequency of national brand coupons used relative to the lor	B
	rSpenAmCoup	total amount of coupons used relative to the total spendings	B
	rSpenAmStore	total amount of store brand coupons used relative to the total spendings	B
	rSpenAmNat	total amount of national brand coupons used relative to the total spendings	B
	RecencyCoup	Number of days since coupon is used during a visit	B
	RecencyStore	Number of days since store brand coupon is used during a visit	B
	RecencyNat	Number of days since national brand coupon is used during a visit	B
	RecencySpec	Number of days since special coupon is used during a visit	B
	InterCoup	Average number of days between coupon use	B
	InterStore	Average number of days between store brand coupon use	B
	InterNat	Average number of days between national brand coupon use	B
	InterSpec	Average number of days between special coupon use	B
	rSIFreqIncr_lor	Relfreq increased (1/0) compared to lor/2 (the situation half of the period till now)	B
	rSIFreqDecr_lor	Relfreq decreased (1/0) compared to lor/2 (the situation half of the period till now)	B
	rSIFreqIncr_last	Relfreq increased (1/0) compared to previous shop incidence	B
	rSIFreqDecr-last	Relfreq decreased (1/0) compared to previous shop incidence	B
	CoupUse_last	coupon used or not (0/1) in last purchase incidence	B
	StoreUse_last	store brand coupon used or not (0/1) in last purchase incidence	SB
	NatUse_last	national brand coupon used or not (0/1) in last purchase incidence	NB
Promotion variables	rSILoyPoints	Total number of loyalty points relative to the number of visits	B
	LoyPoints	Total number of points	B
	Points	Number of points collected during visit	B
	SpecPoints	Total number of extra points	B
	rSISpecPoints	Total number of extra points relative to the number of visits	B
			B
Past Purchase History	Brand(1-2-3-4)	Aggregated spending in 4 brand categories	B
	Brand(1-2-3-4)r	Aggregated spending in 4 brand categories relative to the total spending	B
	Cat(1-2-3-4-5-6-7-8-9-10-11-12)	Aggregated spending in 12 categories	B
	Cat(1-2-3-4-5-6-7-8-9-10-11-12)r	Aggregated spending in 12 categories relative to the total spending	B
	Monetary	Total spendings	B
	Frequency	Total number of visits	B
	Recency	number of days since last purchase	B
	rLorFrequency	Total number of visits relative to the lor	B
	rLorMonetary	Total spendings relative to the lor	B
	Lor	length of relationship (date - firstdate)	B
Demographics	Language	language of the customer	B
	Store	number of the store where one was registrated	B
	Housemember	number of housemembers	B
	Gender	Gender	B
	PostCode	postal code of the customer	B
	Storeid	number of store	B
Others	Meantime	mean timing of the day	B
	Mop(1-2-3-4-5-6-7)	The number of shop incidences where was paid by mode x	B
	Mop(1-2-3-4-5-6-7)r	The number of shop incidences where was paid by mode x relative to the number of visits	B

4.2.1 Coupon Redemption

Several different studies showed that the degree of sensitivity to promotional coupons is a function of previous purchases made with coupons (Bolton, 1989; Krishna, Imran & Shoemaker, 1991; Lattin & Bucklin, 1989; Winer, 1986). As a result, we computed variables capturing information concerning customers' prior sensitivity to coupons. "TotNumCoup" represents the total number of coupons a customer ever redeemed in the past. The same kind of variables was calculated for coupons related to national as well as store brands. "TotAmStore" and "TotAmNat" represent the aggregated face value of all coupons that were redeemed for store and national brand respectively. For these variables several transformations were computed in order to include relative figures as well. We made use of the total number of shop visits ("rSI"), the length of customers' relationship with the store ("rLor") and the total amount of money spent by the customer (rSpen").

"Numcoup" stands for the number of coupons the customer redeemed during his last visit to the shop while "MaxCoup" is the maximum number of coupons that were redeemed by a customer during one shopping incidence. "RecencyCoup" is the number of days that elapsed since a customer redeemed a coupon and "InterCoup" is the average number of days between customer coupon redemption.

In addition, we created variables that indicate whether the redemption of coupons relative to the frequency of store visits increased, decreased or remained stable compared to the situation middle the customers' length of relationship ("rSIFreqIncr_lor") and compared to the last purchase incidence ("rSIFreqIncr_last").

Finally, "CoupUse_last" is a dummy indicating whether or not (1/0) a customer made use of a coupon during his/her last visit.

4.2.2 Promotional behavior

For the same reason as the 'coupon redemption variables', we included variables that contain information concerning customers' sensitivity to promotions that are captured in other behavioral data.

Again, we expect this historical behavior to be determining for future sensitivity to coupons. We included "tpoints", which represents the total number of loyalty points one collected by means of the loyalty card. "rSIPoints" makes this relative to the number of shop visits and "Points" tells how much of these points were collected during the last visit to the shop.

"SpecPoints" and "rSISpecPoints" give an indication of the loyalty points that were gathered thanks to specific purchases.

4.2.3 Past purchase history

We also included variables that represent the spending in different types of brands (national brand, store brand, private brand and exclusive brands) as a representation of brand loyalty (Papatla & Krishnamurthi, 1996). Ortmeyer, Lattin & Montgomery (1991) indicate that customers with a low preference for a brand exhibit a limited sensitivity to promotions for that brand. So, customers with a low level of spending in a brand category are expected to be less interested in a coupon for a product out of that brand than customers with a high spending level. This can be detected from purchase history. We used the absolute aggregated spending level ("Brand1-4") and their relative transformation using the total spending level of the customer ("Brand1r-4r").

Additionally, variables concerning the purchases in twelve different product categories are included to improve model predictions. For some categories more coupons are distributed compared to other categories. The same hypothesis as in the previous paragraph can be made: customers buying products out of categories that frequently make use of coupons as a promotional marketing tool, are expected to have a higher redemption probability. Again, we included absolute ("Cat1-12") as well as relative ("Cat1r-12r") variable transformations.

Finally, we included general RFM related variables as well, like customers' total spending, number of visits and the number of days since they last visited the shop. We also added a variable "Lor" that indicates for how many years a customer already shops at the supermarket. The use of this type of variable is well supported in several studies dealing with response problems (Van den Poel, 2003).

4.2.4 Demographics

Prior research incorporates demographics in their attempt to explain coupon proneness. Whereas some studies support this inclusion, others found demographics to be poor predictors of behavior (Bawa, Srinivasan & Srivastava, 1997). Ailawadi et al. (2001) ascribe the weak explanatory power of demographics to the indirect effect of demographical variables and consider them to be associated with psychographics.

We include a customers' language, number of household members, gender, store of registration and postal code.

4.2.5 Others

These variables represent the way customers paid their bill at the checkout. The different modes of payment are: cash, cheques, lunch-allowance cheques, in-house vouchers, electronic payment, credit cards and the amount of money subtracted from the bill because of returned empty bottles. Absolute and relative versions were taken into account in the models ("Mop1-7" and "Mop1r-7r"). Finally, we included a variable "Meantime" giving an indication of the moment in time a customer normally comes to the shop.

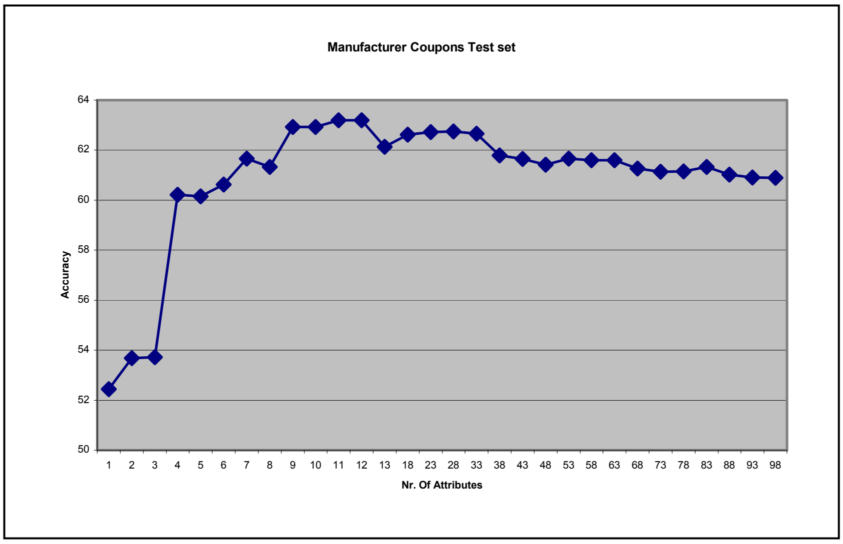
5. RESULTS

5.1 Determination of feature set size.

Figure 2 and Figure 3 show the classification results for different numbers of selected features on the test set. The results were obtained by applying the above-mentioned methodology.

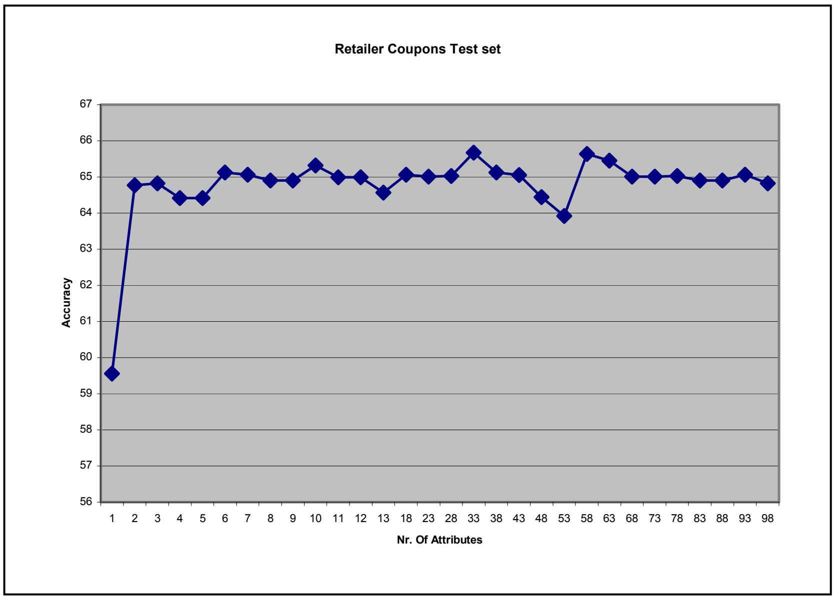
For the prediction of manufacturers' coupon use, the best set of features could be from nine till eleven. However, in order to be able to make an unambiguous conclusion on the number of variables the results on the training set were considered. This indicates that a classification accuracy of 62.92 per cent is to be achieved on the test set when the first nine features are selected versus an accuracy of 63.20 per cent when using eleven variables. The performance of the model when all features are incorporated on the test set is 60.89 per cent. This supports the expectation that as the number of features increases, the feature set starts to capture irrelevant and redundant information that adds noise and as a consequence degrades the performance of the classification technique. Similar results were found by Moons et al. (2001). We opted for the final decision tree using nine variables. This model is built on two variables less when compared to the eleven feature tree, but we only lose 1.23 per cent in accuracy.

Figure 2.2: Classification accuracy vs. number of best-selected features (manufacturer coupons)



Concerning the retailers' coupon use, the highest performance on the training set is achieved when incorporating eight features with an accuracy of 64.90 per cent on the test set. The use of all features results in a performance of 64.82 per cent, which again supports the necessity of making use of the feature selection technique. Taking into account the coupon redemption rate of 50 per cent in the training set, the PCC of both models substantially exceed Morrison's (1969) proportional chance criterion of 0.50 (= $0.50^2 + 0.50^2$), which establishes the no-model benchmark.

Figure 2.3: Classification accuracy vs. number of best-selected features (retailer coupons)



5.2 Interpretation of feature sets.

The results of the feature-selection procedure indicate that it is possible for manufacturers as well as for retailers to make predictions concerning their targets when distributing coupons. Moreover, both are able to detect different segments of customers, which releases the opportunity to avoid each other in their attempt to convince as many customers as possible. This is confirmed when having a look at the different feature sets. Table 2 presents for both coupon types the predictors that were picked by 'Relief-F'.

5.2.1 Manufacturers' coupons

The nine selected features to predict manufacturers' coupon use are to be found in different kinds of variable types (see part 4.2). More specifically, the average number of days between the use of national brand coupons (InterNat), the use of a national brand coupon during customers' last visit (NatUse_last), the use of coupons in general during someone's last visit (CoupUse_last) and the percentage of visits where one used a coupon (rSIFreqCoup) validate the importance of applying variables concerning past coupon redemption.
Concerning the variables that capture information about customers' sensitivity to promotions in general, only the number of loyalty points (rSILoyPoints) seems to be of importance.
In contrast to what we could find in our literature review, also demographics are able to explain a significant part of the variance in the data. First of all, the location of the store (Storeid) and the location of the customer are among the best-selected features (PostCode). Besides, the gender (Gender) of the customer and the number of members in the household (Housemember) define the chance of making use of national brand coupons.
Surprisingly, none of the features that we classified under the 'past purchase history' predictors showed up in the feature selection.

5.2.2 Retailers' coupons

For this classification problem eight features were selected: The average number of days between the use of store brand coupons (InterStore), the number of days since a store brand coupon was used (RecencyStore), the redemption of a store brand during someone's last visit to the shop (StoreUse_last) and the average number of days between the use of special coupons (InterSpec) are the best-selected features in the class of variables concerning past

redemption behavior. In comparison with the manufacturers' coupon use prediction the selected set of variables out of this class is totally different. However, they capture more or less the same information, related to the coupon type under investigation.
Compared to the previous section, none of the variables about promotional sensitivity seems to be important for retailers. Again the irrelevance of past purchase history for coupon-usage prediction is confirmed.
Finally, the same demographics that could be traced in the manufacturer coupon-usage prediction are selected for this application.

Table 2.2: Variable selection results.

Variable Type	*Manufacturer coupon*	*Retailer coupon*
Coupon usage	rSIFreqCoup	RecencyStore
	InterNat	InterStore
	CoupUse_last	InterSpec
	NatUse_last	StoreUse_last
Promotion variables	rSILoyPoints	-
Past Purchase History	-	-
Demographics	Housemember	Housemember
	Gender	Gender
	PostCode	PostCode
	Storeid	Storeid
Others	-	-

5.3 Retailers' versus manufacturers' coupons.

The results show the possibility for retailers and manufacturers to compose a marketing strategy that is better focused to specific types of customers. This confirms the expectation that diverse segments of people exist that differ concerning their coupon-redemption behavior. Consequently, four customer segments can be distinguished. First of all, we are able to identify customers that are only interested in coupons of the retailer. Secondly, there are customers that only redeem coupons of national brands. A third segment of people is interested in both types of coupons and, finally, customers that are interested in no coupons at all can be classified in a fourth segment.

6. CONCLUSIONS

This study examined the use of a feature-selection technique to facilitate and optimize the classification of customers according to their coupon-redemption behavior. Several interesting contributions were realized.

First, we are able to make predictions concerning customers' coupon redemption during their next visit to a supermarket. This way retailers as well as manufacturers can identify their targets when distributing coupons in order to improve their marketing strategies.

The results also confirmed the value of our feature-selection technique 'Relief-F'. Irrelevant information for the models could be detected. This facilitates the modeling since a selection of 9 (manufacturer coupon) and 8 (retailer coupon) features proves to be more accurate than the inclusion of the entire set of 98 explanatory variables.

The selected predictors are different for both predictive models but capture more or less the same information related to the specific coupon type that is examined. The most relevant features for manufacturers are the average number of days between the use of national brand coupons, the use of a national brand coupon during a customers' last visit, the use of coupons in general during someone's last visit, the percentage of visits where one used a coupon and the relative number of loyalty points collected. Retailers should make use of the average number of days between the use of store brand coupons, the number of days since a store brand coupon was used, the redemption of a store brand during someone's last visit to the shop and the average number of days between the use of special coupons in order to classify customers into redeemers or non-redeemers. Both dispensers of coupons, however, are recommended to include the same set of demographics: The number of house members, the gender of the customer, the postal code and the store where one shops.

Moreover, we can conclude that the inclusion of different types of predictors is relevant for the classification of coupon redeemers and non-redeemers.

Neither past purchase behavior nor variables related to the mode of payment or the timing of shopping (others type) are relevant whereas the former proved to be important information in many other studies (Van den Poel, 2003).

Separate models were built to capture a customers' coupon usage for retailers and manufacturers. As a consequence, the entire customer base can be split into four segments: customers who will redeem a store brand coupon, customers who will use a coupon of a national brand product, customers who will use both and finally customers who won't use a coupon of any type at all. A reduction of the conflict between retailers and manufacturers will lead to a better allocation of sources and will save money for both parties.

REFERENCES

Ailawadi, K.L., & Neslin, S.A. (1998). The effect of Promotion on Consumption: Buying more and Consuming it Faster. Journal of Marketing Research, 35(August), 390-398.

Ailawadi, K.L., Lehmann, D.R., & Neslin, S.A. (2001). Market Response to a Major Policy Change in the Marketing Mix: Learning from Procter & Gamble's Value Pricing Strategy. Journal of Marketing, 65(1), 44-61.

Ailawadi, K.L., Neslin, S.A., & Gedenk, K. (2001). Pursuing the value-conscious consumer: Store brands versus national brand promotions, Journal of Marketing, 65(1), 71-89.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., & Dedene, G. (2002). Bayesian neural network learning for repeat purchase modelling in direct marketing. European Journal of Operational Research, 138 (1), 191-211.

Bawa, K., & Shoemaker, R.W. (1987). The Effects of a Direct Mail Coupon on Brand Choice Behavior. Journal of Marketing Research, 24(4), 370-376.

Bawa, K., Srinivasan, S.S., & Srivastava, R.K. (1997). Coupon Attractiveness and Coupon Proneness: A Framework for Modeling Coupon Redemption. Journal of Marketing Research, 34(4), 517-525.

Bell, D.R., Jeongwen, C., & Padmanabahn, V. (1999). The Decomposition of Promotional Response: An Empirical Generalization. Marketing Science, 18(4), 504-526.

Blattberg, R., Eppen, G.D., & Lieberman, J. (1981). A Theoretical and Empirical Evaluation of Price Deals for Consumer Nondurables. Journal of Marketing, 45(Winter), 116-129.

Blattberg, R.C., & Neslin, S. (1990). Sales Promotion, Englewood Cliffs, NJ: Prentice Hall.

Bolton, R. (1989). The Relationship Between Market Characteristics and Promotional Price Elasticities. Marketing Science, 8(2), 153-169.

Bonnici, J., Campbell, D., Fredenberger, W.B., & Hunnicutt, K.H. (1996). Consumer Issues in Coupon Usage: An Explanatory Analysis. Journal of Applied Business Research, 13(1), 31-40.

Buckinx, W., & Van den Poel, D. (2005). Customer Base Analysis: Partial Defection of Behaviourally-Loyal Clients in a Non-Contractual FMCG Retail Setting. Under review European Journal of Operational Research, 164(1), 252-268.

Bucklin, R., & Gupta, S. (1992). Brand Choice, Purchase Incidence, and Segmentation: An Integrated Approach. Journal of Marketing Research, 29(May), 201-215.

Kira, K., & Rendall, L.A. (1992). A practical approach to feature selection. Proceedings of the Ninth International Conference on Machine Learning (pp. 249-256). Aberdeen, Scotland, UK (Morgan Kaufmann Publishers, San Mateo).

Kononenko, I. (1994). Estimating attributes: analysis and extensions of relief. Proceedings of the 7th European Conference on Machine Learning (pp. 171-182). Catania, Italy (Springer Verlag).

Krishna, A., Imran, S., & Shoemaker, R.W. (1991). Consumer Perceptions of Promotional Activity. Journal of Marketing, 55(April), 4-16.

Lattin, J.M., & Bucklin, R.E. (1989). Reference Effects of Price and Promotion on Brand Choice Behavior. Journal of Marketing Research, 26(August), 299-310.

Moons, E., Wets, G., Aerts, M., Vanhoof, K., Arentze, T.A., & Timmermans, H.J.P. (2001). The impact of irrelevant attributes on the performance of classifier systems in generating activity schedules. Proceedings of the 81st Annual Meeting of the Transportation Research Board (Washington, D.C.).

Moreau, P., Krishna, A., & Harlam, B. (2001). The Manufacturer-retailer-consumer triad: Differing perceptions regarding price promotions. Journal of Retailing, 77(4), 547-569.

Morrison D.G. (1969). On the interpretation of discriminant analysis. Journal of Marketing Research, 6(2), 156-163.

Ortmeyer, G., Lattin, J.M., & Montgomery, D.B. (1991). Individual differences in response to consumer promotions. International Journal of Research in Maketing, 8(3), 169-186.

Papatla, P., & Krishnamurthi, L. (1996). Measuring the Dynamic Effects of Promotions on Brand Choice. Journal of Marketing Research, 33(1), 20.

Quinlan J. R. (1993). C4.5 Programs for Machine Learning. San Mateo: Morgan Kaufmann Publishers.

Taylor, G.A. (2001). Coupon Response in Services. Journal of Retailing, 77(1), 139-151.

Van den Poel, D. (2003). Predicting Mail-Order Repeat Buying: Which Variables Matter? Tijdschrift voor Economie en Management, 48(3), 371-403.

Winer, R. (1986). A Reference Price Model of Brand Choice for Frequently Purchased Products. Journal of Consumer Research, 13(September), 250-256

CHAPTER III

PREDICTING ONLINE-PURCHASING BEHAVIOUR[4]

[4] This chapter is based on the following reference: Dirk Van den Poel, Wouter Buckinx, 2005. Predicting Online-Purchasing Behaviour, *European Journal of Operational Research*, Vol 166(2), 557-575.

CHAPTER III:

PREDICTING ONLINE-PURCHASING BEHAVIOUR

1. INTRODUCTION

Since the advent of the Internet, the possibilities with regard to the distribution of goods and/or services have changed substantially. Firms are able to offer goods/services not only through traditional channels such as retail outlets, but also in an online virtual store. But there is more to it than just the addition of a new channel of distribution.

First, whereas data captured from purchases in traditional stores only collect information concerning the *buying* behaviour of their clients, online data provide much more information (Moe and Fader, 2002). Clickstream data typically contain the trajectory of (prospective) clients at the company's website. Subsequently, also the visits that do not result in a purchase of one or more products/services are monitored that makes the customer picture, which firms are attempting to compose, more complete. Clickstreams offer the opportunity to thoroughly improve the understanding of customer activities being an important competitive advantage providing market research as a by-product (Andersen et al., 2000). Bucklin and colleagues (2002) conclude: "The detailed nature of the information tracked about Internet usage and e-commerce transactions presents an enormous opportunity for empirical modelers to enhance the understanding and prediction of choice behaviour".

Secondly, the Internet makes it possible to outline better client relations (Bauer et al., 1999). Customer relationship management (CRM) goes hand in hand with personalization of customer treatments, i.e., alternative strategies can be pursued for different segments as e.g. outlined in Baesens et al. (2004). Through their website, companies can communicate individually with their current clients and prospects. As a result, products, services and even marketing actions can be adjusted to the profile of visitors in order to influence (potential) customers' visiting and shopping behaviour. Finally, Moe and Fader (2001) argue that the more refined the segmentation or profiling of the customer base is, the more efficiently a profitable target segment can be identified.

However, being active in an e-business environment does not necessarily imply a bed of all roses. Clients or visitors of e-commerce websites are rarely loyal to a specific website when searching for a particular product or category (Johnson et al., 2000). Moreover, the conversion rate, defined as the percentage of website visits that lead to a purchase, is very low (Bucklin et al., 2003). One of the reasons is that costs of visiting e-commerce sites are limited compared to the offline world and may result in a delay of purchases (Moe and Fader, 2002). Besides, competition is fierce and clients are able to compare the offers of several companies in an instant's notice. Finally, buying online is not yet well-accepted behaviour and varies widely by product/service category (Van den Poel and Leunis, 1999). Sismeiro and Bucklin (2003) indicate that almost 75 per cent of the Internet users browsed or searched for a specific product but 65 per cent of the visitors never used the Internet to actually buy something. The Internet is most of the time used as an information source (Van den Poel and Leunis, 1999).

Finally, a lot of research still needs to be done concerning Internet usage since Internet choice behaviour is in many respects substantially different from the behaviour that is already thoroughly explored in a traditional store-retail setting (Bucklin et al., 2002). Internet choice behaviour is more dynamic, which provides modelers with more and different types of consumer choices. Besides, the intent of the visitor (browse, search or purchase) is not noticeable. Finally, the marketer has the opportunity to personalise the choice environment and respond in numerous ways at any moment in time. Consequently, other models are needed for understanding Internet behaviour and being able to make predictions about it.

In this study, we develop a model to predict whether a registered website user is going to purchase during the next visit. This enables us to derive individual purchase probabilities for each client in the customer database of an e-business website in order to know their future objective. The purpose is to differentiate customers based on as many as possible dimensions: past customer information concerning general clickstream behaviour and detailed clickstream measures, as well as historical purchase behaviour and customer demographics. To the best of our knowledge, no previous research incorporates variables from all of these categories in one and the same study. This is shown in Table 1, which will be discussed in Section 2.

In summary, we contribute to the existing literature in many respects: (1) We include a large list of predictors from different variable types into one and the same model. This offers the possibility to evaluate various predictor categories concerning their relevance for future purchase forecasts. (2) Since most of our proposed variables were never used before in other studies, we evaluate the gain in predictive power that can be attributed to their inclusion. Thanks to the numerous variables the variability in the model can be reduced so we are able to better classify customers concerning their future purchase behaviour on the Internet. In this process, different variable selection techniques are applied to identify the most important predictors for the model. We evaluate the advantage for online retailers who, in comparison to traditional retailers, have clickstream data at their disposal.

A limitation of this study is that we were restricted to customers who register before surfing the website. This results in limited size of the data set. However, the data we obtained were very elaborate in terms of the information that could be delivered. Therefore we still believe the findings are valuable for e-shops.

The remainder of this paper is structured as follows: Section 2 contains the literature review about predictions of online-purchasing behaviour and describes the topic of this study. The specifics about the dataset are discussed in Section 3, as well as the methodology used. However, the largest portion of this section is devoted to discussing the construction of explanatory variables to convert the massive amount of clickstream data into usable information. Results are reported in Section 4. Section 5 contains the conclusions, and limitations are reported in the final section.

2. LITERATURE REVIEW

As mentioned before in this paper, the conversion rate of a website is one of the major problems for e-commerce marketing managers. Consequently, most recent advanced studies focus on improving that conversion rate by examining the drivers of purchases. Sismeiro and Bucklin stated: "Predicting and understanding online-buying behaviour is of utmost importance for e-commerce website managers". Quantitative models that are commonly used in offline distribution channels prove to be useful in optimizing the use of clickstream data (Montgomery, 2001).

Moe and Fader (2001) were the first ones to investigate customer conversion rates over time. They showed that their more dynamic approach forecasted Internet behaviour significantly better than a model that does not take into account behavioural changes over time. Later on, they focused on analyzing the conversion of store visits into purchases based on historical visiting data (Moe and Fader, 2002) and the type of customer visit (2001). That way, predictions can be made for each customer concerning his probability of purchasing during a visit. Padmanabhan et al. (2001) predicted the probability that the remainder of a visit results in a purchase and if that user would make a purchase in any future session. Sismeiro and Bucklin (2003) show that browsing behaviour and experiences are predictive of online buying.

However, not all studies focus directly on purchase behaviour as the ultimate variable to predict. Bucklin and Sismeiro (2003) focused on the determinants of whether customers continue browsing or prefer to exit the site and examined the drivers of the length of time spent viewing a website page. Emmanouilides and Hammond (2000) constructed models to predict the status of visitors (active or lapsed) and the usage frequency of visitors. Li et al. (2002) developed a model to predict the number of webpages of specific categories viewed in a single session of a customer.

Although all of these studies improve managers' insight in how to approach different types of clients, some space is still left for further investigation. In order to address the problem of low conversion rates it is necessary to understand more in detail the features that control the visitor's decision whether or not to purchase (Bucklin et al., 2002). Site visitors typically are

Table 3.1: Literature review.

Author	Variables								
	Clickstream data							Additional data	
	General clickstream measures					Detailed clickstream measures			
	session frequency	timing	recency	time spent	# pages	other	page content	historical purchase behavior	demographics
Bucklin et al. (2002)	x				x	x			
Emmanouilides & Hammond (2000)									x
Li et al. (2002)					x				x
Moe (2001)							x		
Moe & Fader (2001)	x	x							
Moe & Fader (2002)	x	x	x					x	
Padmanabhan et al. (2001)	x			x					x
Sismeiro & Bucklin (2002)	x			x	x	x			
this study	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**

interested in either browsing for information or have the intention to make a purchase. This, however, is not clear for the online retailer. Moreover, Bucklin and colleagues pointed out that one of the first answers to that problem might be to infer the goal of the individual Internet user. That way, marketers will be able to define the best prospects for online purchasing.

3. DATA AND METHODOLOGY

3.1 Data

We used data of an anonymous commercial retailer selling wine and related products on the Internet. All server log files as well as purchase data and demographical customer information were at our disposal. The site gives visitors the opportunity to view general company information, to visit a community part providing general wine information, to shop wine, attributes and gifts and to participate at a wine auction. Moreover, the visitor can make use of an elaborate search function. Moreover, before people can visit specific parts they have to register using username and password. That way, customers have an individual virtual wine cellar. The data we used were collected from May 25th 2001 till April 18th 2002 so we could exploit almost ten months of clickstream data. Table 2 summarizes the visiting and purchasing behaviour at the specific wine selling site for all registered clients.

Table 3.2: Descriptives Internet behaviour for wine site (May 25th 2001 till April 18th 2002).

Variables	Frequency
Number of visitors	1382
Number of visits	10173
Number of purchases	3539
Number of purchasers	810

3.2 Preprocessing

Before we obtained a meaningful data set, several preprocessing tasks were executed. The method of preprocessing was based on research by Cooley et al. (1999).

In a first step, irrelevant elements in the log files were eliminated. We only want to record the actual visits to pages of the site. However, all files and pictures that appear on a website page are recorded separately. As a result a single request to a page of the website leads to the registration of several record lines in the files. Because the lines concerning files and pictures do not represent actual visitor behaviour, those lines are deleted.
Besides, the most important step in preprocessing the data is to link all data of different individuals and transform them into unique sessions. All log files were stored in a “Common Log Format” so only two variables can be used in order to identify the visitors: the ip address and the date of the visit. Consequently, techniques such as caching and the existence of Internet service providers make it very hard to identify unique visitors because several different clients are registered in the log files with the same ip address. Because this could lead to a great disturbance of the conclusions of this study, we only focus on *registered* clients who identified themselves when accessing the website with a username and password. Only these clients can be identified uniquely. If we would not do this, different visitors would wrongly be seen as one and the same customer. This focus, however, does not undermine the relevance of this study. Several studies confirm the importance of allocating resources first to existing clients rather than putting effort into acquiring new ones (Rust and Zahorik, 1993; Mozer et al., 2000). Moreover it is a well-known phenomenon that a small part of the customer base accounts for a large part of companies’ profit (Niraj et al., 2001) so marketing actions should be lined up with customers’ purchasing potential (Reichheld, 1996). Since registered clients can be expected to be among the most active ones and Moe and Fader (2002) indicated that new visitors of a site and existing customers exhibit a different pattern in behaviour, a specific model for these clients is justified. Certainly in the case of online retailing this is relevant since the overall conversion rate is very low and the difference between active clients and inactive clients may be even bigger than for traditional retailers. Finally, most e-commerce sites make use of a registration obligation before being able to make a purchase (e.g. Amazon.com).

In accordance with what Catledge and Pitkow (1995) indicated, we consider an interruption of 30 minutes and more between two page requests as a signal of a new session. Finally, we removed sessions of only one page request because they are not considered to be real visits (Bucklin and Sismeiro, 2003).

These preprocessing tasks resulted in a rather small dataset of 1382 observations. The data were randomly split in two parts: a training set and a test set (each 50 per cent). In 22.9 per cent of the cases, a purchase was made. This high rate can be explained by our focus on visitors who login when entering the site (analogous to Moe and Fader, 2002). Still, this means that a lot of registered clients do not make a purchase. This potential confirms the relevance of building a choice model for this specific group of customers (see above).

3.3 Model Variables

In order to determine individual purchase probabilities we take into account as many as possible customer data. In comparison with previous studies, we include variables from several different categories: General as well as detailed clickstream measures, customer demographics and past purchase behaviour. By merging this information, we hope to maximize the predictive power of our modelling exercise (Montgomery, 2001) and to detect the most valuable predictors of online purchasing. The paragraphs below describe each of these types. Appendix 3.A - presents an overview of all variables that were used in our model. The next to last column indicates whether the variable already is employed in previous studies and whether it was found to be statistically significant.

3.3.1 General Clickstream Measures

These variables concern data measured at a rather general level of the clickstreams. They represent information at the level of the sessions. A session is a single visit to the website.
As can be concluded from Table 1, the variable most often incorporated in the literature is the frequency of visits. Moe and Fader (2001) show that frequent visitors have significantly higher conversion rates than infrequent visitors.
Later on, they introduced recency in a logistic regression to capture more of customers' past behaviour. But compared to their conversion model the additional information did not

increase the performance of the model. The accumulated visits proved to be the best indicator of purchase potential (Moe and Fader, 2002).

