ELSEVIER

Expert Systems with Applications 29 (2005) 472–484

Expert Systems with Applications

www.elsevier.com/locate/eswa

Predicting customer retention and profitability by using random forests and regression forests techniques

Bart Larivière*, Dirk Van den Poel

Department of Marketing, Ghent University, Hoveniersberg 24, 9000 Ghent, Belgium

Abstract

In an era of strong customer relationship management (CRM) emphasis, firms strive to build valuable relationships with their existing customer base. In this study, we attempt to better understand three important measures of customer outcome: next buy, partial-defection and customers' profitability evolution. By means of random forests techniques we investigate a broad set of explanatory variables, including past customer behavior, observed customer heterogeneity and some typical variables related to intermediaries. We analyze a real-life sample of 100,000 customers taken from the data warehouse of a large European financial services company. Two types of random forests techniques are employed to analyze the data: random forests are used for binary classification, whereas regression forests are applied for the models with linear dependent variables. Our research findings demonstrate that both random forests techniques provide better fit for the estimation and validation sample compared to ordinary linear regression and logistic regression models. Furthermore, we find evidence that the same set of variables have a different impact on buying versus defection versus profitability behavior. Our findings suggest that past customer behavior is more important to generate repeat purchasing and favorable profitability evolutions, while the intermediary's role has a greater impact on the customers' defection proneness. Finally, our results demonstrate the benefits of analyzing different customer outcome variables simultaneously, since an extended investigation of the next buy–partial-defection–customer profitability triad indicates that one cannot fully understand a particular outcome without understanding the other related behavioral outcome variables.

Keywords: Data mining; Customer relationship management; Customer retention and profitability; Random forests and regression forests

1. Introduction

Since the last decade, many companies perceive the retention of the customer as a central topic in their management and marketing decisions (Van den Poel & Larivière, 2004). The emphasis on retention is based on the implicit assumption that there exists a strong association between customer retention and profitability: long-term customers buy more and are less costly to serve (Ganesh, Arnold, & Reynolds, 2000; Hwang, Jung, & Suh, 2004), whereas replacing existing customer by 'new' ones is known to be a more expensive (Bhattacharya, 1998; Colgate & Danaher, 2000) and risky strategy, since it is likely to assume that switched customers are more vulnerable to continue their churning behavior in the near future (Lewis & Bingham, 1991; McNeal, 1999). Nevertheless, there is no clear consensus about the true relationship between customer retention and profitability. In their study, Reinartz and Kumar (2000) argue that the most loyal customers are not necessarily the most profitable ones. As such, it is plausible to assume that some of the most retention-prone customers represent lower profits for the company than some other prosperous customers that divide their money among different financial services providers.

In this study, we investigate both customer retention and profitability outcomes, and we explicitly test for differences with respect to the impact of the same set of explanatory variables on both outcomes.

Unlike previous retention studies that mainly focus on one particular type of retention, this study adopts a more extended approach in the conceptualization of the retention dependent variables; we investigate a repeat purchase as well as a defection outcome. The first retention variable 'next buy' represents whether a customer has bought another product given a particular subset of independent variables. The second retention variable is labeled 'active partial-defection' and expresses the customer's decision to

* Corresponding author. Tel.: +32 9 264 35 24; fax: +32 9 264 42 79.
E-mail address: bart.lariviere@ugent.be (B. Larivière).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.04.043

cancel a product that is characterized by a 'non-ending' status. Contrary to typical grocery products like milk, coffee or cookies, financial products are bought and owned for a specific period in time. As a consequence, you remain a customer until all the products are closed or expired. Regarding the ending status of financial products, there exist two notable types: (i) products that have a fixed duration term and as a consequence automatically end when the expiration date is reached, and (ii) products that do not have a fixed expiration date and hence receive a 'non-ending' label, since they only stop when a customer explicitly asks to cancel that product. With the '*active* partial-defection' retention variable, we emphasize the latter ending status scenario. The 'partial' refers to the fact that the closure of one particular product does not necessarily mean a 'total' defection of the customer, since that customer is allowed to have other products that are still open or not expired.

With respect to the 'customer profitability' dependent variables, we investigate the customer's evolution in profit. Contrary to the existent literature that mainly investigated profitability in a cross-sectional manner by spanning companies and industries, we investigate each customer's profitability longitudinally. As such we are able to analyze the direct relationship between a customer's set of explanatory variables and his generated profits in contrast to previous studies that were often constrained by linking aggregated customer information with, for example, the stock-price performance per firm or the turnover per outlet due to the unavailability of profitability measures at the customer level. In this study, we investigate two measures of customer profitability. The first measure is 'profit evolution' and represents the customers' evolution with respect to the profits generated during the observed window of observation. The second variable 'profit drop' is a deduced version of the former profitability measure. 'Profit drop' is a binary variable expressing whether the customer has become less profitability for the company by the end of observation. The variable is created as an extra tool to validate the accuracy of predicting customers' profitability evolutions and to compare its performance with the other binary retention dependent variables.

In sum, we investigate two major groups of customer outcome: customer retention and profitability. We analyze two measures of retention that both involve an 'active' transaction of the customer: the opening of a new product (next buy) and the decision to end a product that is still open (active partial-defection). Furthermore, we also investigate how customers evolve in terms of the profitability they represent for the company by means of a linear (profit evolution) and a binary (profit drop) dependent variable.

The rest of the paper is organized as follows. In Section 2, we elucidate the methodological underpinnings of the random forests and the regression forests techniques. In Section 3, we present the data set and the explanatory variables under investigation. The study results and its implications are reported in Section 4. In Section 5, we summarize and discuss the results of this study.

2. Methodology

In this study, we use random forests techniques to predict customers' profitability evolution and their next buy and partial-defection decisions. Two types of random forests are used depending on the conceptualization of the dependent variable: that is binary classification and linear prediction outcomes. In the next paragraphs we present the methodological underpinnings of the random forests techniques and the evaluation criteria we use to investigate their performance.

2.1. Random forests

With regard to binary classification tasks, decision trees (DT) have become very popular, thanks to their ease of use and interpretability (Duda, Hart, & Stork, 2001) as well as their ability to deal with covariates measured at different measurement levels (including nominal variables). Nevertheless, conventional decision trees techniques also have their disadvantages. For instance, Dudoit, Fridlyand, and Speed (2002) mention their lack of robustness and the suboptimal performance. Fortunately, many of these disadvantages have been dealt with by some researchers who optimized the DT technique. More specifically, the creation of an ensemble of trees followed by a vote for the most popular class, labeled forests (Breiman, 2001), is the result of such a DT optimization.

In this paper, we also use the more advanced DT technique. We select the random forests as proposed by Breiman (2001), which uses the strategy of a random selection of a subset of m predictors to grow each tree, where each tree is grown on a bootstrap sample of the training set. This number, m, is used to split the nodes and is much smaller than the total number of variables available for analysis.

Since its introduction, random forests have been enjoying increased popularity. The number of applications in fields with large datasets is growing: e.g. in bioinformatics (Deng et al., 2004). On the other hand, the number of applications in economics, and, more specifically in marketing related issues are rather scarce (Buckinx & Van den Poel, 2005). The available applications using random forests reveal that the predictive performance is among the best of available techniques (Luo et al., 2004). Furthermore, an interesting by-product of the technique are the produced importance measures for each variable (Ishwaran, Blackstone, Pothier, & Lauer, 2004) that indicate which variables have the strongest impact on the dependent variables of investigation. Another advantage of the technique concerns the consistent high and robust performance results (Breiman, 2001). Finally, the random forests as proposed

by Breiman have reasonable computing times (Buckinx & Van den Poel, 2005) and are easy to use; 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 (m) to be randomly selected from the available set of variables. In both cases, we follow Breiman's recommendation to pick a large number (5000 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. Since the number of explanatory variables equals to 30 (cf. Table 2) in this study, we fix the number of variables to six.