Padmanabhan and colleagues (2001) introduce the time visitors spent at a specific website during one session. This seemed to be significant in their site-centric model and positively correlated with potential purchase.
Bucklin et al. (2002) concluded that besides the measure for repeat visits also the visit depth, being the cumulative number of session page views, influences visitors' propensity to continue browsing. Li et al. (2002) could explain half of the sample variance while making predictions about which sessions would result in retail visits. They also took into account the page views of customers. Consequently, we included general clickstream measures in our model. Among others, the number of past sessions (FrequencyVisit), the time elapsed since the last visit (RecencyVisit), the average time someone spent during his session (AverageVisitTime), the total time spent at the site during the entire period of observation (TotalVisitTime) and the number of page views (TotalClicks) are taken into account in order to make predictions about future purchase probabilities. We computed some variants on these variables as well. The variable 'Hurry' indicates whether the average time of the clicks during the last session was less than the average over the past.

3.3.2 Detailed Clickstream Measures

Besides the general information about customers' sessions, we like to introduce the use of more detailed information in making predictions concerning purchase incidences. To the best of our knowledge, only a limited number of studies incorporated detailed clickstream data. Moe (2001) used the general content of the pages viewed to make distinctions between customers concerning their purchase likelihood. She categorized pages as buying, browsing, searching or knowledge-building ones. Capturing the percentage of pages that was downloaded in each category, significant differences could be found related to the conversion rate.
As a consequence, in this study the content of the pages that were visited will be taken into account to compose purchase probabilities for each client individually. We assigned each of the retailer website pages to one of the following categories:

- information concerning procedures how to navigate the site;
- information concerning the supply of purchased goods;

- information concerning the company;
- community pages;
- wine (bottles);
- wine accessories;
- bundled products;
- search engine of the site;
- gifts;
- personal pages.

Consequently, we computed for each of these categories the number of page visits by a customer. Besides an absolute number of page views during the last session we also computed aggregated variables representing all past page views and the ratio of the page views in each category with the total pages viewed during the visit. Moreover, relative alternatives for each of these categories were included by taking into account the number of past visits to the retailer's site and the total number of past clicks. Finally, some predictors are made relative to the number of past purchases made (cf Appendix 3.A).

3.3.3 Customer Demographics

Besides behavioural data, also demographic information of the customers seems to contribute in classifying customers as buyers or non-buyers. Padmanabhan and colleagues (2001) support the use of several demographic variables in predicting purchase probabilities. They include gender, age, customers' income, education level, the household size and the presence of children. Besides, Li et al. (2002) include users' demographics as well: customers' age, gender and race. Consequently, customer demographics were taken into account to make predictions concerning future online purchases. The following predictors were included: customers' gender (Gender), age (Age), language (French, Dutch or English are represented by the dummy variables LanguageD and LanguageF) and two trust indicators (Trust and Profsup). These predictors represent whether the visitor did or did not put his telephone number (Trust) or profession (Profsup) at the site owner's disposal when registering. This can be an indication concerning the degree of trust someone has in a e-commerce firm (Buckinx and Van den Poel, 2005). Finally, not every customer did mention his age or gender. Both facts are represented by an additional variable as well (Gendersup and Agesup).

3.3.4 Historical Purchase Behaviour

In the offline world, observed historic purchase behaviour already proved to be commonly used. Past purchase data are widely available and prove to be effective and rich predictors (Schmittlein and Peterson, 1994). The frequency of past purchases is positively related to a customer's future buying behaviour (Lemon et al., 2002). The same relationship is found for customer's past spending at a company. Jones and Sasser (1995) indicate that the amount a customer is buying is commonly used as an indication of loyalty. Moreover, Schmittlein and Peterson (1994) as well as Baesens et al. (2002) confirm this theory by concluding that spending is effective for future purchase predictions. Finally, the time elapsed between purchases might be an indicator for future buying patterns. Wu and Chen (2000) verify that customers who recently purchased are more likely to be active than customers who shopped a long time ago. Of course, all this evidence was found in offline environments. To our knowledge, only Moe and Fader (2002) incorporated past purchase behaviour in making purchase predictions for an e-commerce setting (Table 1).

We include several variables capturing past purchase behaviour: the number of purchases ever made, the percentage of visits that lead to a purchase, the dollar value spent, the average spending during a visit, the spending during the last visit, the average spending when a purchase is made and the number of days that elapsed since the last purchase of the customer. For an overview of all predictors we incorporated in the model we refer to Table 3.A in the Appendices.

3.4 Classification: Logit Modelling

We use logit modelling to answer the question whether or not a purchase is made during the next visit using the set of predictors described in Section 3.3. This choice is justified by the following reasons: (1) logit modelling is well-known, conceptually simple and frequently used in marketing (Bucklin and Gupta, 1992) both at the aggregate market level (Bultez and Naert, 1975), at the segment level (Mela et al., 1997), and at the level of the individual consumer (Jones and Zufryden, 1980); (2) the ease of interpretation of logit is an important advantage over other methods (e.g. neural networks); (3) logit modelling has been shown to provide good and robust results in general comparison studies (Levin and Zahavi, 1998).

The binary logit model is used to approximate a probability P_i, which is constrained to the range from 0 to 1 by the following expression (Aldrich and Nelson, 1984):

$$P_i = \frac{1}{1 + e^{-\sum_{j=1}^{n} b_j X_{ij}}}$$

Whereby: P_i represents the *a posteriori* probability of purchase by customer i;
X_{ij} represents independent variable j for customer i;
b_j represent the parameters (to be estimated);
n represents the number of independent variables.

Once the parameters b_j are estimated using maximum likelihood, this expression allows us to obtain a conditional probability estimate of purchase, which (1) has the properties of a probability, and (2) can be used to rank the customers in terms of their probability of purchase.

3.5 Variable selection procedures

Specifically in our application, the relatively small training set of 690 observations calls for a parsimonious technique, as well as model when considering the 92 available predictors. Therefore, we choose to use a variable-selection procedure. Both a forward and backward variable-selection procedure are used. A forward-selection procedure starts with an empty set of predictors and adds independent variables one-by-one to the logistic regression model choosing the most significant variable based on the score chi-squared statistic. The backward-selection procedure operates analogously but in the opposite direction starting from the total set of predictors and eliminating the least significant independent variable one-by-one. However, both procedures are appealing because they are quick and easy, but do not guarantee that the *best* subset be found. Therefore, to further increase our confidence in the results we use the global score algorithm proposed by Furnival and Wilson (1974) to find the *best* subset of a given number of predictors according to the score chi-squared statistic. They developed an efficient branch and bound algorithm to avoid an exhaustive search of the variable space, which would require, in this study with 92 variables, the estimation of all possible single and multivariate model combinations. After running all three procedures, we checked whether adding a quadratic term for certain variables (based on theory) resulted in improved classification performance. This additional step will be discussed in the results section.

When evaluating the overall model performance, we use two well-established criteria to measure classification performance: (1) percentage correctly classified (accuracy) (PCC; Morrison, 1969); (2) area under the receiver operating characteristic curve (AUC; Hanley and McNeil, 1982). Both measures are complementary because the former is intuitive and the latter is independent of the specific cutoff value chosen (e.g. probability threshold of 0.5).

3.6 Importance of variable types

This is the first study to incorporate such an extensive set of variables in order to predict e-commerce purchase behaviour. All variables can be classified into four variable types (see above). One of our objectives is to investigate the importance of each of the variable types in making purchase predictions. Therefore, we launched a forward-selection like procedure. In a first step we included alternately all the variables classified in each of the groups. Afterwards all the variables of the group that generated the highest AUC on the test set were enclosed in the model (step 1). Next, we checked the change in model performance when including alternately each of the other three groups of variables. This procedure was repeated until all variable groups were included into the model (step 2 - step 4).

In addition, we performed an analysis to identify the performance contribution of the variables that were never used in previous research. Appendix 3.A (one but last column) shows that several variables, especially general clickstream variables, were already examined in past studies. Therefore, we again launched a forward-selection like procedure. Here, we first included all variables that were already used in past studies (step 1). Next, we checked the change in model performance when including alternately each of the four groups of variables and included the group that generated the highest increase in predictive performance (step 3 – step 5). Thanks to this method insight is gathered concerning the contribution of our proposed variables above the existing ones and concerning the importance of each of the variable types.

4. RESULTS

Before discussing the model results, we first have to choose the number of predictors to be included in the logit model according to Furnival and Wilson's (1974) global score algorithm. Figure 1 plots the increase in the global score chi-squared statistic as the number of variables allowed into the model increases (we only plot the performance of the best subset). First differences of these results are shown in Figure 2. We learn from this chart that the performance increase clearly stabilizes after nine predictors.

Figure 3.1: Global score Chi-Squared Statistics

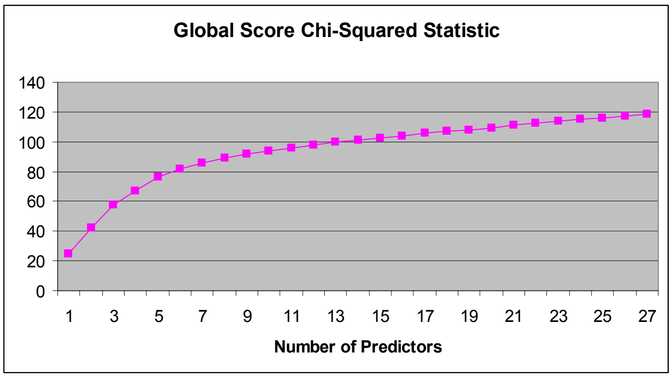
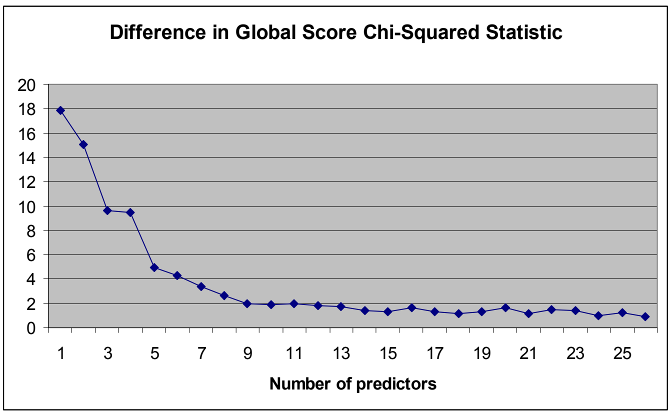


Figure 3.2: Difference in Global Score Chi-Squared Statistics



The results in Table 3 show that the (forward as well as backward) variable-selection procedure results in almost identical models, i.e., containing the same set of predictors with similar parameter estimates. The only difference is that in the forward procedure a relative measure is used as opposed to an absolute measure in the backward selection procedure. As mentioned in the previous section, after running the procedures, we checked whether adding a quadratic term to the different recency variables (RecencyVisit and PurchaseRecency) in the model improved classification performance. This is based on the findings by Gönül and Shi (1998). All three results suggested that the quadratic term was only important for the time since last visit (RecencyVisit), not for the time since last purchase (PurchaseRecency). When considering the parameter estimates of identical variables we observe that they are very close to each other, e.g., the estimate of the 'RecencyVisit' parameter in the forward model (0.0409) does not differ very much from the estimate in the backward model (0.0412).

When interpreting quantitative variables in a logit analysis, it is helpful to subtract one from the odds ratio and multiply by 100 to obtain the percentage increase/decrease associated with a one-unit increase of the variable (Allison, 1999, p. 29). Hence, the odds ratio of 1.042 of RecencyVisit (which stands for the number of days since last visit) for the forward procedure translates into the following meaning: The odds of a purchase are 4.2 per cent higher per additional day since last visit to the website. The significance of both the linear and quadratic term for RecencyVisit, as well as their respective signs (positive linear parameter, negative quadratic parameter) point to the fact that an inverted U-shaped relationship exists between the time elapsed since last visit and the probability of purchase. The next variable from the 'General Clickstream' category, retained in all models, is related to the relative speed of clicking during the previous visit (Hurry): The higher this speed, the lower the probability of purchase.

In the 'Detailed Clickstream' category, the number of pages viewed relating accessories during the last visit (PageAcc) turns out to substantially increase the odds of a purchase during the next purchase occasion (by 90.1 per cent). On the other hand, the number of personal pages viewed during the last visit (PagePers, or alternatively AveragePagePers) lowers the odds of buying on the next visit (by 6.1 per cent for the backward model).

Gender turns out to be an important demographic variable. Based on the odds ratio, males seem to have a higher tendency to purchase during their next web site visit in comparison to

Table 3.3: Results of Forward and Backward Variable Selection Procedure for Logit Analysis

		Forward					*Backward*				
		Estimate	*Std Error*	*Wald Chisq*	*P(W>Chisq)*	*Odds ratio*	*Estimate*	*Std Error*	*Wald Chisq*	*P(W>Chisq)*	*Odds ratio*
	Intercept	-1.0638	0.2137	24.7788	<.0001		-1.1103	0.2116	27.5377	<.0001	
General clickstream	RecencyVisit	0.0409	0.0094	18.9275	<.0001	1.042	0.0412	0.0094	19.2213	<.0001	1.042
	RecencyVisit^2	-0.0002	<.0001	7.4868	0.0062	1.000	-0.0002	<.0001	7.2265	0.0072	1.000
	Hurry	-0.7016	0.2503	7.8612	0.0051	0.496	-0.7386	0.2507	8.6812	0.0032	0.478
Detailed clickstream	PageAcc	0.6426	0.2516	6.5221	0.0107	1.901	0.6717	0.2496	7.2448	0.0071	1.958
	AveragePagePers	-2.5098	0.9756	6.6179	0.0101	0.081					
	PagePers						-0.0629	0.0260	5.8553	0.0155	0.939
	PageProductrPur	-0.0300	0.0083	13.2125	0.0003	0.970	-0.0282	0.0083	11.6665	0.0006	0.972
Customer demographics	Gender	0.5894	0.2145	7.5524	0.0060	1.803	0.6452	0.2116	9.2935	0.0023	1.906
	Trust	0.9720	0.2682	13.1367	0.0003	2.643	0.9329	0.2663	12.2754	0.0005	2.542
Historical purchase behaviour	PurchaseRecency	-0.0051	0.0013	14.4034	0.0001	0.995	-0.0054	0.0013	17.3105	<.0001	0.995

Table 3.4: Results of Best Subset Variable Selection for Logit Analysis

		Best subset procedure				
		Estimate	*Std Error*	*Wald Chisq*	*P(W>Chisq)*	*Odds ratio*
	Intercept	-1.3167	0.2305	32.6426	<.0001	
General clickstream	RecencyVisit	0.0379	0.0092	16.9796	<.0001	1.039
	RecencyVisit^2	-0.0002	0.0001	5.6687	0.0173	1.000
	Hurry	-0.7323	0.2489	8.6548	0.0033	0.481
Detailed clickstream	PagePers	-0.0534	0.0239	4.9879	0.0255	0.948
	TotPageProduct	-0.0093	0.0028	10.9138	0.0010	0.991
Customer demographics	Gender	0.6920	0.2209	9.8157	0.0017	1.998
	Trust	1.0124	0.2658	14.5018	0.0001	2.752
Historical purchase behaviour	TotPurchases	0.0478	0.0141	11.4667	0.0007	1.049
	PurchaseRecency	-0.0040	0.0013	9.4792	0.0021	0.996

women. The result with regard to the 'Trust' variable confirms the finding by Buckinx and Van den Poel (2005) that a desire *not* to report particular information (in this case a phone number) may signal 'distrust' to engage in a particular type of behaviour (i.e., lower probability of purchase).

When combining the results of Tables 3 and 4, we may conclude that all three procedures result in including variables from all four predictor categories (general clickstream, detailed clickstream, customer demographics and historical purchase behaviour). Apart from the recency variable (PurchaseRecency), which is also included in the forward and backward models, the global score best subset variable-selection procedure also includes a frequency variable (TotPurchases). This confirms that at least two from the Recency, Frequency Monetary value (RFM) variables, well-known from direct and database marketing, are also retained in online-purchase predictions (Baesens et al., 2002).

Appendix 3.B shows the correlation matrix of all the variables that are selected by one of the variable selection techniques. None of the variables appearing in one and the same selection do have a high correlation. Moreover, correlations of more than 0.50 only exist between variables that contain more or less the same information and are selected by another technique (AveragePagePers and PagePers, PageproductrPur and TotPageProduct). This supports the efficacy of the applied procedures being the addition of only these variables to the subsets that contribute to a better predictive power without having a high correlation with the already selected predictors.

Table 3.5: Overall Model Performance

	Forward		*Backward*		*Best subset procedure*	
	PCC	*AUC*	*PCC*	*AUC*	*PCC*	*AUC*
Estimation sample	0.7681	0.7605	0.7652	0.7581	0.7565	0.7541
Test sample	0.6893	0.6586	0.6806	0.6497	0.6806	0.6546

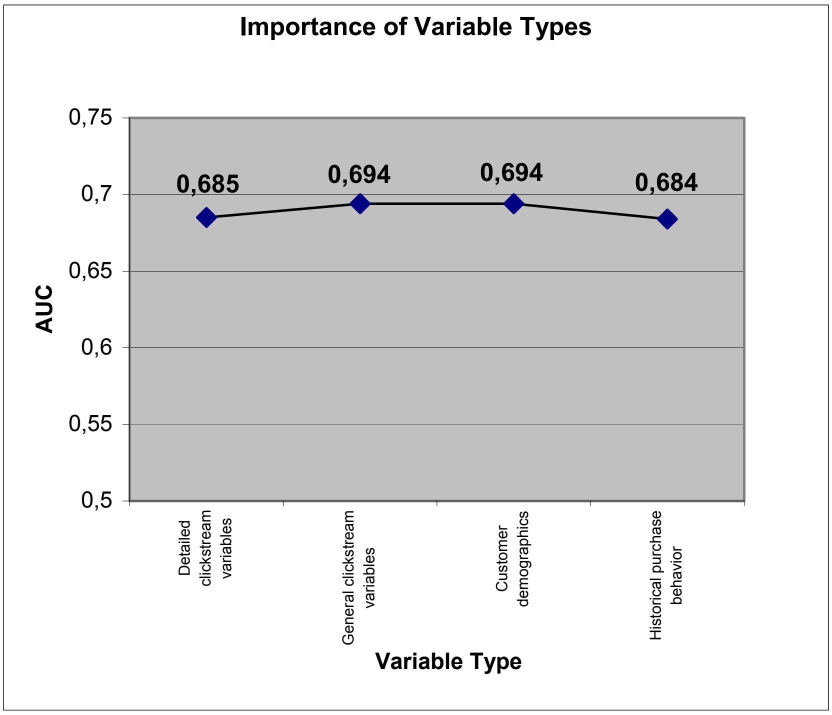
Table 5 contains the overall performance results of both models. We immediately observe the substantial drop-off both in terms of PCC and AUC when moving from the estimation sample to the test sample. Nevertheless, the PCC of both models still substantially exceeds Morrison's (1969) proportional chance criterion of 0.6453 (= $0.771^2 + 0.229^2$), which establishes the no-model benchmark. An analogous conclusion holds for the AUC criterion because all reported AUC results exceed the null-model benchmark of 0.5.

Table 6 and Figure 3 show the results of the above-mentioned procedure that gives insight into the importance of the different variable types.

Table 3.6: Importance of Variable Types.

	Step 1		Step 2		Step 3		Step 4	
	Detailed Clickstream		Customer Demographics		General Clickstream		Historical Purchase Behaviour	
	PCC	*AUC*	*PCC*	*AUC*	*PCC*	*AUC*	*PCC*	*AUC*
Estimation sample	.742	.756	.762	.783	.779	.809	.791	.825
Test sample	.683	.685	.704	.694	.704	.694	.701	.684

Figure 3.3: Importance of Variable Types



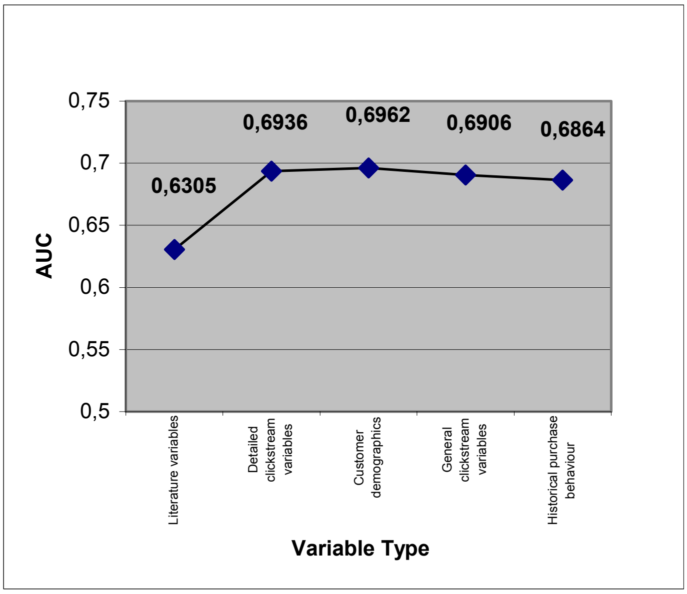
They indicate that the group of detailed clickstream variables generates the best predictive performance. The addition of all general clickstream variables does not improve the results significantly. Besides, neither customer demographics nor historical purchase predictors are able to increase the power of the model. However, these results are obtained when incorporating all variables from each of the categories.

Table 7 and Figure 4 show the performance contribution of the proposed variables that were never discussed in previous research. Detailed clickstream variables again generate the highest increase in terms of model performance. The performance of the model increased from an AUC of 0.630 to an AUC of 0.693. The DeLong et al. (1988) non-parametric test to statistically compare AUC's shows a significant difference between both models[5]. However, neither customer demographics, nor additional general clickstream measures and the historical purchase measures contribute to the variables already proposed in past research.

Table 3.7: Evolution of Model Performance.

	Step 1		Step 2		Step 3		Step 4		Step 5	
	Literature variables		Detailed Clickstream		Customer Demographics		General Clickstream		Historical Purchase Behaviour	
	PCC	*AUC*	*PCC*	*AUC*	*PCC*	*AUC*	*PCC*	*AUC*	*PCC*	*AUC*
Estimation sample	.7043	.708	.7710	.787	.7710	.806	.7739	.810	.8058	.821
Test sample	.6835	.630	.7095	.693	.7038	.696	.7038	.690	.7009	.686

Figure 3.4: Contribution of "new" Variables.



[5] $\chi^2 = 8.78$; p-value = 0.003

5. DISCUSSIONS AND CONCLUSIONS

In this study we predict, based on an extensive set of predictors from different categories, whether a (potential) customer will engage in online-purchasing behaviour during his next visit to the website. The ability to do so provides a powerful predictive tool for (r)etailers that helps them in inferring the goal of their visitors and, consequently, to improve their targeting. As mentioned in Section 2, this is considered to be among the most important steps to improve online conversion rates (Bucklin et al., 2002).

Past research already incorporated some of the presented predictors to examine the relationship with purchase propensity. However, they all considered a small selection of variables in their studies. This study not only incorporates the proposed predictors into one and the same model but also adds several new predictors to those proposed in earlier studies.

To the best of our knowledge we take into account many more variables (92) than all previous studies dealing with online purchasing. Not only general information concerning the visits was captured, but also detailed predictors telling more about the kind of pages visitors are interested in. In addition, data concerning historical purchases were incorporated and, finally, customer demographics were observed too. The results show clearly the magnitude of our contribution since the performance of the model increased significantly by putting them together in the same model. Detailed clickstream variables are shown to be more important than general clickstream variables, which were until now the most frequently used predictors in past studies.

The number of inputs needed to achieve the best predictive performance could be reduced thanks to the use of different selection techniques. This enables managers to avoid the collection of the entire range of predictors and to focus on the most important ones. The results highlight that predictors from all four categories are important when predicting online-purchasing behaviour since variables from all four types were selected by the three selection procedures that were applied. Independently from each other, these techniques selected the most important variables in the model. All of them came to more or less the same subset of predictors. The last column in Appendix 3.A indicates which variables were found to be statistically significant in a univariate logistic regression. Though many variables are relevant, only nine of them are retained by the different selection procedures. The correlation matrix of the variables in these subsets shows the absence of multicollinearity that would be present when incorporating all of the predictors. The most important variables that result from the selection techniques are the number of days since visitors' last visit, the speed of the clickstream behaviour during the last visit, the number of accessories viewed during last visit, the number of personal pages viewed, the number of products viewed, the gender of the customer, the fact of supplying personal information to the company, the number of days that elapsed since visitors' last purchase and the number of past purchases. Seven of them are new variables, never used before in an e-commerce study. This confirms the additional power of our proposed variables.

An examination of each of the inputs shows their relevance since a lot of them are significant when performing a univariate analysis (Appendix 3.A). Most of the earlier proposed variables in past studies turn out to be statistically significant in our study as well. Though some of them are not: The time per click and the percentage of specific pages viewed seem to be less relevant for this application. Moreover, the gender of visitors turns out to be relevant for the prediction of purchasing wine (related) products where it was not in an earlier study from Emmanouilides and Hammond (2000).

So the model presented in this paper offers a more in-depth investigation of conversion behaviour compared to previous studies. This results in a higher predictive ability and a better way to classify customers concerning their future purchase behaviour on the Internet. This is a significant contribution in understanding the features that control a visitor's decision whether or not to make a purchase. Moreover, we can limit the number of necessary inputs based on the different selection techniques.
Based on these findings, marketing managers can define which of the customers will visit their site with purchase intentions, adding to the CRM capability of the company. That way adapted messages can be communicated to the right customers containing recommendations of products (Jonker et al., 2004; Van den Poel and Larivière, 2004). Alternatively, the ad content of webpages can be personalized, certainly in the case of the retailer who owns the site under investigation, since each of the registered customers has access to some personal pages. This makes sense since Mandel and Johnson (1999) prove that what visitors are exposed to has an impact on their purchase behaviour and, consequently, on the site conversion rate. Häubl and Trifts (2000) confirm this and claim that, for example, decision

aids enhance the quality of customers' purchase decisions. Moreover, managers can take action to customers whose purchase probabilities are low. An intensification of the advertising exposure or the offering of specific promotional incentives is recommended (Mandel and Johnson, 1999; Sismeiro and Bucklin, 2003). Alternatively, the retailer could consider avoiding the exposure of unappreciated ads to customers who are not interested in shopping at all (Li et al., 2002). In addition, the company could contact the visitors that never are eager to purchase in order to get more information concerning the reasons for this behaviour.

This study confirms the advantage of online retailers (Anderson et al., 2000) compared to traditional retailers. The former are expected of being able to improve the understanding of the customer and enhance the understanding of choice behaviour since more (detailed) search/browsing information is at their disposal (Bucklin et al., 2002). The selection of relevant predictors out of each of the variable types confirms this expectation. It is not only historical purchase behaviour and customer demographics, being the information at the traditional retailers' disposal, that determine the future purchasing behaviour of visitors. General and mainly detailed clickstream variables enhance the predictive performance of purchasing behaviour.

6. LIMITATIONS

The results are not generalizable to all visitors to a website. In order to have as much customer information as possible at our disposal, we were forced to build a model for registered clients only. As argued before, this does not weaken the relevance of the findings. The focus on registered clients makes us optimising the treatment towards the most active group among customers. Several past studies confirm the usefulness of adapting strategies to customer potential. Moreover, as shown with the variable-selection techniques, customer demographics appear to be relevant information for the predictive power of the model. So being able to convince customers to register will provide useful information. Once also not-registered clients can be identified uniquely in order to join all different data types that were used in this study, it will be possible to build a separate model for these customers as well.

Conclusions were based on the investigation of clickstream data of one website, whereas Padmanabhan (2001) indicates that the performance of predictions concerning future purchases with user-centric information (using clickstream data from multiple websites) outperforms the ones based on site-centric information.

Besides, an application of the same model on clickstream information of other e-commerce sites will improve the insight into the generalizability of the results. Like in many other studies we were also restricted to the data of one online retailer. Getting this elaborate set of data of an online retailer is not obvious.

The results of this study are based on a small dataset. However, the data we obtained were very extensive in terms of the different types of information that could be delivered. Consequently, we still believe the findings are valuable for e-shops. Whether the results only hold for small e-commerce companies or can be generalized to all shops should be tested in additional studies.

We cannot claim the impact of customized treatments that we recommended to undertake for specific groups of customers. Further research, preferably by real-life experiments, has to give more insight into whether personalized website pages or product recommendations have an impact on the performance of the online retailer.

References

Aldrich, J.H., Nelson, F.D. (1984), Linear probability, logit, and probit models. Sage Publications.

Allison, P.D. (1999). Logistic Regression Using the SAS System. SAS Institute Inc., Cary, N.C.

Andersen, J., Giversen, A., Jensen, A.H., Larsen, R.S., Pedersen, T.B., Skyt, J. (2000). Analyzing Clickstreams Using Subsessions. *Proceedings of the 3th International Workshop on Data Warehousing and OLAP,* 25-32.

Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P., Vanthienen, J. (2004). Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. *European Journal of Operational Research,* 156 (2), 508-523.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., Dedene, G. (2002). Bayesian neural network learning for repeat purchase modelling in direct marketing. *European Journal of Operational Research,* 138 (1), 191-211.

Bauer, H.H., Grether, M., Leach, M. (2002). Customer Relations through the Internet. *Journal of Relationship Marketing*, 1(2), 39-56.

Buckinx, W., Van den Poel, D. (2005). Customer Base Analysis: Partial Defection of Behaviourally-Loyal Clients in a Non-Contractual FMCG Retail Setting. *European Journal of Operational Research,* 164(1), 252-268.

Bucklin, R.E., Sismeiro, C. (2003). A Model of Web Site Browsing Behaviour Estimated in Clickstream Data. *Journal of Marketing Research*, 40(3), 249-267.

Bucklin, R.E., Gupta, S. (1992). Brand choice, purchase incidence and segmentation: an integrated modelling approach. *Journal of Marketing Research*, 29(2), 201-215.

Bucklin, R.E., Lattin, J.M., Ansari, A., Gupta, S., Bell, D., Coupey, E., Little, J.D.C., Mela, C., Montgomery, A., Steckel, J. (2002). Choice and the internet: From clickstream to research stream. *Marketing Letters*, 13 (3), 245-258.

Bultez, A.V., Naert, Ph.A. (1975). Consistent sum-constrained models. *Journal of the American Statistical Association*, 70, 529-535.

Catledge, L., Pitkow, J. (1995). Characterizing browsing in the World-Wide Web. *In Proceedings of the Third International World Wide Web Conference*.

Cooley, R., Mobasher, B., Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems,* 1(1), 5-32.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44, 837-845.

Emmanouilides, C., Hammond, K. (2000). Internet usage: predictors of active users and frequency use. *Journal of Interactive Marketing*, 14(2), 17-32.

Furnival, G.M., Wilson, R.W. (1974). Regressions by Leaps and Bounds. *Technometrics*, 16, 499-511.

Gönül, F., Shi, M.E. (1998). Optimal mailing of catalogs: a new methodology using estimable structural dynamic programming models. *Management Science*, 44 (9), 1249-1262.

Hanley, J.A., McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.

Häubl, G., Trifts, V. (2000). Consumer decision making in online shopping environments: the effects of interactive decision aids. *Marketing Science*, 19(1), Winter, 4-21.

Johnson, E., Moe, W., Fader, P., Bellman, S., Lohse, J. (2000). On the depth and dynamics of online Search Behaviour. *Wharton Marketing Department Working Paper,* June.

Jones, T.O., Sasser, W.E.Jr. (1995). Why Satisfied Customers Defect. *Harvard Business Review,* November-December, 88-99.

Jones, J.M., Zufryden, F.S. (1980). Adding explanatory variables to a consumer purchase behaviour model: an exploratory study. *Journal of Marketing Research*, 17(3), 323-334.

Jonker, J.J., Piersma, N., & Van den Poel, D. (2004), Joint optimization of customer segmentation and marketing policy to maximize long-term profitability, *Expert Systems with Applications*, 27 (2), forthcoming.

Lemon, K.N., White, T.B., Winer, R.S. (2002). Dynamic Customer Relationship Management: Incorporating Future Considerations into the Service Retention Decision. *Journal of Marketing*, 66, 1-14.

Levin, N., Zahavi, J. (1998). Continuous predictive modeling: a comparative analysis. *Journal of Interactive Marketing*, 12 (2), 5-22.

Li, S., Liechty, C., Montgomery, A.L. (2002). Modeling Category Viewership of Web Users with Multivariate Count Models, *Carnegie Mellon Working Paper.*

Mandel, N., Johnson, E.J. (1999). Constructing Preferences Online: Can Web Pages Change What You Want? *Wharton Marketing Department Working Paper.*

Mela, C.F., Gupta, S., Lehmann, D.R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research*, 34(2), 248-261.

Moe, W. (2003). Buying, searching, or browsing: differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology,* 13(1), 29-39.

Moe, W., Fader, P.S. (2001). Capturing Evolving Visit Behaviour in Clickstream Data. *Wharton Marketing Department Working Paper.*

Moe, W., Fader, P.S. (2002). Dynamic Conversion Behaviour at e-Commerce Sites. *Wharton Marketing Department Working Paper.*

Montgomery, A.L. (2001). Applying quantitative marketing techniques to the internet. *Interfaces*, 31(2), 90-108.

Morrison, D.G. (1969). On the interpretation of discriminant analysis. *Journal of Marketing Research*, 6(2), 156-163.

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E., Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks*, 11(3), 690-696.

Niraj, R., Gupta, M., Narasimhan, C. (2001). Customer profitability in a supply chain. *Journal of Marketing,* 65(3), 1-16.