2.2. Regression forests

Breiman also extended the concept of random forests to regression cases. Random forests for regression are formed by growing trees depending on a random vector such that the tree predictor takes on numerical values as opposed to class labels (cf. Section 2.1). The random forests predictor is formed by taking the average over a number of the trees specified by the user.

2.3. Evaluation criteria

In this study, we investigate four different dependent variables: next buy, active partial-defection, profit drop and profit evolution. The first three measures involve a binary classification problem of a specific event; that is the event of buying a new product, the event of canceling a 'non-ending' status product and the event of becoming less profitable for the company.

In order to assess the predictive performance of the classification models based on the random forests technique, we use the area under the receiver operating characteristic curve (AUC) criterion. Furthermore, we benchmark the performance of the random forests against the AUC resulting from conventional logistic regression models in which we use the same set of customers, independent and dependent variables. The AUC measure is based on a range of comparisons between the predicted status of the event and the true status of the customer with respect to that event, by considering all possible cut off levels for the predicted values. More specifically, for all the cut off points, the sensitivity (the number of true positive versus the total number of events) and the specificity (the number of true negatives versus the total number of non-events) of the confusion matrix are considered and summarized by means of a two-dimensional graph, resulting in a ROC curve. The area under this curve is used to evaluate the predictive accuracy of the classification models (Hanley & McNeil, 1982). In order to compare the AUC's resulting from the random forests with these of the logistic regression models, we apply the non-parametric test proposed by DeLong, DeLong, and Clark-Pearson (1988) that investigates whether the areas under both ROC curves are significantly different.

With respect to the linear dependent variable, profit evolution, we cannot use the AUC evaluator, since both predicted and real values have more than two (i.e. binary) values. Profit evolution represents the change in the customer's profitability during the observed window of analysis, and consequently can have a wide range of both positive and negative values. In order to evaluate the predicted values, we calculate the mean absolute deviation (MAD)

$$\mathrm{MAD} = \frac{1}{n}\sum_{i=1}^{n}|P_i - R_i| \tag{1}$$

where n is the sample size, P_i the predicted profit evolution for customer i and R_i the real profit evolution for customer i. Similar to the goodness-of-fit evaluation of the random forests models, we also apply conventional linear regression models in order to benchmark its performance against the regression forests results with respect to the profit evolution target variable.

3. Empirical study

A major Belgian financial services company delivered the data for this study. Their data warehouse stores detailed information about customers' banking and insurance acquisitions; that is we know when, what, how much and at which point of sales the customer has bought a specific product. Furthermore, the company gathers demographic information about its customers and provides its customers with a monthly revenue indicator. Since our research setting implies a fourfold analysis of dependent variables, we decided to use the same group of customers, as well as the same set of potential explanatory variables in order to compare their relative and different impact on the customer retention and profitability target variables we emphasize. We decided to take two randomly selected samples of 50,000 customers each of which one is used for the estimation process and the second sample is used for validation. In the next paragraphs, we present the dependent and explanatory variables that are created to perform the customer retention and profitability models.

3.1. Conceptualization of the dependent variables

The following timeline provides some detailed information about the period of analysis in this study.

As it is clear from the timeline, we determine the dependent variables within the time period of 1 June 2003 through 1 February 2004 (=latest release date of the data warehouse). Two measures of retention are created in order to investigate the postulated research objectives. The first measure is 'next buy' and expresses whether the customer has bought a new product during the 8 months of follow-up (i.e. 1 June 2003 through 1 February 2004). The second

dependent variable 'active partial-defection' explores whether the customer has ended himself a product that was still open. Note that with respect to the latter dependent variable, we explicitly focus on 'active' defection, meaning that we do not consider an 'automatic' product defection as the event of investigation (cf. Section 1). Both retention variables are binary and receive the value of '1' when the event happened during the follow-up period ('0' in the other case).