Padmanabhan, B., Zheng, Z., Kimbrough, S.O. (2001). Personalization from Incomplete Data: What You Don't Know Can Hurt. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Reichheld, F.F. (1996). Learning from customer defections. *Harvard Business Review,* 74(2), 56-69.

Rust, R.T., Zahorik, A.J. (1993). Customer satisfaction, customer retention, and market share. *Journal of Retailing,* 69 (2), 193-215.

Schmittlein, D.C., Peterson, R.A., (1994). Customer base analysis: an industrial purchase process application. *Marketing Science,* 13(1), 41-67.

Sismeiro, C., Bucklin, R.E. (2003). Modeling Purchase Behavior at an E-Commerce Website: A Conditional Probability Approach. *Marketing Colloquia, Wharton Marketing Department, March.*

Van den Poel, D., Larivière B. (2004), Customer attrition analysis for financial services using proportional hazard models, *European Journal of Operational Research,* 157 (1), 196-217.

Van den Poel, D., Leunis, J. (1999). Consumer acceptance of the Internet as a channel of distribution. *Journal of Business Research*, 45 (3), 249-256.

Wu, C., Chen, H.L. (2000). Counting your customers: compounding customer's in-store decisions, interpurchase time and repurchasing behaviour. *European Journal of Operational Research,* 127, 109-119.

APPENDICES

Appendix 3.A: List of predictors.

Type	Variable name	Description	Past studies	This study
			S = significant, NS = not significant	
			*** < 0,01 ; ** < 0,05 ; * < 0,1	
General clickstream				
	FrequencyVisit	Total number of past visits	S	**
	RecencyVisit	Number of days since last visit	S	***
	StdDevRecencyVisit	Standard deviation of the time between site visits		***
	MeanRecencyVisit	Average time between site visits		***
	CVRecencyVisit	Coefficient of variance of the time between site visits		**
	AverageVisitTime	The average visit time of a visit		*
	TotalVisitTime	Total past visit time	S	**
	VisitTime	The visit time of the last session	S	*
	TotalClicks	Total number of clicks in the past	S	***
	AverageTotalClickTime	The average time per click	S	NS
	AverageClickTime	The average time per click in the last session		NS
	AverageVisitclicks	The average number of clicks in a session	S	NS
	Hurry	The average time per click in the session is lower than the average		**
Detailed clickstream				
	PageProc	The number of pages viewed concerning the site procedure during the last visit		NS
	PageSup	The number of pages viewed concerning the product supply during the last visit		NS
	PageComp	The number of pages viewed concerning the company during the last visit		NS
	PageComm	The number of community pages viewed during the last visit		NS
	PageWine	The number of wine products viewed during the last visit		NS
	PageAcc	The number of accessories viewed during the last visit		**
	PageComb	The number of bundled products viewed during the last visit		**
	PageSearch	The number of times one made use of the search engine during the last visit		NS
	PageGift	The number of pages concerning gifts viewed during the last visit		NS
	PagePers	The number of personal pages viewed during the last visit		***
	PageProduct	The number of products viewed during the last visit		NS
	NumClicks	The number of clicks during the last visit		NS
	TotPageProc	The total number of viewed pages concerning the procedures of the site		NS
	TotPageSup	The total number of viewed pages concerning the procedures of product supply		NS
	TotPageComp	The total number of viewed pages concerning the company		NS
	TotPageComm	The total number of community pages viewed		*
	TotPageWine	The total number of viewed pages concerning wine products		NS
	TotPageAcc	The total number of viewed pages concerning wine accessories		NS
	TotPageComb	The total number of viewed pages concerning bundled products		***
	TotPageSearch	The total number of times one made use of the search engine of the site		*
	TotPageGift	The total number of viewed pages concerning gifts		NS

TotPagePers	The total number of personal pages that were viewed		***
TotPageProduct	The total number of products viewed	S	***
PercPageProc	The number of pages viewed concerning the site procedure during a visit divided by the total number of clicks during that visit	S	NS
PercPageSup	The number of pages viewed concerning the product supply during a visit divided by the total number of clicks during that visit	S	NS
PercPageComp	The number of pages concerning the company viewed during a visit divided by the total number of clicks during that visit		NS
PercPageComm	The number of community pages viewed during a visit divided by the total number of clicks during that visit	S	NS
PercPageWine	The number of wine products viewed during a visit divided by the total number of clicks during that visit		NS
PercPageAcc	The number of wine accessories viewed during a visit divided by the total number of clicks during that visit		NS
PercPageComb	The number of bundled products viewed during a visit divided by the total number of clicks during that visit		NS
PercPageSearch	The number of times one made use of the site search engine divided by the total number of clicks during that visit		NS
PercPagePers	The number of personal pages viewed during a visit divided by the total number of clicks during that visit		***
PageProcrVisit	The total number of page views concerning the site procedure divided by the total number of visits		NS
PageSuprVisit	The total number of page views concerning the product supply divided by the total number of visits		NS
PageComprVisit	The total number of page views concerning the company divided by the total number of visits		NS
PageCommrVisit	The total number of community pages viewed divided by the total number of visits		NS
PageWinerVisit	The total number of wine products viewed divided by the total number of visits		NS
PageAccrVisit	The total number of accessories viewed divided by the total number of visits		NS
PageCombrVisit	The total number of bundled products viewed divided by the total number of visits		NS
PageSearchrVisit	The total number of times one made use of the search engine divided by the total number of visits		NS
PageGiftrVisit	The total number of personal pages viewed divided by the total number of visits		NS
PagePersrVisit	The total number of personal pages viewed divided by the total number of visits		***
PageProductrVisit	The total number of pages viewed concerning gifts divided by the total number of visits		NS
AveragePageProc	The total number of page views concerning the site procedure divided by the total number of visits		NS
AveragePageSup	The total number of page views concerning the product supply divided by the total number of visits		NS
AveragePageComp	The total number of page views concerning the company divided by the total number of visits		NS
AveragePageComm	The total number of community pages viewed divided by the total number of visits		NS
AveragePageWine	The total number of wine products viewed divided by the total number of visits		NS
AveragePageAcc	The total number of accessories viewed divided by the total number of visits		NS
AveragePageComb	The total number of bundled products viewed divided by the total number of visits		NS
AveragePageSearch	The total number of times one made use of the search engine divided by the total number of visits		NS
AveragePageGift	The total number of pages viewed concerning gifts divided by the total number of visits		NS
AveragePagePers	The total number of personal pages viewed divided by the total number of clicks		***
AveragePageProduct	The total number of products viewed divided by the total number of visits		NS
PageProductrPur	The number of products viewed divided by the number of purchases		*
PageWinerPur	The number of wine products viewed divided by the number of purchases		NS
PageAccrPur	The number of accessories viewed divided by the number of purchases		NS
PageCombrPur	The number of bundled products viewed divided by the number of purchases		**
Customer demographics			
Gender	Gender (0= female, 1 = male)	NS	***
Age	Age of the visitor	NS	NS
LanguageD	Language of the visitor is Dutch (0/1)		NS
LanguageF	Language of the visitor is French (0/1)		NS
Trust	Did the visitor supply the company of his phone number (0/1)		***

ProfSup	Did the customer supply his profession (0= not supplied, 1= supplied)		*
GenderSup	Did the customer supply his sex (0= not supplied, 1= supplied)		***
AgeSup	Did the customer supply his age (0= not supplied, 1= supplied)		***
Historical Purchase behaviour			
Totpurchases	Total number of purchases ever did at the site	S	*
PurchasesrVisit	The number of purchases per visit		***
TotMonetary	Total spending ever at the site		NS
MonetaryrVisit	The average spending per visit		NS
Monetary	Spending during the last visit		NS
MonetaryrPur	The average spending when one did a purchase		NS
PurchaseRecency	The number of days between the visit and the last purchase		***
PurchaseLastVisit	Purchased during last site visit?		**
StdDevPurchaseRecency	Standard deviation of the number of days between a visit and the last purchase		NS
MeanPurchaseRecency	Mean of the number of days between a visit and the last purchase		***
CVPurchaseRecency	Coefficient of variation of the number of days between a visit and the last purchase		NS
CreditcardUse	Ever paid with credit card?		NS
GiftShopper	Ever bought a gift in the past?		NS

Appendix 3.B: Correlation Matrix of Important Predictors

	RecencyVisit	RecencyVisit²	Hurry	TotPageProduct	PageProductrPur	AveragePagePers	PagePers	PageAcc	Trust	Gender	Totpurchases	PurchaseRecency
RecencyVisit	1,00	0,92	0,19	-0,03	0,02	-0,04	0,00	0,07	-0,02	0,03	-0,03	0,04
RecencyVisit²	0,92	1,00	0,13	-0,05	-0,01	-0,03	-0,01	0,08	0,00	0,02	-0,04	0,09
Hurry	0,19	0,13	1,00	0,09	0,07	0,07	-0,02	0,03	-0,03	0,16	0,02	0,02
TotPageProduct	-0,03	-0,05	0,09	1,00	0,52	0,11	0,05	0,08	-0,10	0,23	0,43	-0,02
PageProductrPur	0,02	-0,01	0,07	0,52	1,00	-0,03	0,07	0,12	-0,11	0,03	-0,01	-0,24
AveragePagePers	-0,04	-0,03	0,07	0,11	-0,03	1,00	0,52	0,01	0,14	0,29	-0,15	0,35
PagePers	0,00	-0,01	-0,02	0,05	0,07	0,52	1,00	0,04	0,10	0,08	-0,07	0,12
PageAcc	0,07	0,08	0,03	0,08	0,12	0,01	0,04	1,00	-0,04	-0,04	-0,03	0,03
Trust	-0,02	0,00	-0,03	-0,10	-0,11	0,14	0,10	-0,04	1,00	0,27	0,05	0,13
Gender	0,03	0,02	0,16	0,23	0,03	0,29	0,08	-0,04	0,27	1,00	0,21	0,05
Totpurchases	-0,03	-0,04	0,02	0,43	-0,01	-0,15	-0,07	-0,03	0,05	0,21	1,00	-0,23
PurchaseRecency	0,04	0,09	0,02	-0,02	-0,24	0,35	0,12	0,03	0,13	0,05	-0,23	1,00

CHAPTER IV

CUSTOMER BASE ANALYSIS: PARTIAL DEFECTION OF BEHAVIOURALLY-LOYAL CLIENTS IN A NON-CONTRACTUAL FMCG RETAIL SETTING[6]

[6] This chapter is based on the following reference: Wouter Buckinx, Dirk Van den Poel, 2005. Customer Base Analysis: Partial Defection of Behaviourally-Loyal Clients in a Non-Contractual FMCG Retail Setting, *European Journal of Operational Research*, Vol 164(1), 252-268.

CHAPTER IV:

CUSTOMER BASE ANALYSIS: PARTIAL DEFECTION OF BEHAVIOURALLY-LOYAL CLIENTS IN A NON-CONTRACTUAL FMCG RETAIL SETTING

1. INTRODUCTION

Customers' life cycles are becoming increasingly transitory due to the severe impact of competitors' actions on existing relationships (Reinartz and Kumar, 2000). Nowadays, consumers are offered a tremendous array of choices. Some people restrict their choices, become relationship oriented (Sheth et al., 1995) and have the potential to become long-life customers. Others exhibit switching behaviour in their shopping (Peterson, 1995). Typically, customers split their purchases among several competitive companies (Dwyer, 1997). This may be due to the fact that customers do not experience any switching costs when changing their supplier (Reinartz and Kumar, 2000).

A relationship has the potential to continue only if both parties are satisfied in the normal setting where alternatives are available (Hoekstra and Huizingh, 1999). If customer satisfaction declines for some reason and a competitor is able to offer a similar product or service, the relationship is likely to be broken. Satisfaction and attractiveness of alternatives determine the strength of relationships (Anderson et al., 1994; Morgan and Hunt, 1994; Peelen et al., 1989). So customer retention is driven by customer satisfaction (as well as other drivers) if sufficient valid alternatives exist (Rust and Zahorik, 1993). Lindgreen and

Pels (2002) emphasise that this topic should be studied from a customer's as well as a supplier's perspective. Even if companies are well-equipped to offer relational interaction, some customers prefer not to engage in relationships, i.e. they opt for the 'transactional' exchange as opposed to the 'relational' exchange.

A non–contractual setting suffers from the problem that customers have the opportunity to continuously change their purchase behaviour without informing the company about it. More specifically, in a grocery retail environment (the setting of this study) competition is severe and customers have a wide array of alternatives. This is illustrated by AC Nielsen's (2001) report that more than 70 per cent of all customers shop around in several supermarkets during a month.

Profits can increase because of several reasons (Reichheld, 1996). First of all, by implementing retention programmes, customers are confronted with increasing switching costs, giving them fewer incentives to change their current behaviour (Jones et al., 2000). Secondly, the length of customer relationships influences a firm's profitability. The longer a customer stays the more he spends at the company. Buyers tend to purchase additional services (products) and are more likely to convince others about the positive value the company offers (word-of-mouth effect). They tend to be less price sensitive (Zeithaml et al., 1996) and exhibit a lower responsiveness to competitive pull (Stum and Thiry, 1991).

Retained customers produce higher revenues and margin than new customers (Reichheld and Sasser, 1990). The net return on investments for retention strategies is higher than for acquisitions. So it is supported that companies first spend their marketing resources to keep existing customers rather than to attract new ones (Rust and Zahorik, 1993; Mozer et al., 2000). Recently, however, the argument that customers who purchase steadily from a company over time are necessarily cheaper to serve (or less price sensitive) has drawn substantial criticism (Reinartz and Kumar, 2002).

In summary, customer retention is a valuable strategy to ensure long-term profitability and success of the company. This is illustrated in Table 1. Reducing customer defection can have an enormous impact on companies' results (Mozer et al., 2000; Van den Poel and Larivière, 2003). Suppose 25 per cent of the top clients defect. Considering an average contribution of 2 000 Euro a year and a discount rate of five per cent, an improvement of the retention rate by just one percentage point will cause an increase in profits by 102 923 Euro over five years per 1 000 clients (see last column of Table 1).

Table 4.1: Profit implications.

Retention rate	*Number of customers left*				*Total contribution over 5 years (in Euro)*	*Additional contribution over 75% (in Euro)*
	Year 1	Year 2	…	Year 5		
75%	1 000	750	…	316	5 679 709	0
76%	1 000	760	…	333	5 782 632	102 923 Euro

On top of the lost sales new customers need to be attracted, which requires very costly actions (Athanassopoulos, 2000; Colgate et al., 1996). Advertising efforts as well as promotions and sales costs are significant but necessary expenses to fill up the customer base (Zeithaml et al., 1996) and establish new relationships (Athanssapoulos, 2000). Besides, new clients often are not profitable for some time.

Moreover, defecting (dissatisfied) customers are convinced that the company offers inferior value and might persuade other customers by spreading negative word-of-mouth (Reichheld, 1996; Sonnenberg, 1990; Mizerski, 1982).

In conclusion, retaining customers by avoiding defection is an important issue for marketing/CRM managers. A first step in addressing this issue is finding out who to target in retention actions. *A fortiori*, this is an underresearched topic in the fast-moving consumer goods retail sector. One possible answer is those customers who are most likely to partially attrite. Therefore, in this paper, we investigate whether we are able to predict, at the level of the individual customer, who is going to partially defect. More specifically, we want to find out which of the currently behaviourally-loyal customers are likely to (partially) churn in the future. Moreover, we want to gain insight into which predictors are important in identifying partial defection.

This paper is organised as follows. Section 2 presents an overview of the existing body of literature about churn analysis. Section 3 specifies the methodology including three classification techniques used in this study. The description of the data set, as well as an overview of the attributes used to predict customer attrition is discussed in Section 4. Section

5 presents the results and Section 6 phrases the conclusions. Section 7 and 8 end this paper with a discussion and limitations of this research.

2. DEFECTION OF BEHAVIOURALLY-LOYAL CUSTOMERS: LITERATURE REVIEW

The topic of customer defection has been discussed extensively in recent literature (cf Table 2). Churn analysis typically tries to define predictors of customer defection. In all of the cases, however, switching behaviour is defined as *total* defection. Customers close their accounts (banking) or change their (mobile) phone operator (telecommunications). In these industries it is easy to observe when defection occurs: people totally interrupt their relationship with the company. As these companies are in a contractual setting, they are able to determine the exact point in time when clients interrupt their relationship. In other sectors it is more complex to determine when customers are leaving. However, buyers typically do not defect from the company all of a sudden. They switch some of their purchases to another store, i.e. they exhibit partial defection. There is a real danger that after a while they will switch completely to the competitor. So in the long run partial defection may lead to total defection.

Table 2 reveals that the churn issue has been underresearched in the retail sector. Moreover, all analyses consider total defection. To discover partial defection this study uses company-internal customer data to determine changes in the individual transaction pattern. We may, for example, hypothesise that customers staying true to their existing patterns are likely to stay, whereas deviations in transaction patterns may signal (partial) defection.

Efforts do not need to be made for the entire customer base. Some customers are not worth the effort to develop a long-term relationship (Hoekstra and Huizingh, 1999). Strategies should be in line with the relationship potential of each customer individually (Reichheld, 1996). It is a well-known phenomenon that a small percentage of customers accounts for a large percentage of profits (Niraj et al., 2001). Moreover, a significant part of the customer base is even not profitable. A small example might illustrate these statements. Imagine a company confronted with a defection rate of 25 per cent. In order to set appropriate marketing strategies, they want to discover why customers defect. A churn analysis for their

Table 4.2: Literature review

	Sector							*Defection*		*Customer base*	
	Retail	Finance	Telecom	Computer manufacturer	Insurance	Automotive firm	Other service	Total	Partial	Complete	Partial
Athanassopoulos (2000)		x						x		x	
Bhattacharya (1998)								x		x	
Keaveney and Parthasarathy (2001)							x	x		x	
Lemon et al. (2002)							x	x		x	
Mittal and Kamakura (2001)						x		x		x	
Mozer et al. (2000)			x				x	x		x	
Popkowski et al. (2000)								x		x	
Van den Poel and Larivière (2003)		x						x		x	
Weerhandi and Moitra (1995)			x					x		x	
Zeithaml et al. (1996)	x			x	x			x		x	
This study	**x**								**x**		**x**

entire customer base shows that people leave because of the absence of fast checkouts (e.g., cash registers only available to customers who bought less than ten products). Subsequently, the company decides to invest in such a costly service so more cashiers need to be present at the same moment. However, their most profitable clients are not served with this measure because they typically have more products in their baskets. So only the less profitable customers are satisfied, resulting in a decline of the defection rate, but not necessarily in an increase in profit. In this case, management addressed a reason of customer defection for the unprofitable part of the customer base.

Therefore it is suggested to only focus on those customers in the client base whose future contribution looks promising (Ganesh et al., 2000). Table 2 (last column) reveals that no prior research focused only on the most relevant part of the customer base (in terms of profitability). Instead, they considered all clients.

3. METHODOLOGY

3.1 Behaviourally-Loyal Clients

As argued in the previous section, we do not focus on the entire customer base. We only select the best customers of the company. The core of a valuable customer base consists of loyal customers (Ganesh et al., 2000). Loyal customers are more profitable in the short run as well as in the long run (O'Brien and Jones, 1995). They ensure a continuous stream of profits. In our case we focus our study on those who shop frequently and at the same time exhibit a regular buying pattern. To define that segment of clients we use two behavioural attributes: the frequency of purchases and the time between their purchases (interpurchase time or IPT). Both variables are commonly used to define good customers (O'Brien and Jones, 1995). More specifically, the customers in our segment of attention satisfy the following conditions:

(1) Frequency of purchases is above average.
(2) Ratio of the standard deviation of the interpurchase time to the mean interpurchase time is below average.

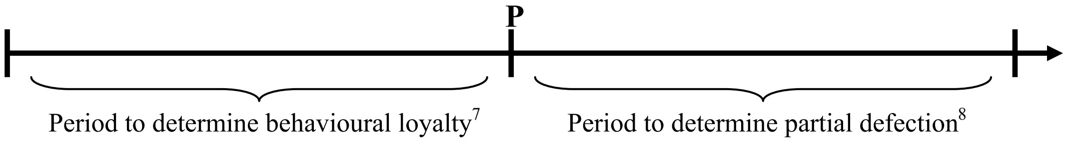
The first criterion provides an indication of a customer's loyalty (Wu and Chen, 2000) and potential profitability. The second attribute ensures that the time between customer visits is regular. To identify behaviourally-loyal customers, we do not take into account any monetary condition. This is to avoid missing those buyers who do not yet belong to the segment of currently profitable customers but do have a high *potential* value (Niraj et al., 2001).

3.2 Partial Defectors

One of the deliverables of this research is an individual-level prediction of the probability to partially defect in the future. In other words, at some specific point in time we want to determine which behaviourally-loyal clients in our database may partially switch their purchases to another store (as indicated by "P" in Figure 1). So, ultimately, for each individual we need to make an unambiguous conclusion about his future behaviour. As a result, the models we build will be all binary classification models where the dependent variable classifies a particular customer either as a partial defector or as a customer continuing his loyal buying pattern.

However, in a non-contractual setting it is not clear when people defect. Therefore, it is very important to clearly define the concept of partial defection. To this end, we again take into consideration both conditions of the previous paragraph that are used to define our segment of interest but this time over a period of observation *after* the period used to determine behavioural loyalty (i.e., after point “P” in Figure 1). So, if one of the abovementioned conditions (1) or (2) is not fulfilled, we classify a customer as partially defective (as the dependent variable) because he deviates from his established transaction pattern.

Figure 4.1: Period of observation



3.3 Classification Techniques

The problem of separating behaviourally-loyal customers from behaviourally non-loyal clients may be solved by any classification technique. In this section we discuss the three techniques we use for this task.

3.3.1 Logistic Regression

Logistic regression modeling is a well-known technique. It is very appealing because: (1) A closed-form solution for the posterior probabilities is available (as opposed to probit); (2) The basic assumption of logit (the logarithm of the ratio of group-conditional densities is linear in the parameters) is satisfied by many families of distributions (Anderson, 1982); (3) It is easy to use and provides quick and robust results.

In this study we include the technique as a benchmark to compare the more advanced techniques against. We refer to other texts for more technical details (Anderson, 1982).

[7] This period of five months (from April 2000 until August 2000) was also used to derive the independent variables of the model (See empirical Study).

[8] This period of five months (from September 2000 until January 2001) was used to derive the dependent variable.

3.3.2 Automatic Relevance Determination (ARD) Neural Network

Artificial neural networks are often credited for achieving higher predictive performance compared to other (statistical) classification techniques (Baesens et al., 2002; Viaene et al., 2001). Within the broad group of neural network architectures we select MacKay's Bayesian ARD neural network framework because it has the appealing property of providing a Bayesian hyperparameter per input variable, representing the importance of the variable (MacKay, 1992). More specifically, we use Nabney's (2001) MATLAB implementation for ARD neural networks. When fixing the number of hidden units, we take into account Penny and Roberts' (1999) recommendation to use a sufficiently large number of hidden units to ensure obtaining a reliable estimate of the predictors' importance.

3.3.3 Random Forests

Decision trees have become very popular for solving classification tasks because they can deal with predictors measured at different measurement levels (including nominal variables) and because of their ease of use and interpretability (Duda et al., 2001, Chapter 8). However, they also have their disadvantages such as lack of robustness and suboptimal performance (Dudoit et al., 2002). Recently, many of these disadvantages have been dealt with by creating an ensemble of trees and letting them vote for the most popular class, labelled forests (Breiman, 2001). Several successful paths have been explored how to grow ensembles of trees: (1) Bagging, where to grow each tree a random selection (without replacement) is made from the examples in the training set (Breiman, 1996); (2) Random split selection, where at each node the split is selected at random from among the K best splits (Dietterich, 2000); and (3) Random subspace method, which does a random selection of a subset of predictors to grow each tree. In this paper we select the random forests as proposed in Breiman (2001) which uses the latter strategy. An interesting by-product of these ensembles of trees is their importance measures for each variable. The only two parameters a user of the technique has to determine are the number of trees to be used and the number of variables to be randomly selected from the available set of variables. In both cases we follow Breiman's recommendation to pick a large number (5 000 in this case) for the number of trees to be used, as well as the square root of the number of variables for the latter parameter.

3.4 Evaluation Criteria

In order to evaluate the performance of classification techniques we use two criteria: percentage correctly classified (PCC) and area under the receiver operating characteristic curve (AUC). Both measures are commonly used as performance criteria (Mozer et al., 2000, Zhang et al., 2002, Chawla et al., 2002). The PCC compares the 'a posteriori' probability of defection with the true status of the customer. The resulting confusion matrix is used to calculate the accuracy of the models. A disadvantage of this measure is that it is not very robust concerning the chosen cut off value in the 'a posteriori' probabilities (Baesens et al., 2002). The AUC measure takes into account all possible cut off levels. For all these points, it considers the sensitivity (the number of true positives versus the total number of defectors) and the specificity (the number of true negatives versus the total number of non-defectors) of the confusion matrix in a two-dimensional graph, resulting in a ROC curve. The area under this curve can be used to evaluate the predictive accuracy of classification models.

4. EMPIRICAL STUDY

4.1 General

For our empirical analysis, one of the largest retailers with worldwide operations offering fast moving consumer goods (FMCG) provided the necessary data. Different purchase occasions could be traced by means of a loyalty card. We refer to Ziliani (2000) for an overview of alternative micro-marketing (which also comprises CRM) strategies using loyalty-card data. Over 85 per cent of purchases at this particular retailer are registered by their loyalty card. Specifically, we used individual records of 158 884 customers from April 2000 until January 2001, which represented a random sample from the entire customer base containing millions of customers within one geographic area. Even though a five-month period may seem short, we believe it is adequate since we are dealing with an FMCG retailer with an average interpurchase time of 12 days, which results in an average visit rate of 30 times a year.

The first five-month period of the available data, from April until August, is used to define the retailers' behaviourally-loyal customers (see Figure 1). Consequently, we select 32 371 customers, which we consider to be behaviourally-loyal clients. This is 20.37 per cent of the total available customer base. These behaviourally-loyal clients visit the retailer each week, which means that their average interpurchase time is only seven days (compared to 12 days for the total customer base). Besides, their spending is a lot higher. The average spending of a customer is 1 417 Euro a year, whereas the behaviourally-loyal customers spend almost twice as much: 2 832 Euro. We randomly separated this group of customers in a training set (16 079 observations) and a test set (16 292 observations). The same procedure is used to determine whether they defected during the subsequent period of five months (from September until January). Applying our partial-defection definition results in 8 140 partial defections. This is 25.15 per cent of the clients under investigation.

4.2 Predictors

The available data consist of behavioural information at the level of the individual customer and customer demographics. Prior research already supports the incorporation of these two groups of predictors. Table 3 reveals that a major part of the existing attrition studies focuses on demographics as antecedents of defection.

Using the observed past purchase behaviour and additional customer information we compile 61 variables to predict (partial) churn behaviour. These variables have the advantage of being widely available and have shown to be effective and rich predictors (Schmittlein and Peterson, 1994; Buckinx et al, 2004). Table 4 summarises all behavioural independent variables supported by former research. The number of purchases (Frequency) and the amount of spending (Monetary) are the most popular predictors in other research. The time of the day (of purchase or consumption), the length of the customer-supplier relationship (LoR), buying behaviour across categories (Category), mode of payment (MoP), usage of promotions and brand purchase behaviour are variables rarely used in past research. Our study, however, will take them into account.

The following paragraphs provide a motivation for including each of these variables. An overview of all variables used in this study can be found in Table 5.

Table 4.3: Predictors of defection in prior research

	Behavioural antecedents	*Demographics*	*Perceptions*
Athanassopoulos (2000)		x	x
Bhattacharya (1998)	x	x	
Keaveney and Parthasarathy (2001)	x	x	x
Lemon et al. (2002)	x	x	x
Mittal and Kamakura (2001)		x	x
Mozer et al. (2000)	x	x	
Popkowski et al. (2000)	x	x	
Weerhandi and Moitra (1995)		x	
Zeithaml et al. (1996)			x
This study	**x**	**x**	

Table 4.4: Behavioural predictors of defection in prior research

	Recency	*Frequency*	*Monetary*	*Timing*	*Lor*	*Category*	*Mop*	*complaints*	*credit*
Bhattacharya (1998)	x	x	x		x				
Keaveney and Parthasarathy (2001)		x	x						
Lemon et al. (2002)		x							
Mozer et al. (2000)		x	x	x	x	x	x	x	x
Popkowski et al. (2000)			x						
This study	**x**	**x**	**x**	**x**	**x**	**x**	**x**		**x**

4.2.1 Interpurchase Time and Related Inputs

We include several variables that are related to the time between customers' shop incidences. First, we include "Recency", which represents the number of days that passed between the last transaction and the end of our observation period. Customers who recently purchased are more likely to be active than customers who shopped a long time ago (Wu and Chen, 2000). Most previous studies find that the lower the value of recency, the higher the probability that a customer stays loyal. In a non-contractual setting this can be the most important variable to indicate an active or inactive relationship (Reinartz and Kumar, 2000). Secondly, the average interpurchase time (IPT), the standard deviation, and the coefficient of variation (ratio of the standard deviation to the average) are incorporated. The average IPT reflects the recency variable over the entire time period. The standard deviation of the IPT and the coefficient of variation measure the irregularity of the time between purchases. We hypothesise the more irregular the less loyal a customer will be.

Table 4.5: Predictors used in this study.

Variable Type	*Variable Name*	*Description*
Interpurchase time	Recency	Number of days since last shop incidence
	MeanRecency	The average number of days between a customers' shop incidences (IPT)
	StdDevRecency	Standard deviation of the IPT
	CVRecency	Coefficient of variation of recency, i.e., ratio of StdDevRecency to MeanRecency
Frequency	Frequency	Number of shop visits (with purchase)
	rFrequency	Number of shop visits relative to the length of relationship (LoR)
	FreqLastMonth	Number of shop visits during last month
	FreqLastWeek	Number of shop visits during last week
Monetary	Monetary	Total monetary amount of spending
	rMonetary	Total spending relative to the length of relationship (Lor)
	rMajorTrip	Percentage of shop visits with above-average spending
Category	rCat (1-12)	Aggregated relative spending in 12 different categories: prepared meals, chemist's, drinks, food, fruit & vegetables, dairy products, meat, non-food, fish, bakery, wine & alcohol, and self-catering.
	Cat 1	Aggregated spending in the self-catering category
	NoCat	Number of categories ever purchased from
Brand	NatBrand	Aggregated relative national brand purchase behaviour
	RetBrand	Aggregated relative retailer's brand purchase behaviour
	LowBrand	Aggregated relative low budget brand purchase behaviour
Length of Relationship	LoR	Number of days since first purchase
Timing	MeanTimeOfDay	Average moment in time of shopping
	StdDevTimeOfDay	Standard deviation of Meantime
	LastTimeOfDay	Time moment of last store visit
Mode of Payment	rMoP (1-6)	Aggregated relative amount of money paid in six different ways: 1. cash, 2. check, 3. lunch-allowance check, 4. in-house vouchers, 5. debit card and 6. credit card
	MoP (1-3)	Aggregated amount of money paid in three of the six different ways: 1. cash, 2. lunch-allowance check, and 3. credit card
	rRetBottles	Aggregated relative value of returned bottles
	RetBottles	Aggregated value of returned bottles
Promotions	FreqPromo	Number of shop incidences coupon used
	NoVisitsLastCoupon	Number of visits since a coupon was used for the last time
	MeanMonCoupon	Average monetary value of coupons (per shopping trip)
	LoyPoints	Number of loyalty points earned because of special product purchase
Demographics	hhs(1-4)	Householdsize: Number of members in the household
	Language	Language (labels a different language group)
	Title (1-2)	Title of the person
	RegionCode (1-6)	Postal code region classification
	Pets	Presence of pet(s): no (0) / yes (1)
	DemoMissing	Dummy indicating whether or not demographic information is missing

4.2.2 Frequency of Purchases

The customer's frequency of purchases may be predictive for their future behaviour (Schmittlein and Peterson, 1994) because it is positively related to customers' expected future use (Lemon et al., 2002). The probability that a customer is alive may be measured by the number of purchases (Reinartz and Kumar, 2000). Again, we propose several alternative operationalisations of this type of variable. "Frequency" is the number of shop visits. Moreover, we use the number of days that a person is already a customer at the retailer to include a 'relative' version of the frequency variable. "FreqLastMonth" and "FreqLastWeek" represent the frequency of purchases during the last month and last week of data respectively. Both variables are included because variables computed over more recent time periods may be (more) important to include as predictors.

4.2.3 Monetary Indicators

These indicators represent the amount of money someone has spent at a company. The monetary value of each customer's past purchase behaviour tends to be effective in predicting purchase patterns (Schmittlein and Peterson, 1994) and is used in the literature to determine future patterns. Mozer et al. (2000) included monthly charges and usage to predict subscriber dissatisfaction and improve their retention rate. We incorporate three monetary indicators: 'Monetary' is the accumulated amount of money spent from April until September, 'rMonetary' is the same as 'Monetary' but takes into account the length of the relationship of a customer with the retailer, and 'rMajorTrip' indicates the percentage of purchases that could be classified as a big shopping incidence.

4.2.4 Shopping Behaviour Across Product Categories

Defection may occur when customers are not pleased (anymore) with a specific product or service (Mozer et al., 2000; Rust and Zahorik, 1993; Mittal and Lassar, 1998). Possible explanations are that prices are too high or quality of the product or service decreases (compared with competitors). If indeed the price or quality of a (category of) product(s) deteriorates and someone intensively purchases this product (category), the probability of defection increases. Consequently, we include inputs representing the spending in each category of the retailer. Literature supports the use of categorical behaviour (Athanassopoulos, 2000). Verstraeten et al. (2002) found preliminary evidence for the existence of a ‘natural’ order of product purchases. Customers may start their relationship with the retailer by buying specific products. The start of buying specific products or products from certain categories may be the indicator of a changing loyalty towards the company.

The retailer’s product-category taxonomy consists of 12 main categories. If numerous customers defect because of the use of a specific category, our model may indicate that the category-spending variable is a predictor of partial defection.

Besides the monetary version we compute the total number of different categories someone purchases from (NoCat). The number of active products/services might be linked to defection (Mozer et al., 2000). The higher this number the more active someone is.

4.2.5 Brand Purchase Behaviour

The retailer classifies each product into a brand category: national brand, retailer’s own store brand, a private label brand, or a (temporal) exclusive brand. For each of these brands a variable is compiled, representing the relative spending of a customer. First, the arguments we used to support the incorporation of the variables summarising their shopping behaviour by category (cf previous paragraph) can be repeated here. If a significant part of the retailer’s top clients defect because of a problem with some brand, the model may indicate that the brand-spending variable is powerful to predict defection. Consequently, management is able to define tailor-made actions. Secondly, concerning the private label/store brand, it is known that qualitative retailer brands can be a tool to differentiate a store and increase store loyalty (Corstjens and Lal, 2000). So we hypothesise that the higher the spending for the store brand/private label brand of the store, the lower the probability that the consumer will leave.

4.2.6 Length of Relationship

Length of relationship represents the number of days an individual is shopping at the retailer. Bhattacharya (1998) found that the extent to which a customer is able to identify himself with a company is positively related to the period he is willing to continue this relationship. Anderson and Weitz (1989) confirmed this expectation and indicated that the length of the relationship is positively associated to the perceived future stability of the relationship. Verhoef et al.'s (2002) findings confirm the impact of age of relationship on number of services purchased in an insurance context.

4.2.7 Timing of Shopping

People do not shop all at the same time during the day or week. This may lead to service quality differences across several moments of the day. For example, employees may be significantly friendlier at noon because in the morning they suffer from morning mood and in the evening they are very busy because the store is too crowded. Under this assumption, people shopping at noon may experience a higher level of service quality than people shopping at other moments. As a result we include a variable representing the average of all points in time when a customer left the shop (check-out time).

4.2.8 Mode of Payment

Customers are offered several possible ways to pay their bill. The use of each of these modes of payment might be useful to classify customers into different segments and consequently might be a predictor for future behaviour. The different modes of payment are: cash, checks, lunch-allowance checks, in-house vouchers, electronic payment and credit cards. The in-house vouchers are distributed by the retailer to reward customers for their loyalty based on the information collected by customer loyalty cards. For example, the intensive use of these vouchers might be predictive for upcoming loyalty. The possession and use of a credit card may indicate that customers like to make use of credit. Literature confirms the use of credit information and rate plans for churn analysis (Mozer et al., 2000). An additional variable in this context is the amount of money subtracted from the bill because of returned empty bottles. People returning their empty bottles to a shop show loyalty and consistency towards the retailer.

4.2.9 Promotional Behaviour

Prior literature supports that the degree of competition between stores has increased over time. Due to the increased merchandising and promotional activities of retailers consumers are trained to compare deals across competitors (Kim and Staelin, 1999). Moreover, Bawa and Shoemaker (1987) proved that customers being deal-prone are less brand loyal and less store loyal. For them, the lower prices are the explanation of their purchases. These customers typically do not develop a relationship with one specific company. Consequently, we hypothesise that people being more sensitive to promotions will have a higher probability of store switching and thus defection.

4.2.10 Customer Demographics

Table 3 indicates the extensive use of customer demographics in other studies of customer defection. Mittal and Kamakura (2001) show that among other things, gender, number of children in a household as well as area of residence are moderating customer characteristics. Vakratsas (1998) confirmed the moderating role of household size: small households are more deal prone than larger-size households (Buckinx et al., 2004). So we expect these clients to be less loyal to the retailer. Mozer et al. (2000) included an indication of the subscriber's location.

Consequently, we incorporate several demographical predictors available in the retailer's data warehouse: "hhs1"-"hhs4" are dummies in order to indicate that a household exists out of one to four or more members respectively (0/1). Secondly, '"Title1" and "Title2" indicate the title of the person who subscribed for the loyalty card of the retailer. "Language" is a dummy representing the mother language of the household. The dummies "RegionCode1"-"RegionCode6" contain geographical information of the customer and finally "Pets" makes a distinction between people having one or more pets at home and people without a pet.

For ten percent of the customers (3 288) these demographics were not available. Consequently, a dummy "DemoMissing" is added in order to take this into account. At the same time, this variable may be an indication of the level of trust in the company because giving personal information to a firm may be an indication of involvement and confidence.

5. RESULTS

Results presented in Table 6 lead us to conclude that predicting partial defection of behaviourally-loyal customers is a viable strategy: First, PCC performance of 0.8040 for random forests on a test sample (i.e. on cases not used during estimation) should be benchmarked to Morrison's (1969) proportional chance criterion[3] of 0.6235 (= $0.2515^2 + (1-0.2515)^2$) or the majority prediction rule of 0.7485 (= 1 - 0.2515); and second, AUC performance of 0.8310 (again for random forests on the test sample) exceeds the 0.5 benchmark of the null model.

Table 4.6: Performance results.

	PCC		*AUC*	
	train	*test*	*train*	*test*
Logistic regression	0.7999	0.8013	0.8278	0.8280
ARD NN	0.8083	0.8040	0.8394	0.8310
Random forests	0.8001	0.8040	0.8249	0.8319

When comparing the different classification techniques they all offer similar performance. Even though random forests consistently come in on top (without the need to tune different parameters, as was the case for ARD neural networks), its performance is not significantly higher than that of the other techniques. Given the recent nature of random forests, we would like to emphasise the attractiveness of this technique for several reasons: 1. Consistent high performance; 2. We confirm Breiman's (2001) observation that the performance results are very robust such that there is not really a need for splitting the sample into an estimation and test sample (similar to logistic regression but unlike neural networks); 3. No need to tune parameters (with the exception of setting the number of trees and the number of variables to be randomly selected from the total set of predictors); 4. Easy computation of variable importance measures; and 5. Reasonable computing times (if logistic regression serves as a reference, random forests are 300 times more 'expensive', which still compares favorably to the 90 000 times more 'expensive' ARD neural networks).

In Table 7 we report the average normalised importance of the 55 most important predictors for the Random Forests method (Breiman, 2001). When comparing the importance measures of the predictors, a Pearson (Spearman) correlation coefficient of -0.345 (-0.313)[4] between

Table 4.7: Importance of variables

No.	*Random Forests*		*ARD Neural Network*	
	AvgNormImp	*Name of Variable*	*Variance*	*Name of Variable*
1	0.99394	Frequency	7.43	Frequency
2	0.86378	MeanRecency	10.65	rFrequency
3	0.82147	rFrequency	19.76	MeanRecency
4	0.74515	LoR	22.18	FreqLastWeek
5	0.67258	FreqLastMonth	34.19	Monetary
6	0.67179	StdDevRecency	44.78	FreqLastMonth
7	0.61325	Monetary	46.31	rMoP2
8	0.56375	rMonetary	54.91	StdDevRecency
9	0.44454	rMajorTrip	59.79	hhs4
10	0.41757	DemoMissing	63.58	Title2
11	0.37740	CVRecency	69.60	LoR
12	0.32867	MeanMonCoupon	77.23	RegionCode6
13	0.31931	Recency	77.65	pets
14	0.30720	rRetBottles	90.41	DemoMissing
15	0.30140	rMoP1	91.75	MoP3
16	0.29828	NatBrand	92.62	rMonetary
17	0.28375	LastTimeOfDay	107.68	rMoP1
18	0.28134	RetBottles	125.55	Title1
19	0.27849	rCat5	127.45	rCat5
20	0.27821	rMoP5	127.57	CVRecency
21	0.27762	rCat1	128.01	rCat2
22	0.27697	rCat2	140.97	RetBottles
23	0.27234	rCat4	146.83	rMoP3
24	0.27167	rCat3	154.23	MeanMonCoupon
25	0.27005	FreqLastWeek	161.17	Recency
26	0.26011	FreqPromo	163.75	RegionCode1
27	0.25709	RetBrand	169.06	hhs2
28	0.25156	LowBrand	174.32	hhs3
29	0.24946	StdDevTimeOfDay	178.45	LastTimeOfDay
30	0.24301	rMoP6	181.00	Cat1
31	0.24226	rCat9	184.23	rMoP6
32	0.23945	rCat10	185.22	rCat4
33	0.23699	rMoP4	192.02	MoP6
34	0.23070	MeanTimeOfDay	194.83	Language
35	0.23057	rCat8	207.85	RegionCode4
36	0.22004	rCat6	211.36	rMoP5
37	0.20848	MoP6	212.20	hhs1
38	0.20727	rMoP3	229.98	RegionCode3
39	0.20334	NoCat	232.07	rRetBottles
40	0.18849	LoyPoints	239.69	FreqPromo
41	0.18286	rCat7	243.18	rCat9
42	0.17623	NoVisitsLastCoup	256.24	NatBrand
43	0.16442	MoP3	265.92	rCat3
44	0.15445	Cat1	270.60	NoCat
45	0.14548	rMoP2	271.67	rMajorTrip
46	0.12864	RegionCode2	292.72	MeanTimeOfDay
47	0.11382	RegionCode4	298.97	rCat1
48	0.11201	RegionCode6	318.03	rCat10
49	0.11173	RegionCode3	338.39	RegionCode5
50	0.10782	Title2	351.97	LoyPoints
51	0.09840	hhs1	395.68	rMoP4
52	0.09252	Language	406.36	rCat8
53	0.09050	RegionCode5	422.38	NoVisitsLastCoup
54	0.08219	RegionCode1	440.58	rCat7
55	0.07765	hhs4	451.59	rCat6

the random forest and the ARD neural network is obtained. The similarity in the ranking of the importances is confirmed by the fact that six of the top-ten variables are the same. We do not report any measures for logistic regression (e.g. standardised estimates) because most measures are prone to multicollinearity, which was clearly present in the dataset, but which is not a problem if the focus is mainly on prediction.

It is clear from the rankings of variable importance that behavioural variables are much more important than demographics. Nevertheless, the latter category cannot be ignored. A model only using behavioural variables (i.e. excluding demographics) results in an AUC of 0.8224 as compared to 0.8319 (see Table 6) in the case of random forests on the test sample. Even though this difference may seem small, it may still translate into a significant impact on the company's profits (cf Table 1). It is remarkable that the most important demographics variable is actually 'DemoMissing'. It gives empirical support to the conclusion that a behaviourally-loyal customer who is not willing to give personal information to the firm may signal future partial defection.

Moreover, within the group of behavioural variables, we find RFM (Recency, Frequency, and Monetary) variables to be the best predictors for separating behaviourally-loyal customers from non/less-loyal clients. RFM variables are well-known predictors from the field of direct marketing (Baesens et al., 2002; Van den Poel, 2003). Nevertheless, other 'signals' of loyalty are similarly important, such as the length of relationship (LoR), as well as returning empty bottles (RetBottles, rRetBottles). On the other hand, the purchase of retailer brands (RetBrand), as well as the number of categories (NoCat) and the number of loyalty points (LoyPoints) are not important in predicting partial churn.

6. CONCLUSIONS

Our empirical results show that classification models can provide an individual's (partial) defection probability given all the individual data collected by the retailer (behavioural as well as customer demographics). Consequently, we are able to track down future (partial) defectors. For managers this classification is very useful in order to establish new marketing strategies towards the companies' clients.

Moreover, we are capable of tracking down partial defection in contrast with past research that focused on total defection. This contribution is substantial for several reasons. First of all, since we consider only behaviourally-loyal clients, the losses in terms of sales may be significant even if customers defect only partially. The average spending of a behaviourally-loyal client is 2 832 Euro a year. Even if these clients switch only ten percent of their expenditure to another store, the effect on turnover is remarkable. So avoiding this switching behaviour is valuable for the retailer (see Table 1: Additional contribution calculation). Secondly, partial defection can escalate and possibly lead to total defection in the long run. Therefore, being able to signal partial defection as early as possible will result in important returns and may even be of greater importance than predicting total defection. Consequently, marketing managers can define which of their customers do have a significant chance to decrease their loyal behaviour towards the company. So they are able to execute specific marketing actions to these clients in order to prevent them from leaving.

The predictive performance of the different classification techniques is very close both in terms of the area under the receiver operating characteristic curve (AUC), as well as for the percentage correctly classified (PCC).

We may conclude that, compared to customer demographics, RFM (behavioural) variables are better in separating behaviourally-loyal customers from those who have a tendency to (partially) defect. This is somewhat in contrast to the expectations we formulated based on existing research, which strongly emphasises the explanatory/predictive power of the demographic variables.

7. DISCUSSION

This attrition research is carried out in a non-contractual setting. This environment suffers from the fact that customers can continuously switch between competitors without feedback to the original company. As a result, it is very hard to define the exact moment in time when clients leave the company. This paper, however, solves the problem by introducing the aspect of ‘partial’ defection. Customers are considered to break their relationship when they interrupt their loyal and stable purchasing pattern that they exhibit during a period of five months. Moreover, this paper contributes to the literature by making use of actual customer behaviour instead of intentions of repurchase. Lemon et al. (2002) and Morwitz et al. (1993) confirm the fact that directly observing the (defective) behaviour reveals greater insights.

This study contributes to the literature by not focusing on the entire customer base. Not all clients deserve to be taken into consideration when establishing a retention programme. This can be illustrated by a quote from Blattberg et al. (2000, p. 70): 'the goal of customer retention management is not zero defections. Instead a firm should manage its retention rate and choose retention strategies and tactics that best support its main focus: optimizing customer equity'. Accordingly, this paper only targets customers whose future contribution looks promising. The companies' targets need to be economically valuable so the increase in tenure should be achieved at a lower cost than the enhancement in customer value (Carroll, 1993). Consequently, behaviourally-loyal clients were selected from a retailer in fast-mover consumer goods. The frequency of purchase as well as the time between purchases are used to distinguish promising shoppers from others. Both variables give an indication of customers' purchasing pattern in terms of occurrence and regularity.

In this paper we focus on identifying partial defectors. However, additional research is required to investigate the actual reasons of the defective behaviour before defining the content of the retention strategy. In other words, the people classified as future defectors can be used to compose focus groups and conduct one-on-one interviews to determine which attributes most determine satisfaction (Rust and Zahorik, 1993).

Once the causes of defection and appropriate strategies are defined, companies still face the complex problem of effective allocation of resources (Rust and Zahorik, 1993). Even knowing what specific steps must be taken, it is hard to determine how much money to spend in order to increase the retention rate and at the same time increase the firm's profitability. Bolton (1998) argues that each method of assessing investments designed to increase retention should take into account the effect of changes on duration lifetimes and lifetime revenues. Mozer et al. (2000) confirm that incentives should be offered to those clients whose probability is above a certain threshold. The threshold should be computed based on the expected savings, the time horizon of evaluation, and the costs of the incentive(s). So, adapted communication actions are needed for different profiles of behaviourally-loyal clients according to their spending and their defection probability. Fortunately, our models can produce these defection probabilities. The only element we are missing to compute the expected savings is the impact of the appropriate marketing actions. Therefore, a real-life experiment with different level actions for future potential partial defectors might be a good follow-up study. This would offer information on the impact of several actions for different levels of defection probabilities.

8. Limitations and Issues for Further Research

This study has several limitations. First of all, results are confined to the retail fast-mover consumer goods (FMCG) sector. To some extent generalisations can be made for all other companies active in a non-contractual setting where defection is difficult to detect.

Demographics as well as past purchase behaviour were used as inputs in the models, based on data from a company-internal data warehouse. However, this predictor list can be extended with customer perceptions in order to increase the performance of the models. Regrettably, this type of data are typically unavailable in data warehouses. Recently, Bloemer et al. (2003) show that customer satisfaction data can provide useful insights into identifying customers ‘at risk’. Even though this fact limits our ability to gain theoretical insight into customer behaviour processes, it can be anticipated that obtaining these data by sending out questionnaires would be a very laborious and expensive exercise (the more so for a database containing millions of customers). Moreover, we anticipate that including these variables would not necessarily improve our predictive capability and would introduce other problems such as non-response bias. Therefore, we leave this as an issue for further research.

We used five months of available data to determine the focus group of the study and five months to evaluate partial defectors. It is unclear to what extent this time window restriction affects our conclusions. Whenever more data are available, more space is left to change the time window. Moreover, we would be able in that case to evaluate the defective behaviour over a longer time period. This will give the opportunity to check what happens after a while to people classified as partial defectors. That way, the expected lifetime value of a customer can be verified more precisely and appropriate actions can be better established. Finally, when more data are available we would be able to investigate the optimal timing of conducting the study. In other words: how frequently should the retention model be updated in order to optimise the retention rate of the retailers.

More fundamentally, identifying customers as potential (partial) defectors is just a starting point for the managerial process of retaining these customers. Alternative tactics or strategies can be formulated and should be tested in the field to find out where and how the marginal marketing euro is best spent (Baesens et al., 2003).

REFERENCES

Anderson, J.A., 1982. Logistic discrimination. In: Krishnaiah, P.R., Kanal L.N. (Eds.), Handbook of Statistics, 2, pp. 169-191.

Anderson, E., Weitz, B., 1989. Determinants of continuity in convential industrial channel dyads, Marketing Science 8 (4) 310-323.

Anderson, E.W., Fornell, C., Lehmann, D.R., 1994. Customer satisfaction, market share, and profitability: Findings from Sweden, Journal of Marketing 58 (3) 53-66.

Athanassopoulos, A.D., 2000. Customer satisfaction cues to support market segmentation and explain switching behaviour, Journal of Business Research 47 (3) 191-207.

Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P., Vanthienen, J., 2003. Bayesian network classifiers for identifying the slope of the customer - lifecycle of long-life customers, forthcoming in European Journal of Operational Research.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., Dedene, G., 2002. Bayesian neural network learning for repeat purchase modelling in direct marketing, European Journal of Operational Research 138 (1) 191-211.

Bawa, K., Shoemaker, R.W., 1987. The coupon-prone customer: some findings based on purchase behaviour across product classes, Journal of Marketing 51 (4) 99-110.

Bhattacharya, C.B., 1998. When customers are members: Customer retention in paid membership contexts, Journal of the Academy of Marketing Science 26 (1) 31-45.

Blattberg, R.C., Getz, G., Thomas, J.S., 2000. Customer equity: Building and managing relationships as valuable assets, Harvard Business School Press, 228 pp.

Bloemer, J.M.M., Brijs, T., Vanhoof, K., Swinnen, G., 2003. Comparing complete and partial classification for identifying customers at risk, International Journal of Research in Marketing 20 (2) 117-216.

Bolton, R.N., 1998. A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction, Marketing Science 17 (1) 45-65.

Breiman, L., 1996. Bagging predictors, Machine Learning 24 (2) 123-140.

Breiman, L., 2001. Random forests, Machine Learning 45 (1) 5-32.

Buckinx, W., Moons, E., Van den Poel, D., Wets, G., 2004. Customer-Adapted Coupon Targeting Using Feature Selection, forthcoming in Expert Systems with Applications 26(4) 509-518.

Carroll, P., 1993. Revisiting customer retention, Journal of Retail Banking, 15 (1) 7-13.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: Synthetic minority over-sampling technique, Journal of Artificial Intelligence Research 16 321-357.

Colgate, M., Stewart, K., Kinsella, R., 1996. Customer defection: a study of the student market in Ireland, International Journal of Bank Marketing 14 (3) 23-29.

Corstjens, M., Lal, R., 2000. Building store loyalty through store brands, Journal of Marketing Research 37 (3) 281-291.

Dietterich, T., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization, Machine Learning 40 (2) 139-157.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern classification, Wiley, NY.

Dudoit, S., Fridlyand, J., Speed, T.P., 2002. Comparison of discrimination methods for the classification of tumors using gene expression data, Journal of the American Statistical Association 97 (457) 77-87.

Dwyer, R.F., 1997. Customer lifetime valuation to support marketing decision making, Journal of Direct Marketing, 11 (4) 6-13.

Ganesh, J., Arnold, M.J., Reynolds, K., 2000. Understanding the customer base of service providers: An examination of the differences between switchers and stayers, Journal of Marketing 64 (3) 65-87.

Hoekstra, J.C., Huizingh, E.K.R.E., 1999. The lifetime value concept in customer-based marketing, Journal of Market Focused Management 3 257-274.

Jones, M.A., Mothersbaugh, D.L., Beatty, S.E., 2000. Switching barriers and repurchase intentions in services, Journal of Retailing 76 (2) 259-274.

Keaveney, S.M., Parthasarathy, M., 2001. Customer switching behaviour in online services: An exploratory study of the role of selected attitudinal, behavioural, and demographic factors, Journal of the Academy of Marketing Science 29 (4) 374-390.

Kim, S.Y., Staelin, R., 1999. Manufacturer allowances and retailer pass-through rates in a competitive environment, Marketing Science 18 (1) 59-76.

Lemon, K.N., White, T.B., Winer, R.S., 2002. Dynamic customer relationship management: Incorporating future considerations into the service retention decision, Journal of Marketing 66 (1) 1-14.

Lindgreen, A., Pels, J., 2002. Buyer-seller exchange situations: four empirical cases, Journal of Relationship Marketing 1 (3) 69-93.

MacKay, D.J.C., 1992. The evidence framework applied to classification networks, Neural Computation 4 (5) 720-736.

Mittal, B., Lassar, W.M., 1998. Why do customers switch? The dynamics of satisfaction versus loyalty, Journal of Services Marketing 12 (3) 177-194.

Mittal, V., Kamakura, W.A., 2001. Satisfaction, repurchase intent, and repurchase behaviour: investigating the moderating effect of customer characteristics, Journal of Marketing Research 38 (1) 131-142.

Mizerski, R.W., 1982. An attribution explanation of the disproportionate influence of unfavorable information, Journal of Consumer Research 9 (3) 301-310.

Morgan, R.M., Hunt, S.D., 1994. The commitment-trust theory of relationship marketing, Journal of Marketing 58 (3) 20-38.

Morrison, D.G., 1969. On the interpretation of discriminant analysis, Journal of Marketing Research 6 156-163.

Morwitz, V.G., Johnson, E., Schmittlein, D., 1993. Does measuring intent change behaviour, Journal of Consumer Research 20 (1) 46-61.

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E., Kaushansky, H., 2000. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry, IEEE Transactions on Neural Networks 11(3) 690-696.

Nabney, I.T., 2001. NETLAB Algorithms for pattern recognition, Springer Verlag, London.

Nielsen, A.C., 2001. Major study to track store switching, Retail World, August.

Niraj, R., Gupta, M., Narasimhan, C., 2001. Customer profitability in a supply chain, Journal of Marketing 65 (3) 1-16.

O'Brien, L., Jones, C., 1995. Do rewards really create loyalty?, Harvard Business Review 73 (3) 75-83.

Peelen, E., Ekelmans, C.F.W., Vijn, P., 1989. Direct marketing for establishing the relationships between buyers and sellers, Journal of Direct Marketing 3 (1) 7-14.

Penny, W.D., Roberts, S.J., 1999, Bayesian neural networks for classification: How useful is the evidence framework, Neural Networks 12 (6) 877-892.

Peterson, R.A., 1995. Relationship marketing and the consumer, Journal of the Academy of Marketing Science 23 (4) 278-281.

Popkowski, L.P.T.L., Sinha, A., Timmermans, H.J.P., 2000. Consumer store choice dynamics: An analysis of the competitive market structure for grocery stores, Journal of Retailing 76 (3) 323-345.

Reichheld, F.F., Sasser, W.E., 1990. Zero defections: Quality comes to services, Harvard Business Review 68 (5) 105-111.

Reichheld, F.F., 1996. Learning from customer defections, Harvard Business Review 74 (2) 56-69.

Reinartz, W.J., Kumar, V., 2000. On the profitability of long-life customers in a noncontractual Setting: An empirical investigation and implications for marketing, Journal of Marketing 64 (4) 17-35.

Reinartz, W.J., Kumar, V., 2002. The mismanagement of customer loyalty, Harvard Business Review 80 (7) 86-94.

Rust, R.T., Zahorik, A.J., 1993. Customer satisfaction, customer retention, and market share, Journal of Retailing 69 (2) 193-215.

Schmittlein, D.C., Peterson, R.A., 1994. Customer base analysis: An industrial purchase process application, Marketing Science 13 (1) 41-67.

Sheth, J.N., Parvatiyar, A., 1995. Relationship in consumer markets: Antecedents and consequences, Journal of the Academy of Marketing Science 23 (4) 255-271.

Sonnenberg, F., 1990. Marketing to win, Harvard Business School Press.

Stum, D., Thiry, A., 1991. Building customer loyalty, Training and Development Journal 45 (4) 34-36.

Vakratsas, D., 1998. Household cost effects on purchase timing decisions: Do demographics matter?, Journal of Consumer Marketing 15 (1) 6-22.

Van den Poel, D., 2003. Predicting mail-order repeat buying: Which variables matter?, Tijdschrift voor Economie en Management 48 (3) 371-403.

Van den Poel, D., Larivière, B., 2003. Customer attrition analysis for financial services using proportional hazard models, forthcoming in European Journal of Operational Research.

Verhoef, P.C., Franses, P.H., Hoekstra, J.C., 2002. The effect of relational constructs on customer referrals and number of services purchased from a multiservice provider:

Does age of relationship matter?, Journal of the Academy of Marketing Science 30 (3) 202-216.

Verstraeten, G., Van den Poel, D., Prinzie, A., Van Kenhove, P., 2002. Detecting sequential patterns for cross-selling fast moving consumer goods, in: Zanasi A., Brebbia C.A., Ebecken N.F.F.E., Melli P. (Eds), Data Mining III, WIT Press, Bologna, 521-530.

Viaene, S., Baesens, B., Van Gestel, T., Suykens, J.A.K., Van den Poel, D., Vanthienen, J., De Moor, B., Dedene, G., 2002. Knowledge discovery in a direct marketing case using least squares support vector machines, International Journal of Intelligent Systems 16(9) 1023-1036.

Weerahandi, S., Moitra, S., 1995. Using survey data to predict adoption and switching for services, Journal of Marketing Research 32 (1) 85-96.

Wu, C., Chen, H.L., 2000. Counting your customers: Compounding customer's in-store decisions, interpurchase time and repurchasing behaviour, European Journal of Operational Research 127 (1) 109-119.

Zhang, Q., Varshney, P.K., Wesel, R.D., 2002. Optimal bi-level quantization of i.i sensor observations for binary hypothesis testing, IEEE Transactions on Information Theory 48 (7) 2105-2111.

Zeithaml, V.A., Berry, L.L., Parasuraman, A., 1996. The behavioural consequences of service quality, Journal of Marketing 60 (2) 31-46.

Ziliani, C., 2000. Retail micro-marketing: Strategic advance or gimmick? The International Review of Retail, Distribution and Consumer Research 10 (4) 355-368.

CHAPTER V

TOWARDS A TRUE LOYALTY PROGRAM: INVESTIGATING THE USEFULNESS AND FEASIBILITY OF REWARDING CUSTOMERS ACCORDING TO THE BENEFITS THEY DELIVER[9]

[9] This chapter is based on the following reference: Geert Verstraeten, Wouter Buckinx, Dirk Van den Poel, 2005. Towards a True Loyalty Program: Investigating the Usefulness and Feasibility of Rewarding Customers According to the Benefits They Deliver, submitted to *Journal of Marketing*, 2[nd] round of review process.

CHAPTER V:

TOWARDS A TRUE LOYALTY PROGRAM: INVESTIGATING THE USEFULNESS AND FEASIBILITY OF REWARDING CUSTOMERS ACCORDING TO THE BENEFITS THEY DELIVER

1. INTRODUCTION

In the previous two decades, marketing has seen a dramatic shift, in which traditional—i.e., product-oriented—marketing has given way to an increasingly customer-oriented view. The best-known theorem underlying this new view states that acquiring a new customer is several times more costly than retaining and selling additional products to existing customers (Rosenberg and Czepiel 1984). In this evolution, to which many authors refer as "the paradigm shift in marketing" (Brodie et al. 1997), loyalty of individual customers has rapidly grown to become the focal point of relationship marketing (Dick and Basu 1994).

Advocates of traditional relationship marketing attribute several advantages to loyal customers. They are said to increase their spending over the course of their relationship with a company (Reynolds and Arnold 2000), generate new customers by their positive word-of-mouth (Reichheld 2003), require diminished costs to serve (Dowling and Uncles 1997), exhibit reduced customer price sensitivities and have a salutary impact on the company's employees (Reichheld and Sasser 1990). In the remainder of this paper, we will refer to such alleged benefits of loyal customers as 'Loyalty Benefits'. An overview of the main findings with respect to these benefits is shown in the literature review part of this paper.

In the development of relationship marketing, different companies have conceived programs, often termed 'Loyalty Programs' or perhaps more accurately 'Reward Programs', in order both to reward and to stimulate such desirable customer behavior (Kivetz and Simonson 2003; Dowling and Uncles 1997). Today, companies ranging from large entities—such as American Airlines[10], American Express, AT&T, Carrefour, Hertz, Hilton Hotels and Shell—to small local merchants, offer reward programs that grant advantages to their customers, proportional to the money spent at their stores. Hence, regardless of the success of relationship marketing, these relationship-building programs are currently focused on rewarding merely repeat-purchase behavior (Nicholls 1989), being just one of the benefits attributed to loyal customers. Conversely, other benefits—which are also considered to be very important for the growth and the continuity of the company—are rewarded to a far lesser degree. Hence, it could be stated that currently, customers are rewarded proportional to a proxy variable of loyalty—spending—instead of loyalty itself. The following paragraphs discuss more in detail why such systems might not be the best method to reward loyal customers. Besides, we propose the use of an alternative reward criterion to overcome these concerns.

From a psychological point of view, rewarding customers can have multiple effects. First, the motivating impact of rewards has long been established in well-known experiments where animals have been proven to persist in the rewarded behavior (e.g., Latham and Locke 1991). Again, this underlines the importance of choosing the desired behavior to be rewarded, henceforth called the *reward criterion*. Accordingly, also in human behavior research, people have proven to be highly motivated to deliver efforts directed at achieving future rewards (e.g., Nicholls 1989). For marketing, it has been suggested that the excitement surrounding relationship marketing has created an expectation that customers who deliver benefits for the company will be rewarded for their loyalty (Dowling and Uncles 1997). In the context of loyalty programs, recent research has shown that customers are attracted more to programs if they feel that they are at an advantage to earn rewards when compared to other customers (Kivetz and Simonson 2003), which can again be related to social comparison theory (Festinger 1954). In summary, the design of the current loyalty programs can be seriously questioned. Indeed, loyal customers who deliver benefits to the company, but who are not big spenders, might feel discriminated against by big spenders who reap benefits without being

[10] The Advantage program of American Airlines is often cited as the first example of such a program.

loyal. This means that companies are rewarding loyal customers as well as spurious loyal customers. Spurious loyal customers show high repeat patronage behavior but have a low relative attitude towards the company. Hence, companies that are able to compensate customers - who exhibit loyalty benefits - to alleviate this discrimination might create a competitive advantage.

Intriguingly, to the best of our knowledge, no previous study has focused on evaluating the extent to which the use of a proxy variable for loyalty sufficiently rewards customers for the benefits often related to loyal customers. In this study, we will evaluate the feasibility of handling a reward scheme that is better able to remunerate loyal customers (for the different types of benefits they deliver) than the schemes that are currently being used. More precisely, our suggestion is to compensate customers in accordance with their share of wallet (SOW). Several reasons can be found why this would be a better reward criterion to compensate customers for their loyalty benefits than spending or length-of-relationship. First, customers who spend most of their budget at a specific company are showing repeat patronage behavior towards the company and, at the same time, are showing a higher relative attitude. This is based on the theory that the degree of attitudinal strength towards the firm and the degree of attitudinal differentiation underlie an individual’s relative attitude (Dick and Basu 1994). Compared to customers with a high spending or length-of-relationship, we consider customers with a high share of wallet to show a considerable differentiated attitude since their purchases are made explicitly at one of the different competitors. In contrast, customers with a high spending or length-of-relationship are not necessarily having a differentiated attitude and therefore these customers’ relative attitude might be lower. So, using share of wallet as a reward criterion might enable companies to reward customer showing both behavioral as well as attitudinal loyalty, which brings along that we expect rewards to go to customers with high loyalty benefits. Secondly, several studies support share of wallet to be one of the most important assets for companies (Verhoef 2003; Magï 2003). In line with the reasoning we made in a previous paragraph concerning the fact that customers will deliver efforts directed at achieving rewards, rewarding share of wallet might entice customers to increase their share of wallet.

So, this study evaluates the use of currently applied proxy variables and SOW as a criterion for rewarding loyalty benefits. Therefore we examine for both reward systems the strength of

the relationship between the amount of rewards that would be distributed and the loyalty behavior (in terms of loyalty benefits) that customers exhibit.
Besides, we propose a viable and feasible solution for each company that administers a customer database, to include SOW in the architecture of a reward scheme. To be precise, we propose a predictive model, which, at the same time, gives insight into the most important available database indicators of SOW. All results are validated in two different store settings: a grocery shopping environment and a general merchandising shopping setting.