With respect to the profitability measures we make use of the company's internal records. Each month, the investigated company computes an individual profitability score for its entire customer base. The monthly score is calculated as a weighted average of the total number of products owned multiplied by the corresponding balance amount (at the end of each month) and the net margin that the product represents for the company. Based on the scores throughout the follow-up period, we were able investigate the customers' evolution with respect to that profitability score. We created two dependent variables. The first profitability measure is 'profit evolution' and represents the shift (expressed in profitability points) in the customer's profitability, whereas 'profit drop' is a binary indicator expressing whether the customer showed a negative evolution with respect to his revenue profile, meaning that he became less profitable for the company by the end of the follow-up period. In Table 1, we provide some insights about the 100,000 customers under investigation in this study and their corresponding retention and profitability measures.

It is clear from Table 1 that some 13% of the customers bought a new product during the follow-up period, whereas fewer customers (6.8% of the customers) decided to cancel a product with a non-ending status. With respect to the binary profitability measure, we observe that approximately a quarter of the customers experienced a negative evolution in the profitability they represent for the firm. This latter finding is intriguing in the context of the second profitability measure that reflects the absolute shift in a customer's profit evolution expressed in profitability points. It is clear from the minimum and maximum values that there is a wide range of movements within the follow-up period with regard to the customers' profit evolutions. Furthermore, the mean and median values are situated around zero, indicating that the extra profits generated by some customers are fully absorbed by the lost revenues of some other profitability defectors. Given the fact that only one quarter experienced a decrease in profits, we can ascertain the need to gain insight into the drivers of the target variable 'profit drop', because on average one customer is likely to absorb the extra profits generated by three other customers.

3.2. Explanatory variables

In this study, we explore three major predictor categories that encompass potential explanatory variables. The three categories are: past customer behavior, observed customer heterogeneity and variables related to intermediaries. In the next paragraphs, we introduce each category by presenting its variables. Note that all explanatory variables are measured at the date of 31 May 2003 (cf. Fig. 1). Table 2 presents the explanatory variables that are investigated in this study.

3.2.1. Past customer behavior

There is ample evidence in the literature that behavioral exchange characteristics are strong predictors of future customer behavior (Baesens et al., 2004; Reinartz & Kumar, 2003) and profitability (Hsieh, 2004). In this study, we investigate the following past customer behavior variables: specific product ownership, self-banking activity, total number of products owned, monetary value and cross-buying.

3.2.1.1. Specific product ownership. Some researchers investigated the impact of specific product ownership on customer outcome. Their findings indicate that specific product ownership is likely to influence future customer behavior (e.g. Athanassopoulos, 2000; Larivière & Van den Poel, 2004). In this study, we test for the impact of seven

Table 1
Insight in the dependent variables for both estimation and validation sample

Dependent variables[a]		Estimation sample ($N=50{,}000$)		Validation sample ($N=50{,}000$)	
		Absolute	Relative (%)	Absolute	Relative (%)
Next buy	Yes	6642	13.3	6644	13.3
	No	43,358	86.7	43,356	86.7
Active partial-defection	Yes	3420	6.8	3386	6.8
	No	46,580	93.2	46,614	93.2
Profit drop	Yes	14,349	28.7	14,167	28.3
	No	35,651	71.3	35,833	71.7
Profit evolution	Min	−2938.28		−3500.86	
	Max	1179.53		2027.58	
	Mean	−1.06		−0.98	
	Median	0.01		0.01	

[a] All dependent variables are measured within the follow-up period (1 June 2003 through 1 February 2004).

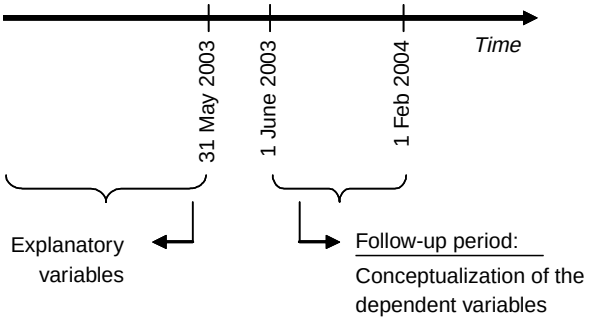

Fig. 1. Period of analysis.

different ownership variables. We introduce six dummy variables that categorize all types of banking and insurance products as well as one variable expressing whether the customer owns credit cards or not.

3.2.1.2. Self-banking by means of internet and phone. Nowadays, more and more financial services providers encourage their customers to perform their daily transactions by means of electronic banking services (such as self-banking with ATM, phone banking or internet banking) in order to minimize their operational working costs. Also, the company of investigation enables its customers to use internet or phone services for both banking and insurance transactions. For this study, we created a dummy variable expressing whether the customer is a self-banking user (by means of internet or phone).

3.2.1.3. Total number of products owned and monetary value. Previous research suggests that there exists a positive association between these two explanatory variables and customers' subsequent customer behavior. For instance, Huber, Lane, and Pofcher (1998) reveal that the more products a customer possesses with the bank, the more retention prone he is. Similarly, the more money a customer invests with a company the more likely he is to stay (Baesens, Viaene, Van den Poel, Vanthienen, & Dedene, 2002; Ganesan, 1994). With respect to the customer's profitability, it is plausible to assume that a higher quantity of products represents higher profits, since previous research found a positive relationship between customers'

Table 2
Explanatory variables used in this study

1. *Past customer behavior*	
Specific product ownership	
Possession of savings and investment products, type low risk (e.g. a savings account, bonds, etc.)	d_SI_low_risk[a]
Possession of savings and investment products, type high risk (e.g. exchange products like stocks, etc.)	d_SI_high_risk
Possession of a typical savings and investment product that is created as the steppingstone between the two other savings and investment groups	d_SI_stepst
Possession of risk products (e.g. fire insurance, car insurance, etc.)	d_risks
Possession of credit products (e.g. a mortgage, etc.)	d_credits
Possession of current account	d_curracc
Possession of (credit) cards	d_card
Self-banking by means of internet or phone	d_self_b
Total number of products owned	nbr_p
Monetary value	mon_val
Cross-buying	cross_b
2. *Customer demographics*	
Age	age
Lifecycle stage: that is, respectively, (1) youngsters, (2) families with young children, (3) midlife and (4) seniors	d_lifec_stage_1 d_lifec_stage_2 d_lifec_stage_3 d_lifec_stage_4
Gender (1 = male, 0 = female)	d_gender
Region (1 = Flanders, 0 = Walloon)	d_region
Geo-demographic data	
Social status of the place of residence	d_soc_status_1 d_soc_status_2 d_soc_status_3 d_soc_status_4 d_soc_status_5 d_soc_status_6 d_soc_status_7 d_soc_status_8
Median income per place of residence	med_income
3. *Intermediary variables*	
Selling tendency	ST
Number of customers served	nbr_cust
Sales assortment	sales_assort

[a] The prefix 'd_' refers to the fact that the corresponding variable is a dummy variable.

spending level and profitable lifetimes (Reinartz & Kumar, 2003). In this study, we also control for the customers' total product ownership and monetary value.

3.2.1.4. Cross-buying. Cross-buying refers to the degree to which customers purchase products from different product categories offered by the company. In this study, we explicitly decided to create a cross-buying variable, since the investigated company is characterized by a large group of mono-product customers. As such, it offers a viable opportunity to investigate the impact of a higher share-of-wallet on both retention and profitability dependent variables.

3.2.2. Customer demographics

It is clear from previous research that accounting for observed customer heterogeneity is warranted. In this study, we control for the customer's age, lifecycle stage, gender, geographical region, and some geo-demographic data.

3.2.2.1. Age and lifecycle stage. It is well known that customers' financial-need priorities and resource availability vary at different stages of his lifecycle, and as such influence the quantity and the sequence in which financial products and services are acquired (Kamakura, Ramaswami, & Srivastava, 1991): e.g. in general, younger customers (i.e. the 'bachelor' stage) have less money to invest than older individuals (i.e. the 'empty-nest' and the 'retirement' stage). As such older people that belong to a later stage in their lifecycle are assumed to have more money available. In this study, the lifecycle stage consists of five stages; as such we create four dummy variables in order to express to which stage the customer belongs. A higher number corresponds with a later stage in the lifecycle.

3.2.2.2. Gender. As in most studies that account for customer demographic data, we also control for the customer's gender. The variable 'gender' is operationalized as a dummy variable that receives the value of '1' when the customer is male, and a '0' when the customer is female.