2. LITERATURE REVIEW

2.1 Loyalty benefits

Advocates of traditional relationship marketing attribute several advantages to loyal customers. Table 1 gives an overview of studies focused on evaluating whether loyal customers do exhibit the alleged loyalty benefits. Some studies in this area are restricted to anecdotal discussions. Reichheld and Sasser (1990) were the first to claim that the length of a relationship makes customers more attractive, whereas Dick and Basu (1994) concluded that comparable benefits were dependent upon customers' loyalty level. In contrast, Dowling and Uncles (1997) did not agree and found arguments to dispute all of the proposed benefits.

These contradictions enticed researchers to search for empirical evidence, which only created more ambiguity. Reinartz and Kumar (2000) undermined nearly all of the benefits suggested by Reichheld and Sasser (1990). In contrast, Reynolds and Arnold (2000) supported the existence of beneficial loyalty behavior in a department-store setting, and Srinivasan et al. (2002) came to similar conclusions in an online setting. Finally, Reichheld (2003) confirmed his earlier findings: "Loyal customers talk up a company to their friends and colleagues". The review shows that ambiguity exists in determining whether loyal customers really deliver loyalty benefits. Our analysis will give more insight into this issue. We examine word-of-mouth, price insensitivity and purchase intentions since these are among the most investigated items.

Table 5.1: Literature Review: Customer Benefits.

Author	Target variable	Measurement	Benefits	Relationship	Supported (s) Not supported (ns) Anecdotal (a)	Data
Dick and Basu (1994)	Loyalty	Attitudinal and Behavioral	Word-of-mouth	+	a	General
			Resistance to counter persuasion	+	a	
			Search motivation	–	a	
Dowling and Uncles (1997)	Loyalty	/	Cost of serving	no	a	General
			Price insensitivity	no	a	
			Profitability	no	a	
			Word-of-mouth	no	a	
Reichheld (2003)	Loyalty	Behavioral	Word-of-mouth	+	s	Six industries
Reichheld and Sasser (1990)	Lifetime duration	/	Profitability	+	a	General
			Cost of serving	–	a	
			Price insensitivity	+	a	
			Word-of-mouth	+	a	
Reinartz and Kumar (2000)	Lifetime duration	Lifetime duration model	Profitability	+	ns	Catalog retailer
			Profit increase	+	ns	
			Cost of serving	–	ns	
			Price insensitivity	+	ns	
Reynolds and Arnold (2000)	Loyalty	Attitudinal (4 items)	Word-of-mouth	+	s	Department stores
			Competitive resistance	+	s	
			Share of wallet	+	s	
Srinivasan, Anderson and Ponnavolu (2002)	Loyalty	Attitudinal (7 items)	Word-of-mouth	+	s	Online B2C
			Price insensitivity	+	s	
			Consideration set size	+	ns	
This study	Loyalty	Behavioral (3 items)	Word-of-mouth	+	s	General merchandise and grocery shopping
			Price insensitivity	+	s	
			Purchase intentions	+	s	

2.2 Current reward programs

As Kivetz and Simonson (2003) note, an important goal of relationship marketing has been the development of customer loyalty. They also mention that loyalty programs have often been used to this end. Hence, while the original design of such programs consisted of rewarding customer loyalty (Dowling and Uncles 1997), in practice, most current reward systems do not use this criterion. Bonus systems like frequent flyer programs and schemes from credit card firms, banks, telephone companies and retailers encourage repeat purchase (Whyte 2004), and are usually rewarding customers for their spending, relationship duration or a combination of both (McMullan and Gilmore 2002). Also in academic research, spending and lifetime are often used to evaluate customers. In their loyalty program evaluation, Dowling and Uncles (1997) only consider reward schemes based on spending level. While Reinartz and Kumar (2000) recommend basing rewards on past spending of customers, in their research they evaluate whether long-life customers exhibit the benefits often attributed to loyal customers. Thus, they clearly evaluate the usefulness of length-of-relationship as an optional reward criterion. Verhoef (2003) makes use of a reward program that gives discounts based on the level of usage and the length of a customer's relationship. Additionally, he suggests that, when the reward structure depends on the length-of-relationship, customers would be less likely to switch, because of the time lag before the same level of rewards can be received by another supplier.

Two main reasons can be found to account for the use of proxy variables such as spending and length-of-relationship. The first reason for companies to make use of behavioral customer information is that such a measure of customer loyalty is not readily available in transactional databases (Jones and Sasser 1995). For a company with many customers, it is impractical to collect the required loyalty data for each of its customers by sending out questionnaires. In contrast, gaining knowledge about customers' spending behavior and lifetime duration is relatively straightforward because all the required data can be found in customer information files (Verhoef, Franses and Hoekstra 2002). Second, the use of these proxies might be justified because it has been shown that these variables are positively related to customer loyalty. East et al. (1995), for example, proved that highly loyal customers spend 32 percent more than other customers. Recently, Reichheld (2003) confirmed the finding that loyal customers spend more money. To our knowledge, however, the relationship between loyalty and length-of-relationship has not been thoroughly researched and , consequently, will be discussed in this study as well.

3. HYPOTHESES

3.1 Comparison of Current and New Reward Criteria

Our introduction casts doubt on the ability of current reward systems to compensate customers in proportion to their loyalty benefits. Consequently, our next step is to check whether the application of another criterion provides a better solution to this shortcoming. More specifically, for the reasons mentioned before, we expect that share of wallet represents a better criterion on which to reward customers for their loyalty benefits. Therefore, our first hypotheses make an efficiency comparison regarding this new criterion and the currently used criteria. The resulting hypotheses are as follows.

H1a(**b**) *If customers are rewarded for their share of wallet, the rewards go more to customers who exhibit benefits related to loyal customers (i.e., word-of-mouth, price insensitivity, purchase intentions) than if customers are rewarded for their spending* ***(length-of-relationship)****.*

3.2 Rewarding Loyals According to Their Predicted Share of Wallet

Even if rewarding based on customers' SOW proves to be more efficient, it is not straightforward to implement this in a reward program. As we mentioned above, individual SOW scores are not directly available in a company's database (Keiningham et al. 2003), whereas behavioral proxy variables like spending and lifetime duration are. To avoid the measurement of SOW for each of its customers, we present a model for predicting actual customer SOW by using a set of predictors derived from a company's database. However, in order to validate the usefulness of this new measure, we need to be sure that the efficiency gains attributed to rewarding according to SOW still hold when rewards are distributed according to these predicted SOW values. Consequently, both previous hypotheses are repeated, but now predicted SOW is used instead of actual loyalty.

H2a(**b**) *If customers are rewarded for their predicted share of wallet, the rewards go more to customers who exhibit benefits related to loyal customers (i.e., word-of-mouth, price insensitivity, purchase intentions) than if customers are rewarded for their spending* ***(length-of-relationship)****.*

4. METHOD

4.1 Data

We use data from four retail stores belonging to the same large European chain, in two middle-sized towns. While two of the stores carried a product assortment normally associated with *grocery* stores (e.g., food and beverages, cosmetics, laundry detergents, household necessities), two other stores carried an assortment usually associated with *general merchandise* stores (e.g., apparel, electronics and household appliances, do-it-yourself (DIY) and gardening equipment). In the remainder of the study, *Setting G* indicates the assortment usually associated with grocery stores and *Setting M* indicates the assortment usually associated with stores selling general merchandise. This partitioning is maintained throughout this study, in order to validate our findings across the two different store settings.
Using different store settings within a common store chain ensured comparability because databases were structured similarly, and recorded identical information in different store settings. Detailed purchase records were tracked for a period of 51 months and a summarized customer table was available that tracked basic customer demographics as well as first purchase date. It is important to mention that all transactions could be linked to customers, as the store requires use of a customer identification card.
In addition to these transactional data, a self-administered survey was used as a complementary data collection method. Data collection took place in each of the four retail stores mentioned previously. Surveys were randomly distributed to customers during their shopping trips, and customer identification numbers were recorded for all customers who received a questionnaire. Respondents were then asked to complete the questionnaire at

home and return the survey in a prepaid envelope. Of the 1500 questionnaires distributed in each setting, we received 875 usable responses in *Setting G*, and 779 usable responses in *Setting M*. A usable response had all fields completed, and the respondent could be successfully linked to his or her transaction behavior in the customer database. Hence, we reached ratios of usable response of 58.33% and 51.93% respectively. Given that customer identification numbers were collected for both respondents and nonrespondents, we tested for nonresponse bias by comparing several database variables between customer groups. We found no significant differences between the groups in terms of their spending, frequency of visiting the store, interpurchase time, length-of-relationship and response behavior towards companies' mailings.

4.2 Measures

In this section, we describe the variables we used, and how they were computed, originating either from our survey or from database records.

4.2.1 Survey-related variables.

We measured word-of-mouth, price insensitivity and purchase intentions, based upon Zeithaml et al. (1996), using seven-point Likert-type items. Consistent with previous research on loyalty programs (e.g., De Wulf, Odekerken-Schröder and Iacobucci 2001), we focus on measuring customer share of wallet to represent customer loyalty. Following Sharp and Sharp (1997), reward systems attempt to maximize customers' share of wallet and should be evaluated in terms of the behavioral changes they create. Hence, in this study, customer's SOW was determined as a composite measure by comparing a customer's spending at the retailer with their total spending in the relevant product category. As a first item, and similar to Macintosh and Lockshin (1997), the percentage of purchases made in the focal supermarket chain versus other stores was assessed on an 11-point scale that ranged from 0% to 100% in 10% increments (i.e., 0%, 10%, 20%, and so on). Additionally, two seven-point Likert-type items assessed the shopping frequency of the customers for the focal store when compared to other stores. We pretested the questionnaire several times and refined it on the basis of pretest results. Table 2(a) gives the exact wording of the items used.

4.2.2 Quality of the measurement model.

We initially performed an exploratory factor analysis using the items of the different scales. Several items were deleted, based on substantial cross-loadings. Because of different cross-loadings in both settings, word-of-mouth (WOM) was represented as a two-item scale in *Setting G*, and as a three-item scale in the *Setting M*. The other items had a consistent pattern of cross-loadings, resulting in a three-item scale for share of wallet (SOW), a two-item scale for price insensitivity (PRINS), and a single-item scale to measure purchase intentions (PINT). However, because the two items measuring price insensitivity had a significant yet weak correlation (*Setting G*: $R = 0.2846$, $\alpha = 0.4431$; *Setting M*: $R = 0.3061$, $\alpha = 0.4687$), we decided to reduce this scale to a single-item measure. After deletion of these items, we achieved a four-factor structure in which items loaded on *a priori* dimensions.

Table 5.2: (a) Wording of the items and (b) Factor Loadings and Construct Reliabilities.

Construct	Item Label	Item Wording
Word-of-mouth	WOM1	Encourage friends and relatives to do business with XYZ.
	WOM2	Say positive things about XYZ to other people.
	WOM3	Recommend XYZ to someone who seeks your advice.
Purchase Intentions	PINT1	Consider XYZ your first choice to buy groceries / general merchandise.
	PINT2	Do more business with XYZ in the next few weeks.
	PINT3	Do less business with XYZ in the next few months (–).
Price Insensitivity	PRINS1	Pay a higher price than competitors charge for the benefits you currently receive from XYZ.
	PRINS2	Take some of your business to a competitor that offers better prices (–).
Share of wallet	SOW1	Buy (much less … much more) grocery / general merchandise products at XYZ than at competing stores.
	SOW2	Visit other stores (much less frequently … much more frequently) than XYZ for your grocery / general merchandise shopping (–).
	SOW3	Spend (0% … 100%) of your total spending in grocery / general merchandise shopping at XYZ.

	SETTING G				SETTING M			
	SOW	WOM	PINT	PRINS	SOW	WOM	PRINS	PINT
SOW1	**0.895**	0.299	–0.204	–0.173	**0.889**	0.388	–0.115	–0.208
SOW2	**–0.880**	–0.257	0.187	0.218	**–0.842**	–0.268	0.205	0.198
SOW3	**0.898**	0.327	–0.270	–0.198	**0.838**	0.301	–0.161	–0.165
WOM1	–	–	–	–	0.312	**0.868**	–0.118	–0.119
WOM2	0.229	**0.892**	–0.160	–0.055	0.279	**0.818**	–0.089	–0.143
WOM3	0.367	**0.872**	–0.122	–0.111	0.352	**0.858**	–0.130	–0.130
PINT3	–0.249	–0.161	**0.999**	0.102	–0.223	–0.155	**0.999**	0.106
PRINS2	–0.221	–0.092	0.101	**1.000**	–0.192	–0.136	0.105	**1.000**
Variance Explained	2.680	1.851	1.198	1.142	2.587	2.514	1.129	1.171
Cronbach's α	0.871	0.715	–	–	0.818	0.805	–	–
Correlation	–	0.556	–	–	–	–	–	–

We tested construct reliabilities of the scales by means of Cronbach's coefficient alpha. Coefficients of all measures clearly exceed the .7 level recommended by Nunnally (1978). The output of the exploratory factor analysis, in terms of factor loadings and cross-loadings, the variance explained by each factor, and the reliability of the final scales, can be found in Table 2(b).

In addition, a maximum likelihood confirmatory factor analysis (CFA) was performed in LISREL 8.5 to evaluate the quality of the original measurement models. Since the initial solution did not fit the data well, we proceeded to increase model fit by excluding items until the model fits were acceptable. After several iterations, CFA obtained very satisfactory four-factor models for both settings; and the resulting measurement models were identical to the outcome of the exploratory factor analysis reported above. Since we used single-item scales to assess purchase intentions and price insensitivity, we accounted for the fallibility of such a scale by introducing some error variance (20%) during estimation, a procedure suggested by Jöreskog and Sörbom (1993, p. 37). Considering that the measurement models were not significant ($p > 0.05$), that all regression coefficients were statistically significant (smallest t: 14.21, $p < 0.01$), that the correlation between every item and the corresponding latent variable exceeds .50 (smallest $R = .6325$) and given the sufficient construct reliabilities

Table 5.3: Model Fit Indexes.

	SETTING G		**SETTING M**	
	Initial Solution	**Final Solution**	**Initial Solution**	**Final Solution**
χ^2	111.52	14.07	84.85	16.79
d.f.	38	10	38	16
P (> .05)	**.00**	**.17**	**.00**	**.40**
TLI (NNFI) (> .9)	.98	1.00	.98	1.00
SRMR (< .05)	.035	.013	.031	.015
AGFI (> .9)	.96	.99	.97	.99

Table 5.4: Correlation Matrix of the Independent Variables

	SETTING G				**SETTING M**			
	SOW	WOM	PINT	PRINS	SOW	WOM	PINT	PRINS
WOM	**0.43**	**1.00**			**0.47**	**1.00**		
	11.59				13.21			
PINT	**–0.31**	**–0.19**	1.00		**–0.27**	**–0.19**	**1.00**	
	–8.24	–4.58			–6.62	–4.42		
PRINS	**–0.26**	**–0.13**	**0.13**	**1.00**	**–0.21**	**–0.17**	**0.13**	**1.00**
	–6.88	–3.10	3.08		–5.13	–4.00	3.04	

reported above, we have tested our final models successfully in terms of unidimensionality, convergent validity and reliability (Steenkamp and van Trijp 1991). The model solutions are presented in Table 3, while the correlation matrices of the independent variables are presented in Table 4.

Finally, discriminant validity was examined by evaluating the decrease in performance when fixing correlations among constructs to 1. All chi-square difference tests (1 degree of freedom) were significant ($p < .01$), which indicates that all pairs of constructs correlated at less than one. For example, the high correlation between word-of-mouth and SOW corresponds to previous findings in the literature (e.g., Reichheld 2003), yet was found to be statistically different from one (*Setting G*: $\Delta\chi^2 = 235.96$, $df = 1$, $p < 0.01$; *Setting M*: $\Delta\chi^2 = 655.77$, $df = 1$, $p < 0.01$).

4.2.3 Database-related variables.

Spending and length of relationship were measured using the company's purchase transaction records. The former variable was computed as the cumulative amount spent by the customer in any of the stores of the focal supermarket chain since the introduction of the current database system. In comparable studies, the computation of length of relationship was complicated by the fact that researchers had to assess whether the customer was still 'alive' (cf. procedures suggested by Schmittlein and Peterson 1994). However, in this setting, all customers who filled in the questionnaire had visited the store during the weeks in which questionnaires were distributed, meaning that all respondents were active customers. This allowed us to compute the length of relationship by simply subtracting the first purchase date for a given customer in the company records from the date of administration of the questionnaire.

4.3 Model

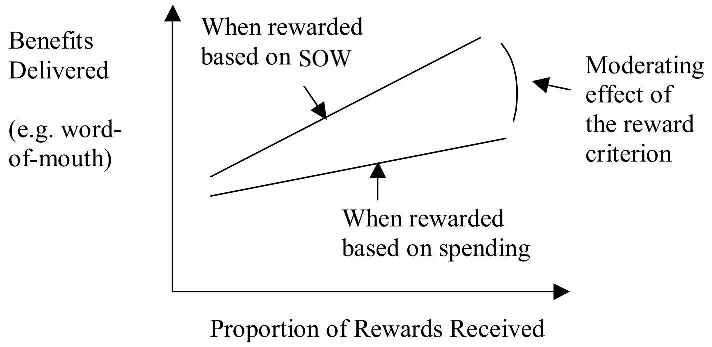
In order to test our hypotheses, we examined the relationship between loyalty benefits delivered and rewards received by the customer. Based on the combination of survey and database information, we are able to compute per customer (i) to what extent the customer delivers each of the benefits usually related to loyal customers, and (ii) the proportion of the rewards received by the customer if this customer was rewarded according to one of the investigated reward criteria. Hence, in this setting, the relationship between loyalty benefits delivered and rewards received is moderated by the reward criterion deployed. Accordingly, we will adapt a multiple regression framework with interaction effects to investigate our hypotheses (e.g., Cohen and Cohen 1983, Chapter 8). Graphically, we can sketch an exemplary regression model containing interaction effects as in Figure 1.

The given relationship could be captured in the following regression equation:

$$(1) \qquad Y = B_0^i + B_0^s X + B_1^i d_1 + B_1^s d_1 X + e \ ,$$

where Y represents one of the benefits delivered by the loyal customer, X represents the proportion of rewards received by the same customer, parameters with a superscript *i* indicate intercept parameters, and parameters with a superscript *s* indicate slope parameters.

Figure 5.1: Example of the Moderating Effect of the Reward Criterion on the Relationship between Rewards Received and Benefits Delivered.



Furthermore, if we suppose that d_1 represents a dummy variable showing a 0 where customers are rewarded for their spending and a 1 where customers are rewarded for their SOW, then B_0^s represents the strength of the relationship between the rewards received and the benefits delivered when customers are rewarded for spending, while $B_0^s + B_1^s$ shows the strength of the relationship between the rewards received and the benefits delivered when customers are rewarded for their SOW. Hence, the test for significance of B_0^s reveals whether customers who deliver benefits (e.g., in terms of word-of-mouth, price insensitivity, or purchase intentions) are rewarded more than others, when all customers are rewarded for their spending. Accordingly, the test for the significance of B_1^s reveals whether the reward criterion is a significant moderator of the relationship between X and Y, or, in other words, whether the relationship between rewards received and benefits delivered is significantly stronger (or weaker) if customers are rewarded for their SOW instead of their spending[11]. While the regression equation defined above delivers sufficient information to construct all necessary parameter estimates (and hence the graph given above), not all useful significance tests can be derived from this definition. Indeed, as Cohen and Cohen (1983, p 183) explain, the group that is represented by $d_1 = 0$ functions uniquely as a reference group here, and all the partial coefficients in fact turn upon it, whereby the relationship does not provide us with a test on the significance of the relationship between rewards received and benefits delivered when customers are rewarded for their SOW. Nevertheless, it is sufficient to adapt the coding scheme, and consider the other possible reward criterion as the reference group, in order to have a different view of the same model. Given this different dummy coding, the significance test of the new parameter B_0^s will reveal whether customers who do deliver benefits are rewarded more than others, when all customers are rewarded for their SOW.

Supposing that this moderator consists of more than two classes (say, *g* classes), we will adapt *g* regression equations to investigate the significance of the *g* slopes and all interactions between the *g* groups, where each of the reward criteria serves once as the reference group. Any of these equations—say equation *k*—can be represented as follows:

$$(2) \qquad Y = B_{0,k}^i + B_{0,k}^s X + \sum_{j=1}^{g-1} (B_{j,k}^i d_{j,k} + B_{j,k}^s d_{j,k} X) + e \ ,$$

where *k* ranges from 1 to *g*. Adding to the previous example, supposing we also wish to evaluate the strength of the relationship where customers are rewarded for their length of relationship or their predicted SOW, then the moderating variable consists of four (or more formally, *g*) groups, that can be represented by three (*g* – 1) dichotomies, d_1, d_2 and d_3, covering the three possible reward criteria (e.g., $d_1 = d_2 = d_3 = 0$: spending; $d_1 = 1,\ d_2 = d_3 = 0$: SOW; $d_1 = 0,\ d_2 = 1,\ d_3 = 0$: length of relationship; $d_1 = d_2 = 0,\ d_3 = 1$: predicted SOW). This procedure is in accordance with procedures discussed by Cohen and Cohen (1983, chapters 5 and 8) for conducting this type of analysis, and carefully considers the pitfalls indicated by Irwin and McClelland (2001) when interpreting the results of moderated multiple regression models.

[11] Note that interpretation of the intercept parameters is similar, but is of less relevance to our research topic.

4.4 Predicting Share of Wallet

In order to make use of our conceptual model, marketing management needs to be able to define customers' SOW. Nevertheless, share of purchases cannot be derived directly from the information in a database, so in a real environment, a predictive model is needed. This part describes how the model is built. The variables, classification technique, validation method and variable-selection procedure are discussed in the following paragraphs.

4.4.1 Variables.

We only used information that is available in the customer database at the individual customer level. These data are collected by the use of a loyalty card. The dependent variable in the model is share of wallet, which is measured by a construct of the three above-mentioned questionnaire items. In total, 33 independent variables were compiled to predict loyalty in the general merchandise store setting and 34 independent variables were computed for the grocery setting. Table 5 summarizes all these variables, together with a brief description of how they are calculated. The results of the model are included in this table and discussed in a later section. It shows that we used more or less the same predictors in both shopping environments. We will, therefore, be able to compare the relevant information for the two settings. The following paragraphs give a short overview of the variables that are taken into account.

Reinartz and Kumar (2002) argue extensively for the inclusion of several predictors in their lifetime duration model. Since their variables are also intended to explain the strength of a relationship, our variable list will be similar. As a consequence, we will not discuss the same literature in detail. First, we focus on variables that are commonly used in scoring models for customer relationship management (Bult and Wansbeek, 1995). The level of customer spending and the frequency of customers' visits prove to be efficient behavioral information for the detection of weak or strong relations. Consequently, we include customers' individual spending and visit frequency derived from data concerning the last month, six months, one year, two years and over our complete data time series. The average spending and customers' spending relative to the length of time since their first purchase are computed to take into account relative figures as well. Related variables in this area are the number of products bought and the amount of money spent on fresh products that need to be weighed by the customers themselves. This last information was only relevant for the grocery setting. Furthermore, we also include the average interpurchase time and the time since the customer's last purchase. All these variables are frequently used to determine loyal customers and to characterize customers who exhibit strong relations with a company (Reinartz and Kumar 2002). Moreover, we include the standard deviation of the interpurchase time as this gives insight into the regularity of customers' visits and turns out to be an important variable for predicting future loyalty (Buckinx and Van den Poel 2005). Some studies support the relation between customers' lifetime and their profitability, while others questioned these results (Reinartz and Kumar 2000). Therefore, we incorporate the length of relationship into our model. Reinartz and Kumar (2002) also incorporate the scope of customers' purchases into their predictive model. Likewise, Baesens et al. (2004) recently showed the variety of products purchased to be a predictor of future spending increases or decreases. Thus, the number of categories from which a customer bought products is included in our model. We summed the same behavior of customers during their previous one, two and three years. Returns of goods can be important information too, though the hypothesis of Reinartz and Kumar (2002) concerning this behavior was not supported. Returns may be a signal for dissatisfaction and consequently a weaker relationship. In contrast, for some products, it is shown that returns signal a positive association with customer loyalty (Buckinx and Van den Poel 2005). We include the total amount of returned goods and two dummies: whether or not a customer ever returned a product or cancelled an order. As earlier in our study, we assume that loyalty is related to price insensitivity (Dowling and Uncles 1997; Srinivasan, Anderson and Ponnavolu 2002). Consequently, we try to derive which customers behave like promotion seekers by computing four promotion-related variables: the number of promoted products bought, the money spent on promotions, the number of visits where at least one promoted product was purchased and, finally, the percentage of products purchased on promotion. The next types of information that we presume to have explanatory power for customer SOW are variables related to customers' response to mailing actions. Though neither the company from the grocery setting nor the general merchandise store is active in direct marketing, their most important communication channel is a biweekly leaflet. Therefore, for each of the customers, we incorporate the percentage of occasions the customer made a visit to the store after having received the leaflet. Because of limited budgets, not all customers receive a catalogue each week. Therefore, we included the percentage of times a customer came to the store even though he or she had not received a catalogue. Finally, we assume a positive relation between the number of times someone visits the store during one and the same promotion period[12] and SOW.

Finally, the strength of a relationship is likely to depend on the costs and benefits experienced. By including the distance between the store and the customers' residence, we test for the influence of living far from or close to the shop.

4.4.2 Classification technique and leave-one-out procedure.

In order to predict customers' SOW, we apply a multiple linear regression model. We will evaluate the predictive power of this model on a validation set that is independent of the information used to build the model. However, the limited number of observations in each of the two settings and the elaborate number of independent variables make it hard to split our data in an estimation and a hold-out test set. As a consequence, we prefer a resampling method called leave-one-out cross-validation because it proves to be superior for small data sets (Goutte 1997). Using this procedure, our data are divided into *k* subsets, where *k* is equal to the total number of observations. Next, each of the subsets is left out once from the estimation set and is then used to estimate a validation score. To get an idea of the power of the model, the final test set is built by stacking together the *k* resulting validations. The performance of the model is evaluated by the adjusted R^2 and the MSE—on the estimation set as on the validation set.

4.4.3 Variable selection.

Considering the number of variables and the rather limited number of observations, we make use of a variable-selection technique. Thanks to this method, the dimensionality of the model can be reduced and redundant variables are removed, which is in favor of the performance of the model. In order to guarantee the selection of the best subset, we apply the leaps-and-bounds algorithm proposed by Furnival and Wilson (1974). Their efficient technique identifies the model with the largest adjusted R^2 for each number of variables and at the same

[12] A promotion period is the period where the offers of one catalogue are valid.

Table 5.5: Description and standardized Parameter estimates of Variables Used for Predicting Share of Wallet.

Variable	Description	Grocery Shopping Multivariate	Grocery Shopping Univariate	Grocery Shopping Ranking	General Merchandise Multivariate	General Merchandise Univariate	General Merchandise Ranking
Spending_1M	Spending during last month.		0.35396 ***	21		0.12996 ***	28
Spending_6M	Spending during last six months.		0.45817 ***	11		0.24760 ***	18
Spending_1Y	Spending during last year.		0.47892 ***	4		0.29260 ***	8
Spending_2Y	Spending during last two years.		0.47424 ***	6	0.23464 **	0.30156 ***	6
Spending	Spending in total history.		0.47144 ***	8	–0.20188 *	0.27243 ***	13
Frequency_1M	Number of purchases during last month.		0.34773 ***	22	–0.07706	0.16882 ***	24
Frequency_6M	Number of purchases during last six months.		0.43562 ***	19		0.27715 ***	12
Frequency_1Y	Number of purchases during last year.		0.44546 ***	16	0.26789 *	0.29420 ***	7
Frequency_2Y	Number of purchases during last two years.		0.44943 ***	14	–0.36123 **	0.28686 ***	10
Frequency	Number of purchases in total history.		0.43889 ***	18		0.26861 ***	15
NumItems	Number of product items bought.	0.09898 **	0.47046 ***	9	–0.11047 *	0.16022 ***	26
Spending_Weight °	Spending in products that need to be weighted by the customer.	0.10086 **	0.43953 ***	17			
rSpend_Freq	Average Spending per visit.	0.07685 **	0.17850 ***	30	0.05450	0.00365	33
rSpend_lor	Spending relative to the length of the customer's relationship.		0.47263 ***	7		0.28188 ***	11
Recency	Number of days since last purchase.		–0.20352 ***	29		–0.13204 ***	27
Ipt	Average number of days between store visits.		–0.29652 ***	25		–0.20427 ***	21
Std_Ipt	Standard deviation of the number of days between the purchases.	–0.05203	–0.32269 ***	23	–0.10788 ***	–0.24934 ***	17
Lor	Length of customer relationship.		0.09398 ***	33		0.06881 *	29
Numcat_LY	Number of different product categories purchased from during last year.	0.21724 ***	0.52211 ***	1		0.33452 ***	2
Numcat_2Y	Number of different product categories purchased from during last two years.		0.47699 ***	5		0.30311 ***	4
Numcat_3Y	Number of different product categories purchased from during last three years.		0.44598 ***	15		0.25404 ***	16
Numcat	Number of different product categories purchased from during the total history.		0.48046 ***	2		0.31897 ***	3
Neg_Inv	Dummy to indicate if the customer ever had a negative invoice (1/0).	0.03956	0.29186 ***	27		0.18993 ***	22
Ret_Item	Dummy to indicate if the customer ever returned an item (1/0).		0.26563 ***	28		0.17951 ***	23
Returns	Total value of returned goods.		0.15719 ***	31	–0.06357 *	–0.00814	32
NumPromItems	Number of items bought that appeared in company's promotion leaflet.		0.45390 ***	13	0.13010 **	0.22736 ***	19
SpenPromItems	Money spent on products that appeared in promotion leaflet.		0.45724 ***	12		0.27082 ***	14
Visitspromitems	Number of visits on which a product is bought that appeared in the promotion leaflet.		0.46804 ***	10	0.11218	0.30203 ***	5
PercNumPromItems	Percentage of products bought that appeared in leaflet.		0.01389	34		0.05476	30
PercResp_Leaf	Percentage of times a purchase is made given that a promotion leaflet was received.	0.13667 **	0.47915 ***	3	0.26732 ***	0.34265 ***	1
PercResp_Noleaf	Percentage of times a purchase is made given that no promotion leaflet was received.		0.30984 ***	24		0.16728 ***	25
MoreThanOnce	Number of times that a customer visits more than once within the same promotion period.		0.43077 ***	20		0.29253 ***	9
PercMoreThanOnce	MoreThanOnce divided by the number of times a customer bought in a promotion period.		0.29400 ***	26	0.06032	0.20873 ***	20
Distance	Distance to the store.		–0.12651 ***	32		–0.03885	31

° This variable was only included for the grocery setting and not in the general merchandise store setting.

* p <.10, ** p <.05, *** p <.01.

time avoids a full search of the variable space. The best subset is chosen based on the adjusted R^2 that can be achieved on the total estimation set.

5. RESULTS

5.1 Introduction

Following Irwin and McClelland (2001), we report the detailed coding scheme used in this research. This coding scheme is represented in Table 6, indicating that Share of Wallet (SOW) was considered as the reference group in the first coding iteration, next Spending (SPEN), Length of Relationship (LOR), and finally Predicted Share of Wallet (PSOW).

Table 5.6: Coding and Recoding of the Interaction Dummies (Dummy-Variable Coding)

	r = 1			r = 2			r = 3			r = 4		
	$d_{1,1}$	$d_{2,1}$	$d_{3,1}$	$d_{1,2}$	$d_{2,2}$	$d_{3,2}$	$d_{1,3}$	$d_{2,3}$	$d_{3,3}$	$d_{1,4}$	$d_{2,4}$	$d_{3,4}$
SOW	0	0	0	0	0	1	0	1	0	1	0	0
SPEN	1	0	0	0	0	0	0	0	1	0	1	0
LOR	0	1	0	1	0	0	0	0	0	0	0	1
PSOW	0	0	1	0	1	0	1	0	0	0	0	0

Since we are interested in the slope parameters in equation (2), they can be summarized as in Table 7(a), where the diagonal represents the slopes of the different relationships, and the off-diagonal figures represent the differences between the slopes. For example, if SOW is considered as the reference group (r = 1), then the relationship between the benefits and the rewards—if customers are rewarded proportionally for their SOW —can be represented as $B_{0,1}^{s}$, while the difference between rewarding for spending versus rewarding for SOW can be represented as $B_{1,1}^{s}$. The corresponding standard estimate of this parameter allows us to interpret whether this difference is significant. Because these differences are symmetric, all information below the diagonal is redundant and will not be repeated. In Table 7(b), we give an overview of all parameters and their standard errors for the different regression equations. The relationships are also represented graphically in Appendix 5.A.

Before discussing the results concerning our hypotheses, we can use the information available in Table 7(b) to draw conclusions concerning issues stated in the introductory part of this study. Namely, in the discourse about loyalty benefits, we can examine to what extent rewards go to customers who exhibit loyalty benefits if they are rewarded in accordance with their share of wallet. Therefore, we consider the parameters $B^{s}_{0,1}$ of the different models. When inspecting the results in Table 7(b), it is clear that these relationships are highly significant.

Table 5.7: (a) Interpreting (Re)Coded Parameter Estimates and (b) Results of Model Estimation.