3.2.2.3. Geographical region. The investigated company provides its financial products and services at the Belgian market. The variable region reflects a geographical cohort and is operationalized as a dummy variable. In general, the Belgian market can be divided into two large geographical areas: Flanders in the north and the Walloon part in the south. Besides the fact that each region has its own language (respectively, Dutch and French), the marketing department of the investigated company reveals that they also have a different way of doing business with a financial services company; that is Flemish people are known to be 'savers', whereas the Walloons make more use of personal loans to acquire the products they want. Also previous research has taken the geographical region into account. For example, Patterson and Smith (2003) investigated the propensity to remain with a service supplier for both Australia and Thailand and found a significant difference. In this study, we also account for this cohort information in order to test whether we observe some significant differences with respect to the profitability and retention proneness for Flemish versus Walloon customers.

3.2.2.4. Geo-demographic data. Besides customer demographic data gathered at the customer level, the company also buys some additional customer information that is gathered based on the place of residence (that is geo-demographic data). In this study, we analyze two different information items: the social status and the median income of the region of residence. The social status consists of nine groups. Therefore, we create eight dummy variables per customer in order to know to which categorical group a customer belongs. We wonder whether these variables provide some additional explanatory information with respect to the dependent variables we emphasize; and as a consequence—in terms of practical reasons for the company—are worth paying for.

3.2.3. Variables related to intermediaries

To date there is still a poor understanding of the impact of salespersons (or intermediaries) on customers' behavior (Guenzi & Pelloni, 2004). Nevertheless, it seems important to investigate the salesperson's role, since he acts as the crucial player who interacts with the company's customers. In this study, we investigate three variables related to these intermediary agents: the selling tendency of the salesperson, the number of customers served by a salesperson and the sales assortment.

3.2.3.1. Selling tendency of the salesperson. In real life, it is likely to assume that not every intermediary is equally skilled in selling financial products and services to the company's customers. With the variable 'selling tendency', we aim to explore the impact of a salesperson's selling capabilities on both the customers' profitability and retention proneness. The variable 'selling tendency' represents the number of products sold in relation to the number of customers served by a specific intermediary. The variable is created by using the information from 1 year preceding the date of 31 May 2003 (cf. Fig. 1); the higher the value for the variable the more products the intermediary had sold to its own customer base.

3.2.3.2. Number of customers served by the salesperson. Although many researchers have suggested that the performance of the salesperson during sales encounters is critical, many of the underlying mechanisms that govern the interaction between salespersons and customers are still unclear (Van Dolen, Lemmink, de Ruyter, & de Jong, 2002). In the financial services setting, it is plausible to believe that some customers experience less personal attention when a salesperson is serving a large customer base, and as a

consequence is unable to know each customer personally. In this study, we account for this information and investigate its impact on both customer's behavior and profitability.

3.2.3.3. Sales assortment. The 'sales assortment' represents the product variety offered by the salesperson. In their study, Hoch, Bradlow, and Wansink (1999) state that the variety in offerings is viewed as the entree fee for maintaining future customer loyalty. With respect to the investigated financial services company, not every intermediary is selling the whole range of financial products to its customers; that is some typical 'banking' intermediaries solely supply banking products, whereas some others only sell a limited variety of insurance products to their customers. As such, it is possible that some customers are unable to acquire all financial products and services they need with their current salesperson. In this study, we explore its impact on both retention and future customer profitability.

4. Findings

The next paragraphs present the findings of the study. First, we report the prediction accuracies of the various models. Next, we present the relative importance of each explanatory variable with respect to the four dependent variables under investigation. Finally, we further examine the signs of the 10 most important covariates for each target variable by means of some descriptive statistics.

4.1. Performance evaluation

The evaluation criteria applied to investigate the predictability of the four dependent variables are presented in Table 3.

It is clear from the table that random forests provide better prediction accuracies compared to logistic regression models. For all three binary classification targets, we observe a significant (DeLong et al., 1988) and better performance in favor of the random forests (cf. all p-values range between <0.0001 and 0.025). Even for the next-buy classification, we find a significant difference although the increase in prediction accuracy is rather low; that is an AUC improvement of 0.006 (0.005) for the validation (estimation) sample. With respect to the predictive performance of the profit drop target variable, we observe a significant difference in AUC of 0.019 and 0.016 for, respectively, the validation and estimation sample. In this study, the most important and outperforming prediction accuracy of random forests can be found in the active partial-defection analysis, where we observe an AUC improvement of 0.106 (0.094) for the validation (estimation) sample when benchmarking its performance against a logistic regression model.

In sum, the classification findings of this study indicate the viable opportunity for both academics and practitioner to consider other than the conventional prediction techniques (such as logistic regression) when investigating a binary-classification problem. Especially, when the obtained goodness-of-fit indices based on conventional prediction models perform rather low—indicating that there is more room for improvement—it is appealing to investigate whether other prediction techniques (such as random forests) perform better, since each major improvement in predictive accuracy is likely to represent major shifts in terms of the effectiveness and the return on investment of marketing actions—that are based on prediction models.

In order to evaluate the performance of the linear dependent variable, we use the mean absolute deviation (MAD) criterion (cf. Section 2.3). The MAD for the regression forests model amounts to 5 (more specifically, 5.099 for the test sample and 4.940 for the estimation

Table 3
Performance results

Dependent variable	Technique	AUC	
		Train	Test
Next buy	Random forests	0.752[a]	0.751[b]
	Logistic regression	0.747[a]	0.745[b]
Active partial-defection	Random forests	0.734[c]	0.742[d]
	Logistic regression	0.640[c]	0.636[d]
Profit drop	Random forests	0.713[e]	0.714[f]
	Logistic regression	0.697[e]	0.695[f]
Dependent variable	Technique	MAD	
		Train	Test
Profit evolution	Regression forests	4.940	5.099
	Linear regression	5.346	5.445

[a] $Chi^2=5.057$; df = 1; $p=0.025$.
[b] $Chi^2=7.157$; df = 1; $p=0.007$.
[c] $Chi^2=305.470$; df = 1; $p=<0.0001$.
[d] $Chi^2=383.140$; df = 1; $p=<0.0001$.
[e] $Chi^2=114.528$; df = 1; $p=<0.0001$.
[f] $Chi^2=145.970$; df = 1; $p=<0.0001$.

sample), meaning that on average we obtain a prediction 'error' of five profitability points per customer. In contrast, when evaluating the performance of the ordinary linear regression model, we observe an average MAD of 5.4 (that is 5.454 for the test sample and 5.346 for the estimation sample), which corresponds with a decline of approximately 7% in terms of prediction accuracy. As such, we find evidence in this study that regression forests outperform traditional linear regression techniques.

4.2. Relative importance indices for the explanatory variables

As stated in Section 2, a welcome feature of the random forests techniques is the importance measures for the explanatory variables. In Table 4, we report these importance indices with regard to each dependent variable of the study. The first three subsections in the table, respectively, present the importance measures with respect to the event of a next buy, active partial-defection and profit drop, whereas the latter part of the table reports the importance measures for the profit evolution target variable that is analyzed by means of regression forests. Per subsection, we ranked the variables in terms of its corresponding importance level.

A number of interesting findings emerge from the table. In the next paragraphs, we elaborate on the top-10 variables for each dependent variable in terms of their ranking.

4.2.1. Next buy

With respect to the binary classification variable that expresses the customer's likelihood to buy another product, we observe that, especially, past customer behavior variables drive subsequent repeat-purchase behavior. It is clear that specific product ownership has a major influence; more specifically: the possession of savings and investment (SI) products characterized by high risks (e.g. stock-market products), the possession of a current account and bank cards, as well as the possession of risk products. Furthermore, we observe the importance of customer's monetary values, the total product ownership and the fact whether a customer owns products from different categories (that is cross-buying). In terms of demographic variables, two variables belong to the top-10 list: the customer's age and the individuals in their senior lifecycle stage (that is d_lifec_stage_4). With respect to the salesperson's role, the selling tendency of the intermediary seems to play a significant and major influence. In sum, with respect to the customer's next buy decision, we can conclude that the stage in the lifecycle as well as the decision to be an active (cf. possession of cards, current account, etc.) and loyal customer (measured by means of cross-buying, monetary value and total product ownership) supported by the intermediary's selling skills are responsible for future purchase behavior.