	B_{SOW}	B_{SPEN}	B_{LOR}	B_{PSOW}
B_{SOW}	$\boldsymbol{B^{s}_{0,1}}$	$B^{s}_{1,1}=-B^{s}_{3,2}$	$B^{s}_{2,1}=-B^{s}_{2,3}$	$B^{s}_{3,1}=-B^{s}_{1,4}$
B_{SPEN}	$B^{s}_{3,2}=-B^{s}_{1,1}$	$\boldsymbol{B^{s}_{0,2}}$	$B^{s}_{1,2}=-B^{s}_{3,3}$	$B^{s}_{2,2}=-B^{s}_{2,4}$
B_{LOR}	$B^{s}_{2,3}=-B^{s}_{2,1}$	$B^{s}_{3,3}=-B^{s}_{1,2}$	$\boldsymbol{B^{s}_{0,3}}$	$B^{s}_{1,3}=-B^{s}_{3,4}$
B_{PSOW}	$B^{s}_{1,4}=-B^{s}_{3,1}$	$B^{s}_{2,4}=-B^{s}_{2,2}$	$B^{s}_{3,4}=-B^{s}_{1,3}$	$\boldsymbol{B^{s}_{0,4}}$

	Setting G				**Setting M**			
	Parameter Estimates (Standard Error)				**Parameter Estimates (Standard Error)**			
	B_{SOW}	B_{SPEN}	B_{LOR}	B_{PSOW}	B_{SOW}	B_{SPEN}	B_{LOR}	B_{PSOW}
Word-of-mouth								
B_{SOW}	462.54 (45.07)***	–413.18 (49.64)***	–424.46 (61.51)***	–224.14 (74.13)***	502.55 (47.05)***	–455.79 (53.01)***	–532.72 (59.65)***	–262.18 (91.61)***
B_{SPEN}		49.36 (20.8)**	–11.28 (46.74)	189.04 (62.42)***		46.76 (24.43)*	–76.93 (44.06)*	193.61 (82.31)**
B_{LOR}			38.08 (41.85)	200.32 (72.21)***			–30.17 (36.67)	270.54 (86.73)***
B_{PSOW}				238.39 (58.85)***				240.37 (78.6)***
Price Insensitivity								
B_{SOW}	343.17 (51.51)***	–234.01 (56.81)***	–310.73 (70.26)***	–111.04 (84.79)	306.12 (56.29)***	–286.52 (63.43)***	–242.21 (71.37)***	–204.98 (109.6)*
B_{SPEN}		109.16 (23.96)***	–76.73 (53.45)	122.96 (71.49)*		19.6 (29.22)	44.31 (52.71)	81.54 (98.47)
B_{LOR}			32.43 (47.78)	199.69 (82.57)**			63.91 (43.87)	37.23 (103.77)
B_{PSOW}				232.13 (67.35)***				101.14 (94.04)
Purchase Intentions								
B_{SOW}	397.47 (51)***	–272.49 (56.17)***	–342.17 (69.6)***	47.15 (83.87)	355.3 (55.9)***	–249.9 (62.98)***	–381.19 (70.86)***	102.95 (108.83)
B_{SPEN}		124.97 (23.54)***	–69.67 (52.88)	319.64 (70.62)***		105.41 (29.02)***	–131.29 (52.34)**	352.84 (97.78)***
B_{LOR}			55.3 (47.36)	389.31 (81.71)***			–25.89 (43.56)	484.14 (103.04)***
B_{PSOW}				444.61 (66.58)***				458.25 (93.38)***

For example, the relationship between rewards received and word-of-mouth in *Setting G* is positive and significant ($B = 462.54$, $p < 0.0001$). By analogy, we can investigate the other parameters, and we conclude that if customers are rewarded for their SOW, the rewards would go more to customers who engage more in word-of-mouth, are less price sensitive, and exhibit higher purchase intentions, in both settings.

Next, as discussed previously, spending and length-of-relationship are commonly used proxies for loyalty in general, and because they are more readily available to the company, they are commonly used as reward criteria. So, the analysis of the correlations between SOW and both proxies suggests a strong significant correlation between SOW and spending in both settings (*Setting G*: $R = 0.4714$, $p < 0.0001$; *Setting M*: $R = 0.2724$, $p < 0.0001$). The correlation between length-of-relationship and SOW, however, proves to hold in the setting of grocery shopping ($R = 0.1150$, $p = 0.0006$), but not in the setting related to general merchandise shopping ($R = 0.0393$, $p = 0.2722$).

Likewise, since both spending and length-of-relationship have been used previously as a reward criterion, we examine whether customers who are rewarded for these also deliver the benefits related to loyal customers. Because the results are rather more ambiguous, we will discuss this relationship for each benefit separately. If customers are rewarded for their spending, the evidence is only moderate that these customers would also deliver more word-of-mouth to the company (*Setting G*: $B = 49.36$, $p = 0.0177$; *Setting M*: $B = 46.76$, $p = 0.0557$). Apparently, this relationship is more pronounced for grocery shopping than general merchandise. This effect is comparable to the effect of the same reward criterion on price sensitivity. If customers are rewarded for their spending, rewards would be distributed significantly more to price insensitive shoppers in the grocery setting ($B = 109.16$, $p < 0.0001$), while no such significant relationship is detected for general merchandise ($B = 19.6$, $p = 0.5024$). Accordingly, those customers rewarded for their previous spending would be customers showing significantly higher purchase intentions towards the store. This effect is consistent in both settings (*Setting G*: $B = 124.97$, $p < 0.0001$; *Setting M*: $B = 105.41$, $p = 0.0003$). If customers are rewarded for their length-of-relationship, the relationships between rewards received and benefits delivered are unambiguous. None of these relationships is significant (significance ranging between $p = 0.1453$ and $p = 0.5524$).

5.2 Hypothesis Tests

In order to validate H_{1a} and H_{1b}, we test whether the slope of the curve based on SOW is significantly higher than the slope of the curves based on spending or length-of-relationship. It is important to notice here that this difference was highly significant in all of the cases ($p < 0.001$ in all cases). Hence, the relationship between the proportion of rewards received and each of the benefits related to loyal customers was significantly higher when customers were rewarded for their SOW instead of their spending or length-of-relationship.

Finally, in order to test the applicability of a reward scheme based on SOW, H_{2a} and H_{2b} test the relationship between rewards received and benefits delivered if the reward criterion was predicted SOW instead of spending or length of relationship. Because the results are again more ambiguous, we will describe the effect per benefit delivered. First, the relationship between rewards received and word-of-mouth delivered by customers is significantly higher if customers are rewarded for their predicted SOW than if they are rewarded for their spending or length-of-relationship (significance ranging between $p = 0.0187$ and $p = 0.0018$). Second, considering price insensitivity, the results are conditional upon the setting: while there is a marginally significant effect in grocery shopping (PSOW vs SPEN: $B = 122.96$, $p = 0.0855$; PSOW vs LOR: $B = 199.69$, $p = 0.0156$), the effect in general merchandise shopping is clearly insignificant (PSOW vs SPEN: $B = 81.54$, $p = 0.4077$; PSOW vs LOR: $B = 37.23$, $p = 0.7198$). Finally, considering purchase intentions, the results across the two settings are again generally consistent: if customers are rewarded for their predicted SOW, those customers with higher purchase intentions will be rewarded significantly more than if they were to be rewarded for their spending or length-of-relationship ($p < 0.001$ in all cases).

5.2 Predicting Share of Wallet

In this section, we describe the performance of the multiple linear regression model used to predict SOW. In Table 8, the performance of the models with all variables—the 'full model'—is compared with the performance of the best performing models in terms of adjusted R^2 and the MSE. We evaluate both the performance of a model where all observations are used for estimation purposes—hence called the 'estimation set'—with a model where the leave-one-out procedure is used to evaluate the real performance of the model. All models are significant considering a significance level smaller than 0.0001.

Table 5.8: Model Performance after Variable Selection Procedure.

	Setting G				Setting M			
	Full Model (v = 35)		Final Model (v = 7)		Full Model (v = 34)		Final Model (v = 13)	
	Estimation Set	Leave-one-out	Estimation Set	Leave-one-out	Estimation Set	Leave-one-out	Estimation Set	Leave-one-out
$R^2_{adjusted}$	0.29256	0.23007	0.30632	0.29416	0.12422	0.04401	0.14119	0.10354
MSE	0.55856	0.61074	0.54770	0.55741	0.63946	0.70856	0.62707	0.65675

As could be expected, the leave-one-out performance decreases slightly compared to the estimation set performance. Additionally, the difference between both performance measures decreases when fewer variables are used in the model; indicating that the variable-selection procedure tempers the negative consequences related to overtraining. Finally, predictive performance increases with the use of a variable selection technique, indicating the usefulness of such a procedure for the prediction of SOW.

Obviously, the most important benefit of the variable-selection procedure lies in detecting a parsimonious subset of database variables that can be used to predict SOW in both store settings. Remarkably, there is a considerable difference in the number of variables selected in each case. In the grocery setting only 7 of the 34 variables are retained, whereas for general merchandise stores more information is needed: the maximum adjusted R^2 was reached with 13 predictors. Table 5 shows the standardized parameter estimates and the significance levels for the variables that are chosen by the feature selection procedure. We represent the multivariate solutions as well as the univariate results of each individual variable since there is clear evidence of multicollinearity in the multivariate model[13]. For the same reason, we also represent the univariate standardized parameter estimates from variables that were not selected for the final model. While the univariate results should be used for interpretation of the signs and significance of the variables, the multivariate solution delivers the best fit to the data, and hence offers the best prediction of SOW.

In order to detect whether different variables are important in the different settings, we investigated the Spearman rank-order correlation, which is a nonparametric measure of association based on the rank of the data values. Given the very large and significant correlation of 0.8915 ($p < 0.0001$), we conclude that the importance of the variables does not differ significantly between the two settings. In order to enhance comparability, we included

[13] For example, several variables that are univariately highly significant are not selected or turn out to be insignificant in the multivariate model.

the ranking of the variables in Table 5. However, considering the multicollinearity we discussed previously, the final predictive models in each setting differ considerably in the variables used. As discussed previously, this should not lead the reader to conclude that different variables are needed to predict SOW in the different settings. The final model for each store setting is shown in Table 5. The importance of each of the variable types for our predictive model is examined in the next section.

6. DISCUSSION

6.1 Loyalty Benefits

Previous empirical research, as well as anecdotal evidence, has focused on the relationship between loyal customers and the alleged beneficial characteristics of such loyal customers. However, considering the conflicting results of these studies, decisive conclusions are lacking. Our research, however, confirms the existence of benefits from loyal customers by examining the relationship between share of wallet and three different benefits. A company's loyal customers actively recommend its services to their peers. Besides, these customers are price insensitive and are motivated to repurchase in the future. Our findings confirm the results of Reynolds and Arnold (2000) and Srinivasan, Anderson and Ponnavolu (2002), who investigated these associations in an online environment. However, they counter the conclusions of Reinartz and Kumar (2000), who could find no support for any of these benefits although both their and our studies focused on a noncontractual setting. What can be the reason for these mixed results? A credible explanation is the way in which loyalty was approached in each of the studies. When considering all empirical evidence, only Reinartz and Kumar (2000) reject any connection between loyalty and loyalty benefits. Table 1 shows that theirs is the only study to examine lifetime duration, while others took behavioral or attitudinal loyalty into account. This might indicate that the conclusions depend on which criterion is used. Indeed, our study agrees with this reasoning, since a significant relationship between customer lifetime and one of the three benefits examined was not detected (H2c). This confirms our assumption that the way in which loyalty is approached drives the studies' conclusions.

6.2 Share of Wallet Outperforms Other Behavioral Proxies as Reward Criterion

This study is the first to question the criteria that are widely used by companies to manage their reward system. Currently, most companies use a reward system where compensations are dependent on customers spending behavior. Past research concerning human behavior has shown that rewards will motivate customers to do what is necessary to get the related returns (Nicholls 1989). Our results show that if companies want to reward customers for more than only repeat-purchase behavior, they are well advised to take into account customers' (predicted) SOW rather than relying on spending or customers' lifetime. This implies that companies that stay dedicated to their current reward strategy are neglecting customers who turn out to be beneficial. These customers positively distinguish themselves from other customers because they actively spread positive word-of-mouth about a company, are willing to pay a superior price and have clear positive intentions to visit the store in the future. Current reward schemes do not compensate for these contributions, while these benefits are extremely valuable for growth, profitability and continuity of a company.

Customers' referrals are very influential in decision-making processes since they seem to be reliable sources of information. Reichheld (2003) emphasizes this reasoning in his last study: "The only path to profitability and growth may lie in a company's ability to get its loyal customers to become its marketing department." Customers who recommend a company to their friends and relatives help to avoid leakage from the customer base (Jones and Sasser 1995). In their recent study, Wangenheim and Bayon (forthcoming) provide evidence that positive word-of-mouth referrals can convince up to 16% of the recipients to switch to the 'advertised' company in a consumer market, and as much as 51% in an industrial market, provided that the source is considered experienced and similar to the receiver. Reichheld (2003) warns of a bad mix of promoters and detractors: the percentage of customers who are promoters has a strong relation with a company's growth. The habit of loyal customers of bringing in new customers is particularly valuable, particularly if the company is competing in a mature market. The second benefit of loyal customers can have direct impact on companies' profits: less price-sensitive customers are indifferent about paying more for the same product/service. As a result, it is not necessary to convince these customers by offering them price cuts and discounts. This means that these customers do not come to a store merely to pick all the 'cherries' but buy products that generate higher margins as well. Finally, customers' purchase intentions guarantee companies' continuity. Bolton et al. (2000) found that purchase intentions do have a strong positive relationship with subsequent repatronage decisions and consequently with retention behavior. This makes them interesting since they assure a steady stream of resources to the company.

The previous paragraph emphasizes the value of the different benefits. In contrast, loyal customers will be discriminated against by companies that apply traditional reward programs. There is a danger that this strategy might motivate loyal customers to leave a company. Feinberg et al. (2002) demonstrate that customers will prefer their favorite firm less when they are put at a disadvantage compared to nonloyal customers—and which company likes to lose customers who deliver substantial benefits? Even worse: promoters of the company can become detractors who will substitute their former recommendations into negative word-of-mouth (Reichheld 2003) that will damage a firm's reputation. Our results suggest that programs that apply (predicted) SOW as a reward criterion are able to give more rewards to customers with diverse loyalty benefits and less rewards to customers having no loyalty benefits. As such, they would compensate customers more effectively for their beneficial behavior, and consequently, such programs are expected to induce a higher retention rate. Customers who experience appreciation for their contribution and feel recognized in a reward program will weigh comparisons with competitors less heavily in making purchase decisions (Bolton et al. 2000).

Hallberg (2004) reports that the success of companies' reward systems is not only dependent on results that have an immediate financial impact. The extent to which these reward systems attach customers emotionally to a brand or a store is as important. The newly proposed reward criterion in this study will focus management's attention on different types of benefit.

6.3 Effect of Reward Programs

In addition to marketing research on the profitability of loyal customers, a number of other studies have concentrated on the effects of reward programs on customer behavior. A literature review confirms Dowling and Uncles' (1997) theory that it is hard to influence customer behavior with the current reward schemes. The limited number of studies investigating this topic shows diverse effects of reward programs on behavioral loyalty. Mägi (2003) investigated the effect of loyalty card programs on share of purchases in a grocery shopping environment. Her results confirm the mixed results and suggest that at the store level, no effect must be expected on the share of purchases. The conclusions of Verhoef

(2003) indicated a marginal effect of relationship marketing instruments (RMI) on share development. Even more importantly, the outcomes revealed that loyalty programs' effect was, for the most part, explained by past customer behavior: "Customers with a small (past) customer share are more likely to increase their customer share in the next period." These findings emphasize the need for a reward criterion such as the one we propose in this study. More specifically, Verhoef (2003) investigates the impact of a reward program on the change in share of purchases. However, as for most companies, this study included a reward system that distributed price discounts based on the level of purchases and the length-of-relationship. Such schemes do not take into account a customer's behavioral loyalty, which offers a potential explanation for their marginal effect. Customers exhibiting an already high level of SOW are not likely to increase their spending, since they already make all their purchases in a particular store. This is supported by the conclusions of Verhoef (2003) on the importance of the initial customer share in explaining the (small) effect (see above). In general, the mixed effects of relationship programs might be explained by this phenomenon. Selection criteria, which define the level of incentives or rewards, should be in accordance with the goals of the marketing program. On that reasoning, spending as a reward criterion to increase customers' share of wallet is not the best option. Instead, making use of (predicted) SOW to manage reward programs, as suggested in this study, seems a valid solution. Other studies that value customer loyalty for marketing action purposes are those of Dowling and Uncles (1997) and Reinartz and Kumar (2002). Though these last authors examine the value of a lifetime duration framework, their managerial implications emphasize the need for loyalty, measured by share of wallet, to fine tune companies' actions and to deal with different types of customers. Nevertheless, they did not empirically check the advantages related to that proposition, nor did they offer a model to define share of wallet for the total customer base. Therefore, ours is virtually the first study to show empirically the importance of using SOW in a reward system and to propose a feasible solution that incorporates individual customer loyalty into a relationship-marketing program.

6.4 Model Results

The outcomes from the predictive SOW models end in several interesting contributions. First, the significance of the overall predictive models in both settings points to the ability of marketing management to compute a customers' SOW to an acceptable extent from his or her transactional data. Without this feature, a company is forced to send out questionnaires to all of its customers in order to know their exact SOW. Using the method presented above, however, it is sufficient to interrogate a limited number of randomly chosen customers from the database. In this model, we only incorporated data that can be derived directly from the customer database and that is available for all customers thanks to their customer identification cards. This enables companies to create a SOW score for every customer at any given moment. Given the satisfactory predictive performances of our models, efficiency in rewarding customer benefits validates the usefulness of our new proxy measurement. The results confirm the findings concerning actual customer SOW: rewarding in accordance with predicted SOW is significantly better than rewarding in proportion to commonly used proxy variables (see previous paragraph).

Second, the difference in predictive ability between the two store environments is remarkable. Apparently, it is more complex to define SOW in a general merchandise shopping environment than in a grocery shopping environment. While it is very likely that these differences can be explained by different purchasing patterns in both settings, more research is required to investigate and explain these differences.

As mentioned above, our feature selection procedure proved to be useful for the prediction of SOW since the multiple regression models achieved an increased performance with fewer predictive variables. In order to draw conclusions on which kind of data explains SOW, we focus on the univariate models' standardized parameter estimates for each of the predictors. Both store settings are very comparable in terms of the ranking of the explanatory variables, which suggests that our results may be generalizable to different, yet similar, store settings. Nearly all variables feature a significant influence that confirms the findings of Verhoef (2003), that past customer behavior explains most of customer share development. Intriguingly, the most valuable customer information for defining SOW is the variety of products purchased and responsiveness to direct mails. These variables can be detected within the top three predictors in both settings (the number of different product categories purchased during last year, Numcat and the percentage of times a purchase is made given that a leaflet was received). Our study is the first to show the great importance of this type of customer information when explaining SOW. In previous research, purchase depth (captured in variables such as the frequency and monetary value of previous purchases) has received more attention than purchase width (i.e., the purchase variety). However, our findings suggest that the predictive capacity of the latter type of information should not be neglected. Indeed, the more a customer is interested in purchasing a large variety of product categories, the stronger the relationship with the company and hence the more loyal the customer. This conclusion is consistent with the importance of this type of variable for predicting the strength of the customer's relationship and future developments in this relationship (Baesens et al. 2004). Representing another important predictor, the degree of response to leaflets is a signal of loyal customer behavior. This means that the level of past interest someone has shown in a company's communication is related to the fraction of that customer's total household budget that he or she spends at that company. Remarkably, variables related to customer spending or length-of-relationship are not found to be the best predictors, despite these being widely used in companies' reward schemes. The former type of variable shows up in the top 10 importance ranking. Their significance validates much past research that already suggested a relationship between loyalty and customers' spending level (Reichheld 2003). Moreover, buying more promotional products seems to be an indicator of increased SOW. An explanation for this surprising relation is that these variables correlate highly with the number of items bought and the frequency of visits. Customers who buy more items are expected to exhibit a higher absolute level of promotional purchases as well. Therefore, the parameter estimates of the multiple regression models are biased because of multicollinearity, and the univariate outcomes are driven by the number of items[14] and visit frequency[15] and not by the promotional nature of the products. This is supported by the insignificance of the percentage of promotional products bought (PercNumPromItems) in both settings. Furthermore, information concerning customers' last purchase date and the time between their purchases are significant in our models. The standard deviation of the time between customers' purchases also explains SOW. The effect suggests that regular customers, who show a low standard deviation, are more loyal to the store. This finding is in line with the loyalty definition of Buckinx and Van den Poel (2005), who incorporated this standard deviation to distinguish loyals from nonloyals. To our knowledge, this is the first study to confirm empirically the value of this behavior for classifying customers in accordance with the strength of their relationship. Surprisingly, the length of customers' relationship is ranked at the bottom of the results. Moreover, in the general merchandise store setting, only a marginal effect can be found. This supports the findings of Reinartz and

---

[14] Pearson Correlation Coefficients between 'Numitems' and 'NumPromItems': Grocery shopping .84 ($p < 0.01$); General merchandising .80 ($p < 0.01$).

[15] Pearson Correlation Coefficients between 'Frequency' and 'VisitsPromItems': Grocery shopping .96 ($p < 0.01$); General merchandising .93 ($p < 0.01$).

Kumar (2000), who doubt the value of lifetime duration for the characterization of valuable customers. Furthermore, the distance to the store is of minor importance for SOW.

Perhaps the most notable conclusion from this overview is that customers' spending, frequency and lifetime are not the only sources of information to explain SOW. This study shows the importance of other behavior when classifying customers according to their SOW. These findings point to the limited ability of currently used criteria to approximate customer SOW. The significant explanatory power of just about all variable types explains why our predicted SOW measure is more efficient in rewarding loyalty benefits than spending and lifetime. The more relevant customer behavior is taken into account, the better SOW can be approximated and the better the benefits related to loyal customers can be rewarded.

6.5 Limitations and Directions for Further Research

As in any other study, this study has its limitations and encourages further research on the issue and related topics.

First, although we validated this study in two different store settings, we cannot claim that our findings can be generalized to all environments. The results show small differences between the store formats considered: some hypotheses that are supported in the grocery setting are not supported, or only partially supported, in the general merchandise setting. Therefore, further research is needed in order to confirm our results in other industries—not necessarily restricted to consumer markets.

Second, our predictive model included little demographic customer information to explain SOW. Only the customers' distance to the store was incorporated. Since the European store chain that provided the data does not collect this type of information when customers register, no social demographics were at our disposal for the predictive model. Therefore, the predictive ability of our models might even increase when demographics are available from the company's internal data files.

Third, in this study, we provide evidence that loyalty, measured by SOW, can be predicted from the company's internal data records to an extent where it provides a more efficient criterion for rewarding loyalty benefits than spending or length of relationship. Hence, we have only shown that it is feasible to reward customers for their loyalty, and that the currently designed reward schemes do not fully reward loyalty. Indeed, in the present study, we were unable to test the effect of rewarding customers based on different reward criteria in the field. To this end, an economic decision about the most appropriate reward criterion would have to reside on a full cost–benefit analysis, whereby all consequences and benefits related to the reward criteria are quantified. Further increasing complexity, it is not impossible that the optimal reward program may be constructed by forming a segmented reward criteria approach, using different rewards for different customer groups—based on their scores on different reward criteria. However, considering the involvement of customers in reward programs and the need for clear communication about the reward criterion, companies are extremely reluctant to perform such a real-life test.

Finally, rewarding customers for their predicted SOW can prove to be difficult to communicate to the total customer base. An operational advantage of the currently used schemes lies in the fact that customers can trust the objectivity of the system: every dollar spent is translated into a certain reward. However, the application of SOW as a reward criterion does not necessarily imply that successful current systems should be changed. A potential solution would be to maintain the current reward systems and in addition target those customers who are highly loyal but are currently not rewarded for their loyalty, in order to prevent these customers from weakening their relationship owing to a feeling of neglect.

To conclude, a number of further studies can be designed to determine the full potential of using predicted SOW as a (complementary) reward criterion.

REFERENCES

Baesens Bart, Geert Verstraeten, Dirk Van den Poel, Michael Egmont-Petersen, Patrick Van Kenhove and Jan Vanthienen (2004), "Bayesian Network Classifiers for Identifying the Slope of the Customer Lifecycle of Long-Life Customers," *European Journal of Operational Research*, 156, 508–523.

Bolton, Ruth N., P.K. Kannan and Matthew D. Bramlett (2000), "Implications of Loyalty Program Membership and Service Experiences for Customer Retention and Value," *Journal of the Academy of Marketing Science*, 28 (1), 95–108.

Brodie, Roderick J., Nicole E. Coviello, Richard W. Brookes and Victoria Little (1997), "Towards a Paradigm Shift in Marketing? An Examination of Current Marketing Practices," *Journal of Marketing Management*, 13 (5), 383–406.

Buckinx, Wouter and Dirk Van den Poel (2005), "Customer Base Analysis: Partial Defection of Behaviourally-Loyal Customers in a Non-Contractual FMCG Retail Setting," *European Journal of Operational Research,* 164(1), 252-268.

Bult, Jan Roelf and Tom Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science*, 14 (4), 378–394.

Cohen, Jacob and Patricia Cohen (1983), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.

De Wulf, Kristof, Gaby Odekerken-Schröder and Dawn Iacobucci (2001), "Investments in Consumer Relationships: A Cross-Country and Cross-Industry Exploration," *Journal of Marketing*, 65 (October), 33–50.

Dick, Alan S. and Kunal Basu (1994), "Customer Loyalty: Toward an Integrated Conceptual Framework," *Journal of the Academy of Marketing Science*, 22 (2), 99–113.

Dowling, Grahame R. and Mark Uncles (1997), "Do Customer Loyalty Programs Really Work?" *Sloan Management Review*, 38 (Summer), 71–82.

East, Robert, Patricia Harris, Gill Willson and Wendy Lomax (1995), "Loyalty to Supermarkets," *International Review of Retail, Distribution and Consumer Research*, 5 (1), 99–109.

Feinberg, Fred F., Aradhna Krishna and John Z. Zhang (2002), "Do We Care What Others Get? A Behaviorist Approach to Targeted Promotions," *Journal of Marketing Research*, 39 (August), 277–291.

Festinger, Leon (1954), "A Theory of Social Comparison Processes," *Human Relations*, 7, 117–140.

Furnival, George M. and Robert W. Wilson (1974), "Regressions by Leaps and Bounds," *Technometrics*, 16, 499–511.

Goutte, Cyril (1997), "Note on Free Lunches and Cross-Validation," *Neural Computation*, 9, 1245–1249.

Hallberg, Garth (2004), "Is Your Loyalty Programme Really Building Loyalty? Why Increasing Emotional Attachment, not Just Repeat Buying, is Key to Maximising Programme Success," *Journal of Targeting, Measurement and Analysis for Marketing*, 12 (3), 231–241.

Irwin, Julie R. and Gary H. McClelland (1991), "Misleading Heuristics and Moderated Multiple Regression Models," *Journal of Marketing Research*, 38 (February), 100–109.

Jones, Thomas O. and Earl W. Sasser (1995), "Why Satisfied Customers Defect," *Harvard Business Review*, November–December, 87–99.

Jöreskog, Karl G. and Dag Sörbom (1995), *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.

Keiningham, Timothy L., Tiffany Perkins-Munn and Heather Evans (2003), "The Impact of Customer Satisfaction on Share-of-Wallet in a Business-to-Business Environment", *Journal of Service Research*, 6 (1), 37-50.

Kivetz, Ran and Itamar Simonson (2003), "The Idiosyncratic Fit Heuristic: Effort Advantage as a Determinant of Consumer Response to Loyalty Programs," *Journal of Marketing Research*, 40 (4), 454–467.

Latham, Gary P. and Edwin A. Locke (1991), "Self Regulation Through Goal Setting," *Organizational Behavior and Human Decision Processes*, 50 (2), 212–247.

Macintosh, Gerrard and Lawrence S. Lockshin (1997), "Retail Relationships and Store Loyalty: A Multi-Level Perspective," *International Journal of Research in Marketing*, 14 (5), 487–497.

Mägi, Anne W. (2003), "Share of Wallet in Retailing: the Effects of Customer Satisfaction, Loyalty Cards and Shopper Characteristics," *Journal of Retailing*, 79, 97–106.

McMullan, Rosalind and Audrey Gilmore (2002), "The Conceptual Development of Customer Loyalty Measurement: A Proposed Scale," *Journal of Targeting, Measurement and Analysis for Marketing*, 11 (3), 230–243.

Nicholls, John G. (1989), *The Competitive Ethos and Democratic Education*. Cambridge, MA: Harvard University Press.

Nunnally, Jum C. (1978), *Psychometric Theory*. New York: McGraw-Hill.

Reichheld, Frederick F. (2003), "The One Number You Need to Grow," *Harvard Business Review*, December, 46–54.

—— and Earl W. Sasser (1990), "Zero Defections, Quality Comes to Services," *Harvard Business Review*, 68 (September-October), 105–11.

Reinartz, Werner J. and Vikas Kumar (2000), "On The Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (October), 17–35.

—— and ——(2002), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (January), 77–99.

Reynolds, Kristy E. and Mark J. Arnold (2000), "Customer Loyalty to the Salesperson and the Store: Examining Relationship Customers in an Upscale Retail Context," *Journal of Personal Selling and Sales Management*, 20 (2), 89–98.

Rosenberg, Larry J. and John A. Czepiel (1984), “A Marketing Approach to Customer Retention,” *Journal of Consumer Marketing*, 1, 45–51.

Schmittlein, David and Robert A. Peterson (1994), “Customer Base Analysis: An Industrial Purchase Process Application,” *Marketing Science*, 13 (1), 41–67.

Sharp, Bryon and Anne Sharp (1997), “Loyalty Programs and Their Impact on Repeat-Purchase Loyalty Patterns,” *International Journal of Research in Marketing*, 14 (5), 473–486.

Srinivasan, Srini S., Rolph Anderson and Kishore Ponnavolu (2002), “Customer Loyalty in E-commerce: An Exploration of its Antecedents and Consequences,” *Journal of Retailing*, 78, 41–50.

Steenkamp, Jan Benedict E.M. and Hans C.M. van Trijp (1991), “The Use of LISREL in Validating Marketing Constructs,” *International Journal of Research in Marketing*, 8, 283–299.

Verhoef, Peter C. (2003), “Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development,” *Journal of Marketing*, 67 (October), 30–45.

——, Philip Hans Franses and Janny C. Hoekstra (2002), “The Effect of Relational Constructs on Customer Referrals and Number of Services Purchased from a Multiservice Provider: Does Age of Relationship Matter?” *Journal of the Academy of Marketing Science*, 30 (3), 202–216.

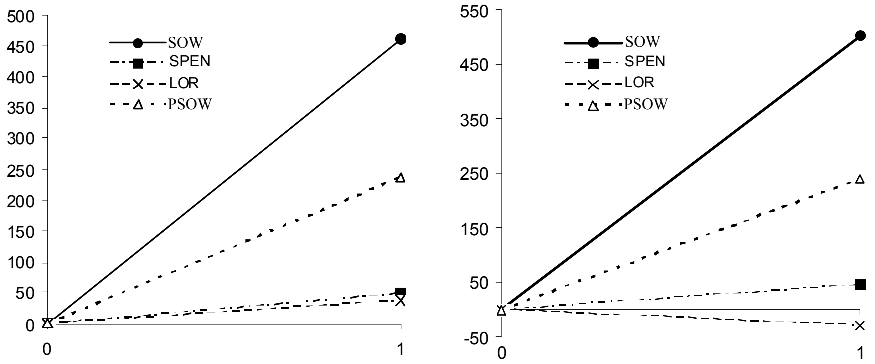
Wangenheim, Florian and Tomás Bayón (2004), “The Contribution of Word-of-Mouth Referrals to Economic Outcomes of Service Quality and Customer Satisfaction,” *Journal of the Academy of Marketing Science*, forthcoming.

Whyte, Randall (2004), “Frequent Flyer Programmes: Is it a Relationship, or do the Schemes Create Spurious Loyalty?” *Journal of Targeting, Measurement and Analysis for Marketing*, 12 (3), 269–280.

Zeithaml, Valarie A., Leonard L. Berry and A. Parasuraman (1996), “The Behavioral Consequences of Service Quality,” *Journal of Marketing*, 60 (April), 31–46.

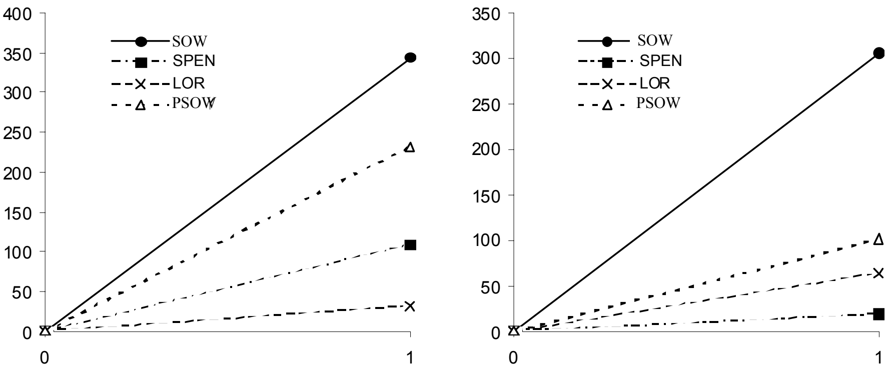
APPENDICES

Appendix 5.A: Relationship between Rewards Received and Benefits Delivered



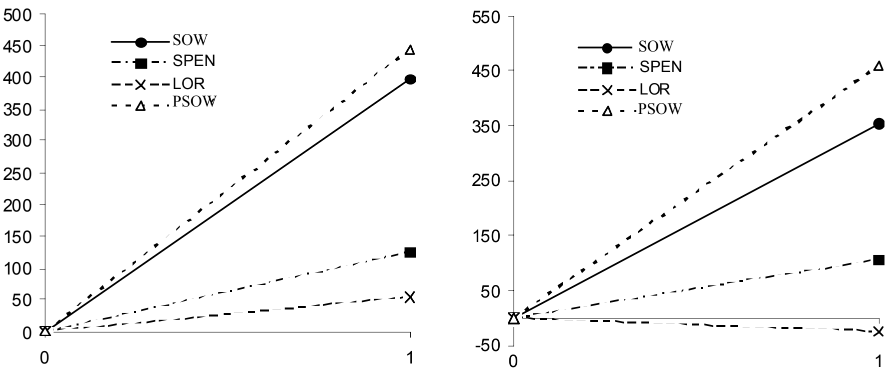
Setting G **Setting M**

A: Word-of-mouth



Setting G **Setting M**

B: Price Insensitivity



Setting G **Setting M**

C: Purchase Intentions

CHAPTER VI

SUCCESFULLY PREDICTING CUSTOMER LOYALTY USING COMPANY-INTERNAL TRANSACTIONAL DATABASE INFORMATION[16]

---

[16] This chapter is based on the following reference: Wouter Buckinx, Geert Verstraeten, Dirk Van den Poel, 2005. Succesfully predicting customer loyalty using company-internal transactional database information, ready for submission.

CHAPTER VI:

SUCCESFULLY PREDICTING CUSTOMER LOYALTY USING COMPANY-INTERNAL TRANSACTIONAL DATABASE INFORMATION

1. INTRODUCTION

In the two latest decades, Customer Relationship Management (CRM) has grown to be one of the major trends in marketing, both in academia and in practice. This evolution took form in a dramatic shift in the domain, evolving from transaction-oriented marketing to relationship-oriented marketing [12], and builds strongly on the belief that it is several times less demanding – i.e. expensive – to sell an additional product to an existing customer than to sell the product to a new customer [23]. Hence, it has been argued that it is particularly beneficial to build solid and fruitful customer relationships, and in this discourse, customer loyalty has been introduced as one of the most important concepts in marketing [20].

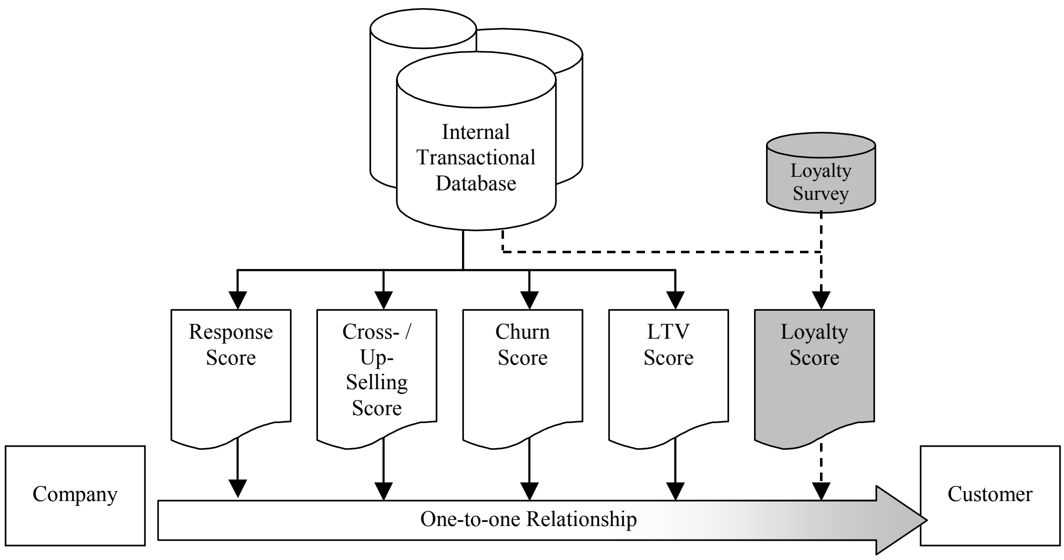
From an analytical point of view, several tools have emerged in recent years that enable companies to strengthen their relationships with customers. Besides, the rise of new media such as the World Wide Web, and the continuous improving technological conditions have further increased the opportunities to communicate in a more direct, one-to-one manner with customers [26]. Response modeling – i.e. predicting whether a customer will reply to a specific offer, leaflet or product catalogue – represents the most central application in this domain, and serves as a tool to manage customer relationships. Indeed, it would be beneficial for the company-customer relationship that the latter party would receive only information that is relevant to him/her, hence allowing the company to present only those offers for which the individual customer shows a high response probability [2]. Related to this, cross-selling analysis is involved with finding the optimal product to offer to a given customer [7, 15]. Additionally, upselling analysis is focused on selling more – or a more expensive version – of the products that are currently purchased by the customer. Both techniques share a similar goal, i.e. to intensify the customer relationship by raising the share of products that is purchased at the focal company, and to prevent these products from being purchased at competitive vendors. The fear of losing sales to competitors also features in churn analysis, which is focused on detecting customers exhibiting a large potential to abandon the existing relationship. Churn analysis has received great attention in the domain ever since it has been proven that even a small improvement in customer defection can greatly affect a company's future profitability [21, 27]. Finally, lifetime value (LTV) analysis is a widely used technique to predict the future potential of customers, in order to target only the most promising customers [13]. While these techniques can each serve individually to enhance customer relationships, it should be clear that additional advantages reside in the combination of these analytic techniques. Some attempts to integrate such techniques can be found in recent literature (see, e.g. [1, 14]).

2. THE NEED FOR PREDICTING CUSTOMER LOYALTY

In sum, we could state that both the focus on customer loyalty and the analytic tools described above have emerged from the CRM discourse. However, it is very unusual that actual customer loyalty is used to either devise or evaluate a company's targeted marketing strategies. The major cause of this deficiency lies most likely in the unavailability of information. Currently, while companies are maintaining transactional databases that store all details on any of a given customer's contacts with the focal company, these databases cannot capture the amount of products that this customer purchases at competing stores. Indeed, a recent study showed that only 7.5 % of companies involved in database marketing activities collect such purchase behavior [28]. Hence, the real behavioral loyalty of a certain customer is generally unavailable in the company's records, whereby the full potential of the customer (i.e., the total needs of the customer for products in the relevant category) is unknown to any specific company. However, this information could prove to be extremely valuable in different applications.

First, the knowledge of a customer's loyalty would be useful for improving CRM. We illustrate this with an example from a banking context. It would most likely be more lucrative to offer an additional savings product to a customer who has a high balance at the focal bank and at the same time has large amounts invested at other banking institutions, than to offer the savings product to a customer that has an equally high balance, but where all his/her money is invested at the focal bank. Secondly, a notion of a customer's loyalty could be used for adapting the usefulness of the model-building process. For example, currently, cross-selling models are being built on the total customer database, whereby the users will estimate the probability of purchasing this product *at the focal company*, whereas from a cross-sales point of view, it would be more interesting to estimate whether they are interested in the product category *in general*. To overcome this, it could be interesting to build a cross-selling model on loyal customers only, because only for these customers, their total product needs are known. In this context, when attempting to model the real – and total – product needs of customers, it might seem suboptimal to include unloyal customers into the analysis. Thirdly, the knowledge of a customer's loyalty and the evolution therein could be useful for evaluating the results of CRM-related investments, and monitoring whether certain actions lead to the desired results in the relevant customer segments.

Figure 6.1: Creating a loyalty score from transactional data and loyalty survey.



While such loyalty information can be obtained through a questionnaire, it would prove to be financially infeasible to obtain this information for each individual customer, especially when customers would have to be surveyed regularly in order to track changes in their loyalty profile. Consequently, in this paper, we will prove that it is sufficient to survey a sample of the company's customers, since we will combine the information stemming from the survey and the internal transactional database in order to create a loyalty score for all individual customers. Hence, as summarized in Figure 1, this score could provide additional information to the scores based on the transactional data only, and form a valuable expert tool for managing customer relationships.

The remainder of this paper is structured as follows. The next section covers the methodology used, and focuses on a description of the applied predictive techniques, the need for adequate cross-validation, and the variable-selection procedure we propose. Next, we will describe the data used for this study. In a subsequent section, we discuss the results of the proposed predictive modeling study. Finally, we end the paper with a section covering the conclusions and directions for further research.

3. METHODOLOGY

3.1 Predictive techniques

Technically, in this study, we will predict this loyalty for customers that do not belong to the surveyed sample by use of the data that are available for all customers, i.e. the transactional data. In essence this is a problem of predictive modeling. It is not our ambition to compare all possible predictive techniques. Instead, we will compare three techniques that show interesting differences and similarities. Because of the need for an accurate prediction as well as an understanding of the model – in order to explain the findings to management – we only considered models that were expected to (i) deliver adequate predictive performance on a validation set and (ii) provide an insight into the most important variables in the model. As a benchmark predictive technique, we have used a multiple linear regression (MLR) model [8], because of the widespread usage of this statistical technique in industry and academia. We compared this benchmark with two state-of-the-art techniques from the machine learning and data mining domain. First, given the widespread use of decision trees in prediction problems where the user seeks insight into the predictive process, we have implemented Random Forests (RF). This technique focuses on growing an ensemble of decision trees using a random selection of features to split each node (i.e. the random subspace method), where the final prediction is computed as the average output from the individual trees [4]. RF models have been argued to possess excellent properties for feature selection, and to avoid overfitting given that the number of trees is large [4]. In this approach, we will grow 5000 trees, as in other applications [11]. Finally, since Artificial Neural Networks (ANN's) have often been credited for achieving higher predictive performance, we selected MacKay's Automatic Relevance Determination (ARD) neural network because it additionally reveals a Bayesian hyperparameter per input variable, representing the importance of the variable [17]. To this end, the relevance of the features is detected by maximizing the model's marginal likelihood. We respected the author's view that a large number of hidden units should be considered in order to build a reliable model. The use of the ARD model is made possible using Markov Chain Monte Carlo techniques, hence avoiding overfitting due to the use of a Bayesian 'Occam's razor' while allowing an interpretation of the variables' importance [17].

3.2 Cross-validation

An important early topic in predictive modeling consists in validating the predictive power of a model on a sample of data that is independent of the information used to build the model. In this study, the limited number of observations in each of the two settings and the elaborate number of independent variables make it hard to split our data in an estimation and a hold-out validation set. As a consequence, we prefer a resampling method called leave-one-out cross-validation because it proves to be superior for small data sets [10]. Using this procedure, our data are divided into k subsets, where k is equal to the total number of observations. Next, each of the subsets is left out once from the estimation set and is then used to estimate a validation score. To compute the real-life power of the model, the final validation set is built by stacking together the k resulting validations and the predictive performance is computed on this stacked set. The performance of the model – on the estimation set as well as on the validation set – is evaluated by computing (i) the correlation between surveyed loyalty and its prediction, (ii) R^2, (iii) adjusted R^2, (iv) Mean Squared Error (MSE) and (v) the Root of the MSE (RMSE).

3.3 Variable selection

In the current study, it is likely that we can compute a large number of database-related variables in comparison with the number of observations (i.e. the number of respondents of this questionnaire). While both the RF and ARD models claim to avoid overfitting, this effect does provide a reasonable threat to the multiple regression model [8]. To overcome this problem, we will make use of a variable-selection technique. Thanks to this method, the dimensionality of the model can be reduced and redundant variables are removed, which is in favor of the model's performance. Additionally, a variable-selection procedure will allow us to gain insight in selecting the variables with the best predictive capacities, and allows us to interpret the parameter estimates due to the exclusion of multicollinearity.

Figure 6.2: Model selection and validation for the multiple linear regression model

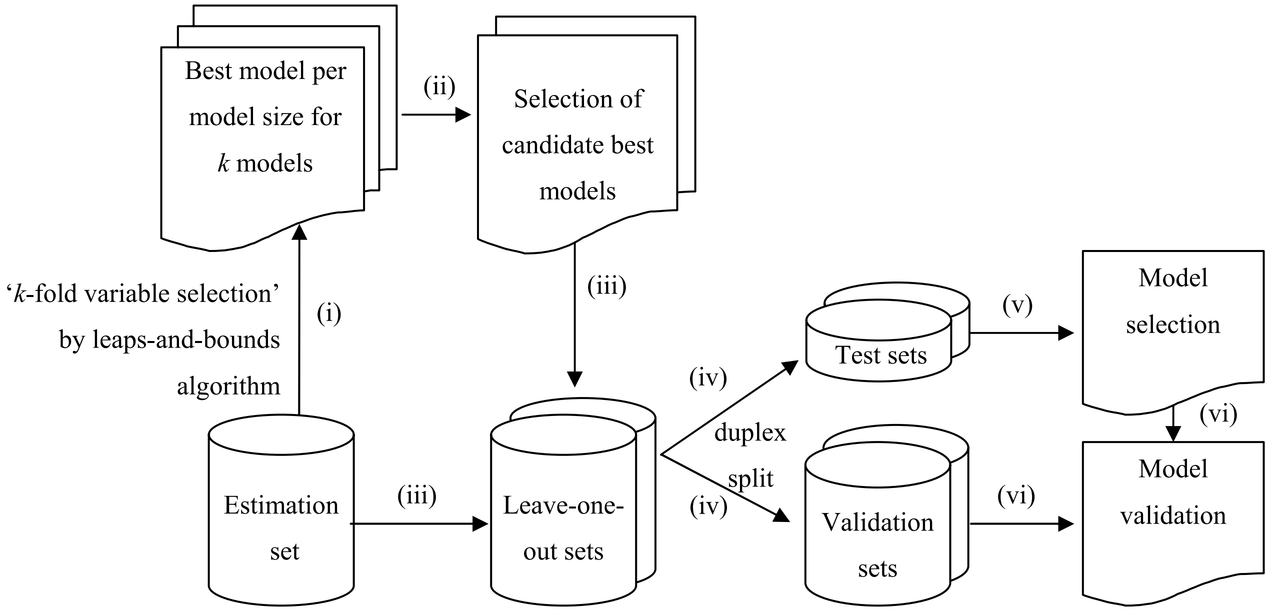


Figure 2 partitions the variable-selection procedure that was used in this study into six disjoint steps. In step (i), we apply the leaps-and-bounds algorithm proposed by Furnival and Wilson [9] on the estimation set. Their efficient technique identifies the model with the largest adjusted R^2 for any given model size (i.e. starting from the best model with only one variable to the full model) and at the same time avoids a full search of the variable space. However, because of the leave-one-out procedure described previously, in this case, we cannot simply perform this procedure on the total estimation set. Indeed, in order to allow for a validation of the model, the estimated models should be built when at least one observation is set aside for validation. Since it would be suboptimal to select this observation randomly, in this study we propose an iterative process in which we set aside one observation at a time, such that we create *k* new estimation sets, where *k* equals the total number of observations in the original estimation set. Hence, the outcome of this procedure – to which we refer as '*k*-fold variable selection' – will consist in a list of *k* best models per model size. Next, in step (ii) to ensure tractability and to avoid the choice of selecting an unstable model, we reduce this list by selecting, per model size, only those models that were 'winners' in at least 5% of the occasions. In step (iii), we create the leave-one-out predictions for each candidate model using the procedure described in the previous paragraph. In the following steps, we are concerned with selecting the best models, and validating the performance of these models. Because of this dual need, in step (iv) we divide the leave-one-out data set per candidate model into a test set containing 25 % of the observations, that will be used for model selection; and a validation set consisting of the remaining 75 % of the observations, that will be used for detecting the real predictive performance of the model. Considering both the importance of a good split and the low number of observations available, we do not perform a random split, but rather complete the division via the Duplex algorithm [24], which performs best in separating a dataset into two sets covering approximately the factor space. Concretely, here, this factor space is composed of the set of independent variables created for the study. Next, in step (v), based on the leave-one-out test set performance, we select the best-performing model per model size among the selection of candidate models. Additionally, we select the model with the highest overall performance. In the final step (vi), we validate the real predictive performance of the models selected in the previous step on the unseen data.

4. DATA DESCRIPTION

We use data from two retail stores belonging to the same large European chain which were considered, according to management, to be representative for the entire chain. The stores carried a product assortment normally associated with grocery stores (e.g., food and beverages, cosmetics, laundry detergents, household necessities). Detailed purchase records were tracked for a period of 51 months and a summarized customer table was available that tracked basic customer demographics as well as date of first purchase.

Table 6.1: Description and predictive performance of variables used.

Variable	Description	MLR Standardized Parameter Estimates		RF Variable Importance	ARD Alpha (Importance)
Spending_1M	Spending during last month.	0.3540	***	0.0086	21.49
Spending_6M	Spending during last six months.	0.4582	***	0.1136	13.00
Spending_1Y	Spending during last year.	0.4789	***	0.2246	15.58
Spending_2Y	Spending during last two years.	0.4742	***	0.0228	23.63
Spending	Spending in total history.	0.4714	***	0	32.08
NumItems	Number of product items bought.	0.4705	***	2.3071	16.27
Spending_Weight	Spending in products that need to be weighted by the customer.	0.4395	***	0.2985	17.52
rSpend_Freq	Average Spending per visit.	0.1785	***	0.0055	7.11
rSpend_Lor	Spending relative to the length of the customer's relationship.	0.4726	***	0.4104	0.16
Frequency_1M	Number of purchases during last month.	0.3477	***	0	2.41
Frequency_6M	Number of purchases during last six months.	0.4356	***	0.035	3.76
Frequency_1Y	Number of purchases during last year.	0.4455	***	0.0544	3.91
Frequency_2Y	Number of purchases during last two years.	0.4494	***	0	2.77
Frequency	Number of purchases in total history.	0.4389	***	0	3.87
Recency	Number of days since last purchase.	–0.2035	***	0	24.44
Ipt	Average number of days between store visits.	–0.2965	***	0.6045	17.23
Std_Ipt	Standard deviation of the number of days between the purchases.	–0.3227	***	0.292	13.20
Lor	Length of customer relationship.	0.0940	***	0	29.92
Numcat_LY	Number of different product categories purchased from during last year.	0.5221	***	0.543	6.32
Numcat_2Y	Number of different product categories purchased from during last two years.	0.4770	***	0.2001	3.09
Numcat_3Y	Number of different product categories purchased from during last three years.	0.4460	***	0.1434	5.27
Numcat	Number of different product categories purchased from during the total history.	0.4805	***	0.2233	10.28
Neg_Inv	Dummy to indicate if the customer ever had a negative invoice (1/0).	0.2919	***	0.1115	2.42
Ret_Item	Dummy to indicate if the customer ever returned an item (1/0).	0.2656	***	0.0293	1.56
Returns	Total value of returned goods.	0.1572	***	0	11.90
NumPromItems	Number of items bought that appeared in company's promotion leaflet.	0.4539	***	0.9065	9.62
SpenPromItems	Money spent on products that appeared in promotion leaflet.	0.4572	***	0.0064	11.79
Visitspromitems	Number of visits on which a product is bought that appeared in the promotion leaflet.	0.4680	***	0.0342	5.35
PercNumPromItems	Percentage of products bought that appeared in leaflet.	0.0139		0.06	8.48
PercResp_Leaf	Percentage of times a purchase is made given that a promotion leaflet was received.	0.4792	***	0	0.22
PercResp_Noleaf	Percentage of times a purchase is made given that no promotion leaflet was received.	0.3098	***	0	1.32
Perc_Noleaf_Freq	PercResp_Noleaf divided by shopping frequency	–0.2235	***	0.1258	2.57
MoreThanOnce	Number of times that a customer visits more than once within the same promotion period.	0.4308	***	0	2.92
PercMoreThanOnce	MoreThanOnce divided by the number of times a customer bought in a promotion period.	0.2940	***	0	0.34
Distance	Distance to the store.	–0.1265	***	0.0457	6.07

4.1 Computation of database-related variables

It is important to mention that all transactions could be linked to customers, as the store requires use of a customer identification card. In total, 35 independent variables are computed, that are related to the following topics: (i) monetary spending, (ii) frequency of purchasing, (iii) recency of last purchase, (iv) length of the customer-company relationship, (v) interpurchase time, (vi) returns of goods, (vii) purchase variety, (viii) promotion sensitivity, (ix) responsiveness on mailings and (x) distance to the store. The inclusion of these variables was mainly based on previous literature in the domain of predicting the strength of the relationship between a company and its customers [1, 5, 6, 22, 25]. Table 1 summarizes all these variables, together with a brief description of how they are calculated.

4.2 Loyalty survey

In addition to these transactional data, a self-administered survey was used as a complementary data collection method. Data collection took place in each of the retail stores mentioned previously. Surveys were randomly distributed to customers during their shopping trips, and customer identification numbers were recorded for all customers who received a questionnaire.

Table 6.2: Wording of the items of the loyalty scale.

Item 1	Buy (much less … much more) grocery products at XYZ than at competing stores.
Item 2	Visit other stores (much less frequently … much more frequently) than XYZ for your grocery shopping (–).
Item 3	Spend (0% … 100%) of your total spending in grocery shopping at XYZ.

A customer's behavioral loyalty was determined as a composite measure by comparing a customer's spending at the retailer with their total spending in the relevant product category. As a first item, and similar to [16], the percentage of purchases made in the focal supermarket chain versus other stores was assessed on an 11-point scale that ranged from 0% to 100% in 10% increments (i.e., 0%, 10%, 20%, and so on). Additionally, two seven-point

Likert-type items assessed the shopping frequency of the customers for the focal store when compared to other stores. We pretested the questionnaire and refined it on the basis of pretest results. Table 2 gives the exact wording of the items used. After rescaling the second item (due to its expected negative correlation with both other items), we standardized the 3 loyalty-related questions, and averaged them to represent the behavioral loyalty construct.

5. RESULTS

5.1 Survey response

Of the 1500 distributed questionnaires, we received 878 usable responses (i.e. a ratio of usable response of 58.33%). We successfully tested for nonresponse bias by comparing database variables such as spending, frequency of visiting the store, interpurchase time, length-of-relationship and response behavior towards companies' mailings between respondents and nonrespondents. A usable response had all fields completed, and the respondent could be successfully linked to his or her transaction behavior in the customer database. We tested construct reliabilities of the loyalty scale by means of Cronbach's coefficient alpha. The resulting coefficient of .871 clearly exceeds the .7 level recommended by [19], which proves it is a reliable scale, especially given the fact that reverse coding was used to measure one item of the 3-item scale.

5.2 Predictive Performance

In terms of predictive performance, in Table 3, we compare the results of the different models. Considering the MLR models, we compared the full model with the final model resulting from the variable-selection procedure described previously, which resulted in a selection of just 4 variables. Regarding the results from the RF model, all variables were introduced, yet only 24 variables were selected by the technique. In terms of the ARD model, we reached an optimal performance by using 24 hidden units. No variables were selected by the latter technique so each variable contributes, to some extent, to the predictive performance.

Table 6.3: Model performances.

	MLR Full Model (v=35)		MLR Final Model (v=4)		RF Full Model (v=35)		ARD Full Model (v=35)	
	Estimation	Validation	Estimation	Validation	Estimation	Validation	Estimation	Validation
R	0.5664	0.5107	0.5535	0.5442	0.5186	0.5238	0.5714	0.4935
R²	0.3208	0.2608	0.3064	0.2962	0.2689	0.2744	0.3265	0.2435
R²adj	0.2926	0.2301	0.3032	0.2919	0.2385	0.2442	0.2985	0.2121
MSE	0.5586	0.6107	0.5502	0.5569	0.6023	0.5969	0.5586	0.6237
RMSE	0.7474	0.7815	0.7417	0.7463	0.7761	0.7726	0.7474	0.7898

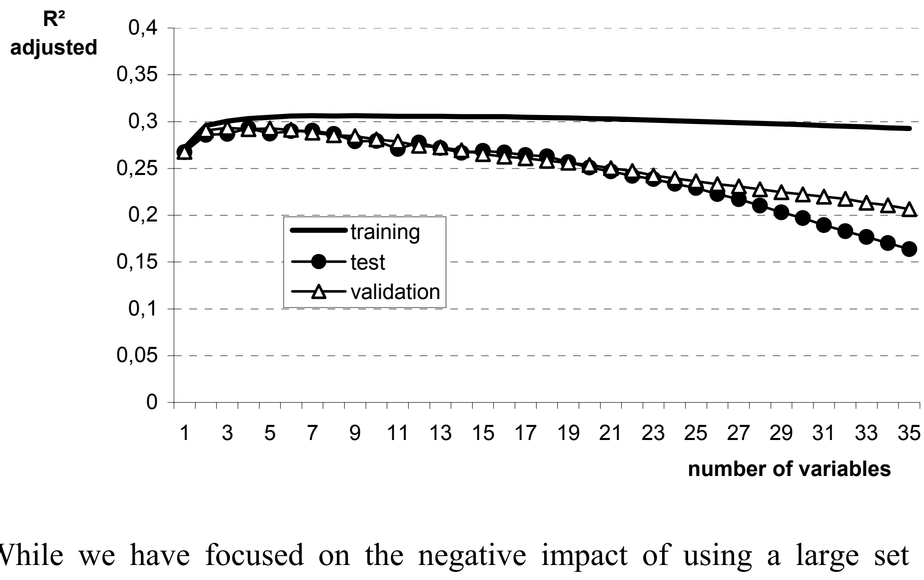
Different interesting conclusions can be drawn from Table 3. First, it is clear that – as was expected – overfitting prevails in the MLR model, and does not appear in the RF model. This finding is in line with Breiman's initial claims [4] as well as findings by other authors [5]. Indeed, the adjusted R^2 of the full MLR model drops from 0.2926 on the estimation set to 0.2301 on the validation set, which introduces skepticism on the validity of this model. Second, the variable-selection procedure we described previously succeeds in reducing the negative impact related to overfitting. Indeed, the difference between the adjusted R^2 on the estimation set (0.3032) versus the test set performance (0.2919) is sufficiently small. Thirdly, contrarily to what might have been expected using the Bayesian 'Occam's razor' [17], the ARD model also proves to be sensitive to overfitting, as the performance on the estimation set is substantially higher than the performance after cross-validation. Fourth, given that an efficient variable-selection procedure is performed to the regression model, this model clearly outperforms the other models in terms of predictive performance. Fifth, in order to test whether this result is significant, we tested whether the correlations (R) differ significantly using a test of the difference of dependent samples described in [8, p. 57]. From this test, we can conclude that the MLR model performs significantly better than the RF ($t = 2.57$, $p = 0.01022$) and ARD models ($t = 2.68$, $p = 0.00747$). However, the difference in performance between the RF and ARD models is not significant ($t = 1.39$, $p = 0.16421$).

In sum, given that the adjusted coefficient of determination of the final MLR model is fairly high (0.2919) for cross-sectional data, and given its significance ($F = 96.39$, $p = <.0001$), we can state that it is possible to predict a customer's loyalty to a reasonable degree from the internal transactional database using a regression model – provided that an elaborate variable-selection procedure is performed. Because of the importance of the latter procedure, we discuss its implications in detail in the following paragraph.

5.3 Usefulness of the variable-selection technique

In Figure 3, we illustrate the effect of the variable-selection technique by plotting the estimation, test and validation performance of the best-performing model per model size. While the adjusted R^2 of the estimation data set does not decrease substantially as the number of variables increases, the validity of these models is severely hampered. However, the splitting of the leave-one-out sample into a test and validation set does clearly allow us to select the best-performing model and validate this model, while efficiently exploiting the available observations. Hence, the test set reached its highest level with the use of only four variables, whereby overfitting is reduced. The Appendix 6.A features a similar graph illustrating overfitting in terms of the RMSE.

Figure 6.3: Evidence of overfitting when the number of variables is increased (adjusted R²).



While we have focused on the negative impact of using a large set of variables on the predictive performance of the model, an additional threat resides in the occurrence of multicollinearity. Indeed, it is likely that, when using a large number of predictors, several predictors that are jointly used might be severely correlated. Hence, the affected parameter estimates might become unstable and may exhibit high standard errors, reflecting the lack of properly conditioned data [3]. In this section, we will illustrate the existence of multicollinearity graphically. To this goal, we follow the procedure of Belsley et al [3], and hence we present the evolution of the condition index of the best performing model per model size in Figure 4. Considering the author's informal suggestion that, at an index larger than 15, weak dependencies may start to affect the regression estimates [3, p. 153], those models incorporating more than 7 variables might exhibit unstable estimates and high standard errors. In order to validate this rule of thumb we have attempted to provide a graphical representation of the stability of the estimates. To this effort, we have computed the parameter estimates of all variables when they are used separately in univariate predictive models. Next, we compared the signs of these parameters – to which we refer as the 'correct' signs – with the signs of the best multiple regression models, and we plotted the percentage of 'correct' signs in Figure 5. The results confirm the previously offered rule-of-thumb, as at least some parameter signs differ in models that contain more than 7 variables. Hence, in these models, the parameter estimates can be considered as unstable.

Figure 6.4: Detecting multicollinearity by the condition index.

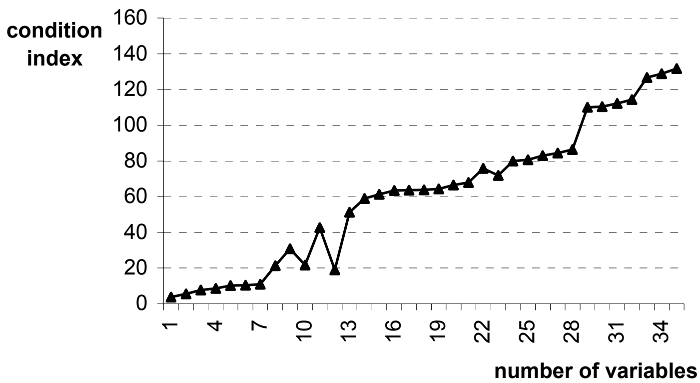
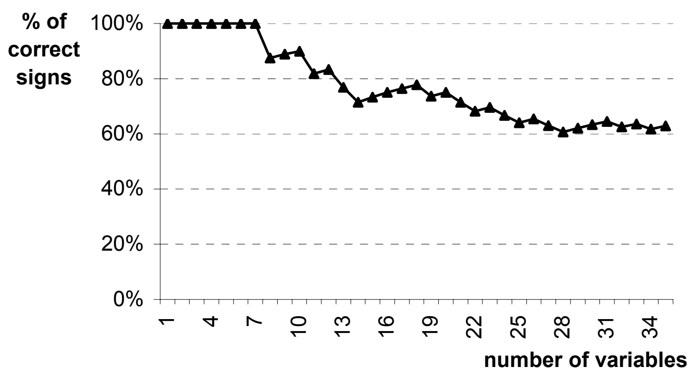


Figure 6.5: An illustration of the effect of multicollinearity on the parameter signs.



To conclude this section, the full model – containing all variables – shows evidence of multicollinearity that is manifested in a condition index of 131.6 and the fact that only 63% of the parameter signs correspond to their univariate counterparts. However, these problems seem efficiently solved in the final model – containing only the four selected variables – showing a condition index of only 8.5 and a proportion of 100% ‘correct’ parameter signs.

5.4 Variable Importance

In order to discuss the importance of the variables to predict behavioral loyalty, we will look both at the univariate performances as well as the inclusion of these variables into the MLR models. First, in terms of the univariate importances, Table 1 illustrates that the different models emphasize different variables. For example, in the ARD model, the length of relationship is considered as the second most important variable, while in the MLR model it features as the second least important variable, and the variable was not selected in the RF model. The difference between the models can be evaluated more formally through the computation of the correlation between the variable importances. The correlation between the MLR model and RF model is 0.08862 (p=0.6127), between the MLR model and the ARD model -0.16933 (p=0.3308), and between the RF model and the ARD model 0.12051 (p=0.4905), so we conclude that the models really emphasize different predictors. Since the MLR model outperforms the other models, in the remainder of this paragraph, we will focus on the importance of variables according to the MLR model. From the univariate performances, we note that the purchase variety clearly forms the best predictors of loyalty. However, several groups of variables have only a slightly lower performance. Variables related to the spending, frequency, promotion behavior and response on mailings all have a good predictive performance. The other variables, such as recency, interpurchase time, length of relationship, average spending per visit, returns of goods and distance to the store clearly exhibit lower univariate predictive performance.

An additional insight can be gained from the inclusion of the variables in the best performing multivariate models. Hence, in Table 4, we present the variables of the selected models that contain up to seven variables. This confirms the fact that purchase variety, spending and a customer’s response on mailing folders present the most useful information for predicting behavioral loyalty.

Table 6.4: Parameter estimates of the best predictive models.

Number of variables	Variable	Standardized Estimate	t Value	Pr > |t|	R²adj Validation
1	Intercept	0	-15.69	<.0001	0.2678
	Numcat_LY	0.5221	18.12	<.0001	
2	Intercept	0	-14	<.0001	0.2905
	Spending	0.2154	5.56	<.0001	
	Numcat_LY	0.3751	9.67	<.0001	
3	Intercept	0	-14.3	<.0001	0.2934
	Numcat_LY	0.2979	6.16	<.0001	
	PercResp_Leaf	0.1240	2.64	0.0084	
	rSpend_Lor	0.1859	4.59	<.0001	
4	Intercept	0	-13.62	<.0001	0.2919
	Spending_Weight	0.0887	2.12	0.0343	
	Numcat_LY	0.2741	5.54	<.0001	
	PercResp_Leaf	0.1145	2.43	0.0151	
	rSpend_Lor	0.1468	3.31	0.001	
5	Intercept	0	-11.91	<.0001	0.2926
	Spending_Weight	0.0994	2.41	0.0162	
	Numcat_LY	0.2389	4.54	<.0001	
	NumItems	0.1017	2.16	0.031	
	PercResp_Leaf	0.1651	3.07	0.0022	
	rSpend_Freq	0.0739	2.21	0.027	
6	Intercept	0	-8.46	<.0001	0.2911
	Spending_Weight	0.1024	2.48	0.0133	
	Numcat_LY	0.2193	4.06	<.0001	
	NumItems	0.1043	2.22	0.0269	
	PercResp_Leaf	0.1487	2.72	0.0066	
	rSpend_Freq	0.0732	2.2	0.0284	
	Std_Ipt	-0.0553	-1.64	0.1007	
7	Intercept	0	-8.53	<.0001	0.2881
	Spending_Weight	0.1009	2.44	0.0147	
	Neg_Inv	0.0396	1.2	0.2313	
	Numcat_LY	0.2172	4.03	<.0001	
	NumItems	0.0990	2.09	0.0365	
	PercResp_Leaf	0.1367	2.46	0.0141	
	rSpend_Freq	0.0769	2.3	0.0219	
	Std_Ipt	-0.0520	-1.54	0.1237	

6. CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

Following the prevalence of the CRM discourse, companies have started to realize the value of loyal customers, and have acquired the competences to manage customer relationships through targeted communications. Intriguingly however, these relationships are currently managed almost unanimously based on transactional data (such as recency, frequency, and monetary value of a customer) while the behavioral loyalty and hence the full potential of a customer is generally unavailable. In this study, we have constructed a reliable three-item scale to measure behavioral loyalty, and we have proven that it is possible to predict a customer's behavioral loyalty to a reasonable degree based on his/her transactional information. Hence, we have provided a viable methodology for building a loyalty score for all customers, based on a limited sample of customers for which behavioral loyalty was surveyed. This additional customer knowledge can be useful in many marketing applications within the area of customer relationship management, be it direct marketing, model building and customer evaluation.

To this end, we compared three techniques that have been argued to show a good predictive performance and an interpretation of the importance of the predictors. More specifically, we compared multiple linear regression with two state-of-the-art techniques, namely Breiman's regression forests and MacKay's automatic relevance determination. The predictive modeling we propose in this study is different from the general situation of predicting transactional behavior by use of historic transactional behavior in the sense that here, the target variable is only known for a limited set of customers. Because overfitting is more likely to occur when the observations are limited compared to the number of variables, and since overfitting is a well-acknowledged problem in multiple linear regression, the major contribution of this study lies in designing an effective variable-selection procedure. Hence, considering the limited sample size, we propose a model selection and validation procedure that is based on the leaps-and-bounds algorithm using an intelligent split of a leave-one-out cross-validation sample. In a real-life study, we show that this procedure effectively increases the validation performance to an extent that the linear regression model outperforms the other models in terms of predictive accuracy, and that multicollinearity is removed to an adequate degree in the resulting model, allowing for a sound interpretation of the parameters. Hence, we show that purchase variety is the best performing predictor of behavioral loyalty, and that a customer's spending, frequency, promotion behavior, response to mailings and regularity of purchasing all provide useful information to deliver an adequate prediction of a customer's behavioral loyalty.

As any other study, this study has its limitations which may lead to further research. First of all, in this paper it was not our ambition to compare all possible predictive modeling techniques. Hence, it is not excluded that other techniques serve even better to predict behavioral loyalty. Instead, we have confirmed that a proper use of sound statistical techniques is at least able to compete with two state-of-the-art predictive techniques. Second, contrarily to what was expected, we gained evidence of overfitting in the ARD model. While again it was not the focus of this specific study, this finding seems at least intriguing. Hence, further research might focus on performing a (possibly similar) variable-selection technique for the ARD model to account for the overfitting that was detected. Thirdly, in this case, we have used a leave-one-out cross-validation sample. It is not unlikely, however, that for future usage, the procedure could be applied in a more resource-efficient way by applying a leave-*k*-out cross-validation, where *k* is increased while carefully monitoring the validity of the results. Finally, in this procedure, due to financial constraints, it was not possible to perform an out-of-sample cross-validation to account for any possible model drift. Indeed, a subsequent survey of the behavioral loyalty would prove useful in evaluating the stability of the model for future loyalty predictions.

ACKNOWLEDGEMENTS

The authors are grateful to Leo Breiman [4] for the public availability of the random forests software, and to Ian T. Nabney [18] for his implementation of the technique of automatic relevance determination using neural networks.

REFERENCES

[1] Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P., & Vanthienen, J. (2004). Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. European Journal of Operational Research, 156 (2), 508-523.

[2] Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J. & Dedene, G. (2002). Bayesian Neural Network Learning for Repeat Purchase Modelling in Direct Marketing. European Journal of Operational Research, 138 (1), 191-211.

[3] Belsley, D.A., Kuh, E., & Welsch, R.E. (1980). Regression Diagnostics. New York: John Wiley & Sons, Inc.

[4] Breiman, L. (2001). Random Forests. Machine Learning, 45, 5-32.

[5] Buckinx, W. & Van den Poel, D. (2005). Customer Base Analysis: Partial Defection of Behaviorally-Loyal Clients in a Non-Contractual FMCG Retail Setting. European Journal of Operational Research, 164 (1), 252-268.

[6] Bult, J.R. & Wansbeek, T. (1995). Optimal Selection for Direct Mail. Marketing Science, 14 (4), 378–394.

[7] Chintagunta, P.K. (1992). Estimating a multinomial probit model of brand choice using the method of simulated moments. Marketing Science, 11 (4), 386-407.

[8] Cohen, J. & Cohen, P. (1983). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (2nd ed). Hillsdale, NJ: Erlbaum.

[9] Furnival, G.M. & Wilson, R.W. (1974). Regressions by Leaps and Bounds. Technometrics, 16, 499–511.

[10] Goutte, C. (1997). Note on Free Lunches and Cross-Validation. Neural Computation, 9, 1245–1249.

[11] Geng, W., Cosman, P., Berry, C.C., Feng, Z. & Schafer, W.R. (2004). Automatic Tracking, Feature Extraction and Classification of C Elegans Phenotypes. IEEE Transactions on Biomedical Engineering, 10 (51), 1811-1820.

[12] Grönroos, C. (1997). From marketing mix to relationship marketing—towards a paradigm shift in marketing. Management Decision, 35 (4), 839-843.

[13] Hwang, H., Jung, T. & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. Expert Systems with Applications, 26 (2), 181-188.

[14] Jonker, J.J., Piersma, N. & Van den Poel, D. (2004). Joint Optimization of Customer Segmentation and Marketing Policy to Maximize Long-Term Profitability. Expert Systems with Applications, 27 (2), 159-168.

[15] Larivière, B. & Van den Poel, D. (2004). Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. Expert Systems with Applications, 27 (2), 277-285.

[16] Macintosh, G. & Lockshin, L.S. (1997). Retail Relationships and Store Loyalty: A Multi-Level Perspective. International Journal of Research in Marketing, 14 (5), 487–497.

[17] MacKay, D.J. (1992). Bayesian Interpolation. Neural Computation, 4, 415-447.

[18] Nabney, I.T. (2001). Netlab Algorithm for Pattern Recognition, Springer.

[19] Nunnally, J.C. (1978). Psychometric Theory. New York: McGraw-Hill.

[20] Reichheld, F.F. (1996). The Loyalty Effect. Harvard Business School Press, Cambridge, MA.

[21] Reichheld, F.F. & Sasser, W.E. Jr (1990). Zero defections: quality comes to service. Harvard Business Review, 68 (5), 105-111.

[22] Reinartz, W.J. & Kumar, V. (2000). On The Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. Journal of Marketing, 64 (October), 17–35.

[23] Rosenberg, L.J. & Czepiel, J.A. (1984). A Marketing Approach to Customer Retention. Journal of Consumer Marketing, 1, 45–51.

[24] Snee, R.D. (1977). Validation of regression models: Methods and examples. Technometrics, 19 (4), 415-428.

[25] Srinivasan, S.S., Anderson, R. & Ponnavolu, K. (2002). Customer Loyalty in E-commerce: An Exploration of its Antecedents and Consequences. Journal of Retailing, 78, 41–50.

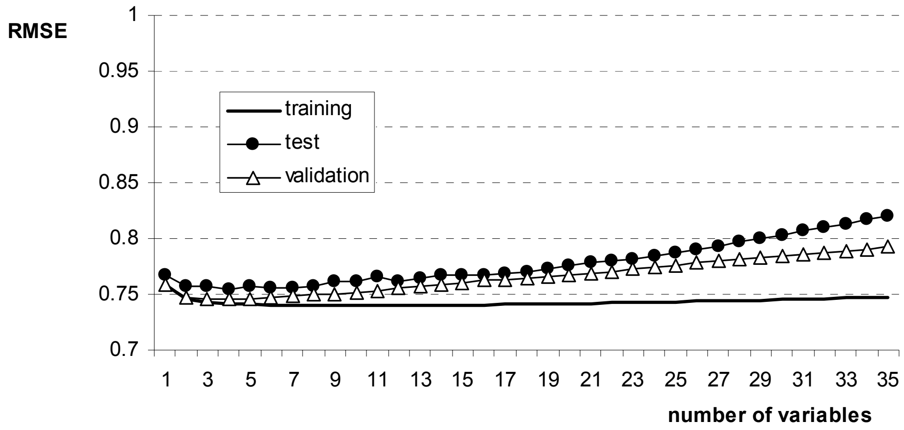
[26] Van den Poel, D. & Buckinx, W. (2005). Predicting Online Purchasing Behaviour. European Journal of Operational Research, 166(2), 557-575.

[27] Van den Poel, D. & Larivière, B. (2004). Customer Attrition Analysis for Financial Services Using Proportional Hazard Models. European Journal of Operational Research, 157 (1), 196-217.

[28] Verhoef, P.C., Spring, P.N., Hoekstra, J.C. & Leeflang, P. (2002). The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands. Decision Support Systems, 34, 471-481.

APPENDICES

Appendix 6.A: Evidence of overfitting when the number of variables is increased (RMSE).



DISCUSSION AND MAIN CONCLUSIONS

1. RECAPITULATION

The customization of marketing activities has known an extensive evolution during the last decades. This shift in marketing from a product-oriented to a customer-oriented policy yields benefits for both customers and companies: marketing costs can be restrained and clients are not interfered by inappropriate actions. Targeted marketing is enhanced by making use of individual customer information which is, typically, stored in company's internal transactional database. The management of marketing activities by using these data in combination with analytical models is called database marketing. Database marketing is among the fastest growing channels of marketing thanks to the evolution in information technology.

This doctoral dissertation researched methods to improve targeted marketing strategies by applying predictive modeling techniques. First, we investigated three topics related to direct marketing: the optimisation of direct mailing by assessing the individual profit functions to define customer ranking and the optimal mailing depth, the distribution possibilities of promotional coupons for retailers and manufacturers and, finally, the evaluation of site visitors' future purchase intentions based on their clicking behavior on a website. In addition, we studied how loyal customers can be tracked based on internal database information and based on additional information of a survey. In both studies the usefulness of knowing customer loyalty is examined. We investigated to what extent it is feasible to detect partial defectors among loyal clients and we queried current companies' reward programs by exploring an alternative criterion for compensating loyalty benefits.

The following paragraphs discuss our research topics and describe their most relevant findings. Conclusions about variable importance for each application and a summary concerning the analytical techniques that were used are made in separate sections.

2. DIRECT MARKETING

Direct marketing is applied to affect a measurable response by using advertising media. It has received a lot of attention in CRM literature, and emphasizes the importance of direct mailing and coupon targeting.

Therefore we present an improved direct mailing method which assesses and exploits the profit function, used to define expected customer values (Study 1). The accuracy by which individuals' contributions can be estimated has a direct impact on the ranking of the customers in the segmentation list and on the optimal mailing depth. So, we developed current theory in this matter by including customers' expected behavior in case they are not being targeted. That way, contributions are determined by accounting only for the net effect of a targeting action. Besides, we are the first study that investigated the substitution of all the elements of the profit function by the outcomes of separate predictive models. Most studies only account for purchase propensities. Our findings show that valid predictive models can be built for each of the aspects in the profit function. We indicate that the prediction of expected expenses has a better fit with real behavior than applying a mean expenditure, which definitely has a positive impact on the precision of the customer ranking. Moreover, the use of our advanced function is beneficial for companies as profits increase thanks to a reduction of their mailing costs and a modified ranking of their clients. This is demonstrated by incorporating this method into the direct mailing system of a European retailer. The optimal number of mailings was substantially reduced by sixty-five per cent. This was to be expected since contributions were based on the net effect of an action so customers who would make purchases anyway were consciously left out of the mailing list. As a result, their total profits increased by five per cent.

Several studies support the distribution of promotional coupons. Consumers tend to increase their purchase volume, coupons have a positive effect on repeat purchases and some authors support that promotions result in brand-switching behavior. However, the redemption rate of coupons is low and coupon strategies are told to be unprofitable. Besides, retailers and manufacturers got stuck in a competitive battle and are devising money-consuming actions to convince as much customers as possible to buy their products. In contrast, both type of products are said to attract different kinds of people. Consequently, we examined the use of predictive models to define the proneness of customers for both types of coupons (Study 2). Our findings point to the ability to classify customers with respect to their coupon redemption behavior during their next visit to the supermarket. So, retailers and manufacturers can identify their targets and improve their marketing strategies. Moreover, the entire customer base can be split into four segments: customers who are sensitive to store brand coupons, customers who will redeem coupons dispensed by manufacturers, customers who are interested in both types of coupons and clients who are not interested in coupons at all.

Furthermore, we studied how online retailers can enhance their targeting strategies towards their site visitors (Study 3). CRM opportunities seem elaborate in an e-commerce setting since much more data are available regarding customers' behavior on the website. Besides, companies are able to outline better client relations since they can communicate individually with clients and prospects. In contrast, online purchase behavior is very limited. For these reasons we investigated the features that control the visitors' decision whether or not to purchase. We show that feasible predictions can be made about which visitors will engage in online purchasing during their next visit to a website. This provides a powerful tool for marketing managers to fine-tune customized targeting strategies towards high and low scoring customers. Adapted messages like product recommendations and personalized advertising contents can be communicated. Many more variables of different types were taken into account than is done in previous studies, which results in a higher predictive ability and a better understanding of the relevant variable types. Detailed clickstream behavior appeared to be the most important customer information, compared to general clickstream behavior, demographics and past purchase data. This confirms the advantage of online retailers compared to traditional retailers.

3. LOYALTY

Several benefits are attributed to loyal customers. They increase their spending, they spread positive word-of-mouth about the company, they can be served at diminished costs, are less convinced by competitive pull, become price insensitive and have a positive impact on employees. As a consequence it seems appropriate to put effort in treating these clients. This requires the ability to distinguish loyal clients in a customer base so specific marketing actions towards this segment can be set up. Typically, however, information about customers' behavior at competitive stores is not available so no insight can be obtained on their loyalty level. Our work provided different tools to distinguish loyal customers and showed how both approaches can be put in practice to develop different suitable marketing activities.

A first study (study 4) focused on loyals by employing two behavioral attributes: the frequency of purchases and the time between purchases (standard deviation divided by the mean). Both these data elements can be found in the transactional database of a company and were confirmed in a second study to be relevant proxies for customer loyalty. In a non-contractual retail setting, customers can continuously change their purchasing behavior without informing a company about it. Besides, competition is severe and switching costs are low to nonexistent. Our empirical results show that our models can provide a viable method to track down future loyal defectors. Moreover, we introduce the aspect of partial defection so companies are signalled as early as possible about loyals' disadvantageous intentions. Avoiding this switching behavior of behaviorally-loyal clients is valuable for the retailer since the losses in terms of sales may be significant. Besides, partial defection can lead to total defection in the long run.

Study 5 presented the use of a predictive model to derive customer loyalty in two different store settings. Therefore, a survey was conducted to get complementary data of a random sample of customers. A three-item construct measured their behavioral loyalty to define the dependent variable in the model. The explanatory variables were computed based on data that can be derived from the internal customer base. Our findings point to the ability of marketing management to model customers' loyalty to an acceptable extent. So, an interrogation of a limited number of customers enables companies to create a loyalty score for every customer at any given moment. Besides, the results show that it is more complex to predict loyalty in a general merchandise setting than in a grocery environment. In both settings, the most valuable information to define loyalty is the variety of purchases and the responsiveness to direct mails. So, the more a customer is interested in a larger variety of product categories, the more loyal the customer. Customer information like spending and length-of-relationship are not found to be the best predictors, though they are generally used in loyalty-reward schemes. In the same study, we question the criteria that are widely used by companies to manage their reward systems. In most systems, customers are rewarded in accordance with their spending behavior. Our results show that if companies want to reward customers for more than repeat-purchase behavior, which is only one of the benefits attributed to loyal customers, they are recommended to take into account customer loyalty or the just-explained predicted loyalty. Other contributions, attributed to loyal customers are considered to be valuable for the growth, profitability and continuity of a company. The danger exists that loyal customers who are not heavy spenders and exhibit benefits that are currently not rewarded, will be motivated to switch their buying behavior. However, to choose an optimal reward criterion, a more in-depth cost-benefit analysis should be done. The final solution might be a segmented approach, in which several reward schemes are combined to reward different customer segments.

In our final study, we benchmarked several techniques with respect to their ability to predict customer loyalty. Consequently, we compared the performance of a multiple linear regression, the technique used in the previous study, with the performance of two state-of-the-art techniques: Random Forests and MacKay's Automatic Relevance Determination neural network. We show that, thanks to a valid variable selection procedure, our multiple linear regression outperforms the other models in terms of predictive accuracy. The result of this feature selection is discussed in a subsequent paragraph.

Our second loyalty segmentation method, proposed in studies 5 and 6, indicated that behavioral characteristics like frequency of purchases and the standard deviation of the time between customers' shopping incidences, are both relevant attributes to distinguish promising shoppers from others. Both features were used in study 4 to detect loyal customers based on the internal customer base, which confirms that a meaningful loyalty segmentation was employed. However, since these predictors were not the only significant explanatory variables in study 5 and 6, and additional customer information was selected in the final models, we recommend marketing managers to make use of our second approach in which transactional database information was enriched with survey information and predictive models were built to track loyal customers.

4. MODELING TECHNIQUES

In the course of this work, several different analytical techniques were used to model each of the targeting problems at hand: multivariate linear regressions, logistic regressions, C4.5 decision trees, Random Forests and Automatic Relevance Determination neural networks (ARD). We had not the intention to analyze the performance of an exhaustive number of modeling techniques. In most studies we rather preferred to use more than one technique in order to get a second or a third view on the predictive ability of the topic. In summary, we conclude that the predictive performance of straightforward techniques like multiple linear regressions and logits are not inferior to the ones of more state-to-the-art algorithms. In Study 1, Random Forests could only outperform the multivariate regression when assessing the expected expenses after customers are being treated. The other elements of the profit function are best predicted by using regressions. In Study 4, we could not make a distinction between the accuracies of logit, ARD or Random Forests. What's more, for the segmentation of loyal customers, our multiple linear regression did outperform ARD and Random Forests.

5. FEATURE SELECTION METHODS

In most studies we made use of a feature selection procedure in order to make a selection of variables or to avoid overfitting problems and increase the predictive performance on the validation set. We applied Forward selection, Backward selection, Relief-F, and Furnival and Wilson's (1974) global score algorithm. We did not observe a substantial difference between the selections made by Furnival and Wilson and a Forward and a Backward selection procedure (Study 3). However, the variable importances reported by Random Forests, ARD neural networks and Furnival and Wilson were considerably diverse (Study 6).

Most importantly, in generally all studies, the selection of inputs realized valuable results. The estimation of customers' expenses (Study 1) suffered overfitting problems since the accuracy on the validation set could be enhanced by including only a subset of predictors. Our research concerning the estimation of manufacturer coupon usage (Study 2) indicated that the inclusion of all variables did decrease the results so redundant variables could be excluded. Besides, we suggest marketing managers to predict customer loyalty by applying a multiple linear regression model in combination with a feature selection technique, since this procedure did outperform the accuracies of Random Forests and ARD neural networks (Study 6). Finally, in Study 6 the ARD model showed overfitting problems as well while this is not to be expected. An overview of the final models is shown in Appendix B.

6. MODEL PREDICTORS

In all of the models, we incorporated as much explanatory information as possible. First, the more information is included in a model, the more variance is explained and the higher the predictive power. Second, thanks to our substantial amount of data, we are able to evaluate different types of customer information with respect to their relevance for each of the targeted marketing topics. As a consequence, an analysis across all chapters gives insight into a) which data are relevant for each of the marketing problems, and b) which combination of customer data is optimal to get the highest predictive performance.

The available data were extensive but inconsistent across the studies. We grouped variables according to their variable type in order to formulate general conclusions. In total, eleven variable groups concerning past purchase history are considered (recency, frequency, monetary value, length-of-customer-relationship (lor), brand purchase behavior, category purchase behavior, promotional behavior, coupon usage, mode of payment, timing of shopping, response to mailing actions and return of goods). Besides, we distinguished demographical variables: distance to the store and all other demographical data, which are grouped in a category called 'other'. Finally we also considered general and detailed clickstream information in case online data were available. To analyze the importance of each of the variable types we use two different data sources. In studies 1, 3, 5 and 6 we calculated the univariate standardized parameter estimates. In studies 4 and 6 we dispose of the variables' importance reported by the Random Forests. Besides, for almost every topic, we applied a feature-selection procedure to overcome overfitting problems and to increase predictive performances. This resulted into final models, which are the combinations of variables that yield the best predictive power. Appendices A and B summarize these findings in two general tables.

Our studies confirm that the traditional RFM variables, widely used in many marketing studies, are relevant when predicting customers' future purchase propensity. However, information about responses to past targeting actions and the extent to which customers return goods are important information for this problem as well. Length-of-customer-relationship and demographical information have the least explanatory power. Besides, the relevance of the variables is identical for modeling purchase probabilities for targeted and not targeted customers.
However, there is a considerable difference with modeling customers' future expenditures. In these cases, it is particularly customer spending which is significant. Frequency and recency are less relevant predictors, just as the response to mailings and the return of goods. Again, demographical customer information is of minor importance. Length-of-customer-relationship is even totally irrelevant.

The prediction of online purchase behavior is explained by general and detailed clickstream behavior. This means that marketing managers can track relevant information about future purchase intentions based on the visiting behavior on their site. Our study indicated that detailed clickstream information is the most important data to collect. Even more important than past purchase behavior, in which recency and frequency are the best predictors. In comparison to most other applications, demographical information is important for online purchase predictions.

To define (partially) defectors, the variable importances of the Random Forests indicate that mainly RFM variables and customers' length-of-relationship are key predictors. Nevertheless, every variable type has some explanatory power.

There is only a small difference between grocery and general merchandise settings when defining the relevant variables for predicting customer loyalty. This suggests that our findings might be generalizable to other settings as well. The most relevant information is the response to past mailings and the purchase variety of customers. In both store environments, the traditional RFM variables and customers' promotional behavior are significant as well. In general, demographics have the least explanatory power. Remarkably, distance to the store is only valuable for grocery retailers and not for retailers selling durables. Length-of-customer-relationship is found at the bottom of our ranking.

The relevance of a variable does not mean that it is essential to the final model that yields the best predictive performance. Appendix B shows an overview of the final models, selected for each of our targeted marketing topics. The most notable conclusions are the appearance of almost all variable types in the prediction of purchase propensities and the selection of only spending related variables in the prediction of customer expenditures. Besides, defining future coupon usage needs to be done by combining past coupon usage information, demographics and promotional information (only for manufacturers). RFM variables, brand purchase behavior and data about purchase variety are not included. E-commerce retailers should include variables of each variable type to assess customers' purchase probabilities: general and detailed clickstream information, purchase history data and demographics. Finally, whereas almost every variable type showed to be important for the prediction of loyalty, retailers active in a grocery setting must only put purchase variety, responsiveness to past mailings and spending behavior into their final model. RFM variables, promotional behavior and the return of goods are to be included by companies who are selling durables.

To conclude, across the investigated targeted marketing activities, there is a substantial difference concerning which customer data are relevant for the problem at hand and which variables need to be included in the final model to optimize predictive performances. In summary, RFM variables are relevant for practically every application, though are not consistently selected in each of the final models. Purchase variety is especially important for FMCG retailers when predicting loyalty. Conversely, promotional behavior is important for retailers selling durables for the estimation of customer loyalty and for the distribution of manufacturer coupons. Further, past coupon usage is to be incorporated when defining future coupon redemption. The responsiveness to past mailing actions is relevant for the prediction of purchase probability and for the determination of customers' loyalty level. For both these applications, the return of goods is relevant as well, but these data are hardly ever restrained in the final models. Remarkably, length-of-customer-relationship is of major importance only for detecting partial defectors and increases the accuracy when predicting purchase propensity. The distance to the store explains the prediction of customer loyalty in grocery settings but is only selected for the prediction of future purchases. Demographics are to be incorporated for the prediction of future purchases, both in the traditional as in the online retail environment. Finally, clickstream data are only available and relevant for modelling online purchase intentions.

6. General conclusions

The management of targeted marketing strategies is still open to improvement. Direct marketing methods in a traditional and an online environment can be enhanced by customizing the distribution of product catalogs and coupons, and by detecting the future purchase intentions of visitors to a company's website. Besides, the detection of loyal customers offers a valuable expansion of the available individual customer data, used to tailor customized marketing actions. Different analytical techniques and feature selection methods have proven to be effective for the optimization of modeling results. The performance of frequently used and straightforward techniques is not always inferior to more complicated algorithms. Finally, the relevance of and the necessity for individual customer information differ across marketing activities and is dependent on the purpose of each action.

7. Further research

In each of our studies we already reported several issues for further research. In this section we present how general future research can build on our findings and how targeted marketing can be enhanced.

First, our research examined manners to enhance specific targeting problems in order to increase companies' marketing performances. However, in real-life, such actions are not undertaken mutually exclusively. Therefore, it would be of interest to investigate how different targeted marketing actions can be evaluated together in order to devise an integrated customized marketing approach. We expect that the effects of targeted activities will influence each other if they are performed simultaneously. The challenge then is to find out which actions - or which combinations of actions - are best employed for each individual. Suppose, for example, that both direct mailings, retailer coupons and online advertisements are all media to generate store traffic. What will be the optimal targeting approach for each of the customers individually to optimize the impact on total profits: employ all advertisement means, distribute only one of these media, or dispense none of them? In that respect, a customer might already be convinced to make a purchase after receiving a leaflet, which would imply that the distribution of a coupon on top of this leaflet would only mean lost revenue (value of the coupon) and more mailing costs, without having an additional effect on customer's purchase behavior. Or, for some customers, it might be sufficient to tease them with a personalized online advertising when visiting the company site to make them spend money at the store. We expect that the optimal combination of media tools will be different for each customer and therefore needs to be assessed individually. In summary, the examination of an integrated system, which balances different simultaneous targeted marketing actions to maximize company profits, is a subject for further research.

Besides, our work proposed an advanced profit function, which was extended by accounting for the net impact of direct mailings. We suggest to take other adjustments into consideration as well, and to examine the further elaboration of the profit function. Thereby, the proceeds of targeted marketing actions might be increased if companies would take into account to what extent a person will feel attracted by the content of a specific action, be it a coupon, a direct mail or other advertising media. Further, purchase propensities might be influenced by the degree to which the content of an advertisement differs from the ones that are sent during the previous mailing periods. Additionally, some customers make purchases at a company only at specific moments in time throughout the year. Therefore, it might be recommended to include such information when determining individual expected contributions. And, as mentioned in the previous paragraph, we expect that there will be an influence of simultaneously conducted advertisements as well.

Moreover, further research needs to examine to what extent the application of advanced profit functions are also useful for other targeted marketing actions. For example, the composition of segmentation lists for the distribution of coupons can be compensated by the extent to which specific products would be bought without the reception of a reduction. That way, customers having the intention to purchase a product anyway, can be - consciously - not given a coupon in order to save redundant promotional costs.

Furthermore, we expect that our proposed loyalty model has a lot of potential for targeted marketing research. First, we suggest to examine the usefulness of employing this additional information about customers' loyalty level in the assessment of other predictive models. Our expectation is that the incorporation of these individual data will enhance the explanation of model variance for other targeted marketing actions. Second, we constructed a loyalty score which reflects customer behavior with respect to the entire product range of the company. However, this score is like an overall loyalty indication and we have no data about which product categories a customer is loyal to. This could be of importance to define, for example, the content of direct marketing strategies. Further research needs to give insight into the ability of making use of similar data enrichment methods to define loyalty scores for product classes.

Besides, data enrichment seems a valuable tool for marketing applications. Therefore it is worthwhile to study whether also other customer behavior can be predicted by internal transactional company information. It might be interesting to gain insight into reasons of customers' disloyalty. By estimating customer loyalty, the interpretation of the parameter estimates gives an indication, to a very limited extent, why people are not loyal. For example, the negative sign of 'distance to the store' explains that certain customers are not loyal because of the location of the store. However, much more reasons will exist why customers purchase at competitive stores: level of prices, product quality, lay-out of the store, product assortment, friendliness and competence of store personnel, image, shopping pleasure, … . Nevertheless, these data are not at companies' disposal either. Information about customers' loyalty level completed with reasons why they are not loyal might enhance a customized marketing policy. As a consequence, further research should study to what extent these reasons for unloyalty can be estimated by sending out a questionnaire and building scores based on individual transactional data from the customers.

APPENDICES

Appendix A: Variable importance across all studies.

		Study 1 Assessment of profit function				**Study 2 Coupon use**		**Study 3 Online**	**Study 4 Defection**	**Study 5 True loyalty**		**Study 6 Loyalty**
Data type		Purchase with treatment uni[1]	Purchase without treatment uni	Expenses with treatment uni	Expenses without treatment uni	Manufacturer coupon -	Retailer coupon -	Online purchase uni	Partial defection RF	Grocery uni	General merchandise uni	Loyalty uni / RF
Purchase history	**Recency**	***	***	*	*	na	na	***	***	**	***	***/***
	Frequency	***	***	**	*	na	na	**	***	***	***	***/*
	Monetary	***	***	***	**	na	na		***	***	***	***/***
	Brand	na[2]	na	na	na	na	na	na	*	na	na	na
	Category	na	na	na	na	na	na	na	*	***	***	***/***
	Promotions	na	na	na	na	na	na	na	*	***	***	***/***
	Coupon usage	na	na	na	na	na	na	na	**	na	na	na
	MoP	na	na	na	na	na	na	na	**	na	na	na
	Timing	na	na	na	na	na	na	na	*	na	na	na
	Mailing response	***	***	*	*	na	na	na	na	***	***	***/**
	Retour	***	***	*	*	na	na	na	**	**	**	**/**
	Lor	*	*			na	na	na	***	*	*	*/
Demographics	**Distance**	*	*	*	*	na	na	na	na	**		**/*
	Other	*	*	*	*	na	na	***	**	na	na	na
Clickstream	**General**	na	na	na	na	na	na	***	na	na	na	na
	Detailed	na	na	na	na	na	na	***	na	na	na	na

1:
uni: Importances based on ranking and significance of univariate standardized parameter estimates
RF: Importances based on ranking and significance of variable importance from Random Forests

2:
na:data was not available or not applicable

*** variables are significant and at top of importance ranking
** variables are significant and in the middle of importance ranking
* variables are significant and at bottom of importance ranking

Appendix B: Variable selection in final models across all studies.

Data type		Study 1 Assessment of profit function: Purchase with treatment F&W[1]	Study 1: Purchase without treatment F&W	Study 1: Expenses with treatment F&W	Study 1: Expenses without treatment F&W	Study 2 Coupon usage: Manufacturer coupon Relief-F	Study 2: Retailer coupon Relief-F	Study 3 Online: Online purchase F&W	Study 4 Defection: Partial defection -	Study 5 True loyalty: Grocery F&W	Study 5: General merchandise F&W	Study 6 Loyalty: Loyalty F&W
Purchase history	**Recency**	***[2]	***					***	na		***	
	Frequency	***	***					*	na		**	
	Monetary	***	***	***	***				na	**	**	***
	Brand	na[3]	na	na	na			na	na	na	na	na
	Category	na	na	na	na			na	na	***		***
	Promotions	na	na	na	na	***		na	na		**	
	Coupon usage	na	na	na	na	***	***	na	na	na	na	na
	MoP	na	na	na	na			na	na	na	na	na
	Timing	na	na	na	na			na	na	na	na	na
	Mailing response	***	***			na	na	na	na	**	***	***
	Retour					na	na	na	na		*	
	Lor	***						na	na			
Demographics	**Distance**	***	**			na	na	na	na			
	Other	***	***			***	***	***	na	na	na	na
Clickstream	**General**	na	na	na	na	na	na	***	na	na	na	na
	Detailed	na	na	na	na	na	na	***	na	na	na	na

1: F&W: Furnival and Wilson selection procedure
Relief-F: Relief-F selection procedure

2: * p < .10
** p < .05
*** p < .01

3: na:data was not available or not applicable