4.2.2. Active partial-defection

With regard to the active partial-defection prediction, it is striking that the three explanatory variables related to the salespeople show the highest importance measures. Especially, the selling tendency variable that is intuitively more affiliated with the previous dependent variable (next buy) appears to be the number one predictor in terms of importance, whereas the variable was attributed a 10th place for the next buy dependent variable. As such, the selling capabilities of the agent do not only influence the customers purchase decisions, they also strongly drive customer's vulnerability to cancel non-ending status products. Furthermore, it is clear that the monetary value, the total product ownership, the customer's cross-buying behavior and age belong to the top-10 list of the most important variables. Moreover, Table 4 reveals that these four variables are the only ones that appear in the top 10 for all the dependent variables under investigation in this study; suggesting its importance for a variety of CRM applications. Table 4 also reveals the importance of risks products such as fire and car insurances as well as SI products characterized by high capital and revenue risks. Finally, the cohort variable indicating whether a customer belongs to the Walloon or Flemish part of Belgium seems to influence active partial-defection. As such, we find evidence that besides the difference in language, there also exist differences with respect to the cancellation of financial services products.

4.2.3. Profit drop

For the profit drop dependent variable, we observe that again the past customer behavior variables play an important role as well as some customer demographic variables. On the other hand, none of the variables related to the intermediaries show up in the top-10 list. With respect to the past behavior variables, we observe a stronger influence of being an 'active' customer, since the self-banking variable has joined the top-10 importance list. As a consequence, our findings give evidence that the interactiveness of customers strongly influence the company's future profits.

4.2.4. Profit evolution

A first thing that strikes the attention when investigating the importance indices for the profit evolution variable is that more than half of the explanatory variables reveal no significant impact on the prediction of customers' profit evolutions. Hence, our results suggest that it is more difficult to understand the customer's absolute changes in profitability, compared to the deduced format that only indicates a customer's profit direction by means of a binary classification. With respect to the most important variables, we find similar findings compared to the binary profit drop variable; that is past customer behavior variables are most important followed by customer demographic variables.

Table 4
Importance of variables

Random forests									Regression forests		
Dependent variable = Next buy			Dependent variable = Active partial-defection			Dependent variable = Profit drop			Dependent variable = Profit evolution		
No.	Importance measure[a]	Variable name	No.	Importance measure	Variable name	No.	Importance measure	Variable name	No.	Importance measure	Variable name
1	149.536	d_SI_high_risk	1	273.260	ST	1	174.616	mon_val	1	1.490	d_credits
2	147.642	d_curracc	2	240.702	nbr_cust	2	148.870	d_curracc	2	1.004	age
3	141.566	age	3	191.625	sales_assort	3	125.883	d_card	3	0.623	mon_val
4	128.041	mon_val	4	167.320	age	4	114.565	nbr_p	4	0.575	nbr_p
5	124.833	nbr_p	5	144.917	mon_val	5	95.953	age	5	0.273	cross_b
6	118.025	d_card	6	137.873	nbr_p	6	95.284	cross_b	6	0.232	nbr_cust
7	106.901	d_risks	7	132.024	d_region	7	94.376	d_credits	7	0.171	d_curracc
8	92.611	d_lifec_stage_4	8	126.118	d_risks	8	89.111	d_lifec_stage_4	8	0.144	d_soc_status_6
9	91.995	cross_b	9	123.324	cross_b	9	83.848	d_self_b	9	0.137	d_risks
10	75.584	ST	10	94.772	d_SI_high_risk	10	81.666	d_lifec_stage_2	10	0.091	d_lifec_stage_2
11	74.022	d_SI_low_risk	11	93.972	d_lifec_stage_2	11	79.941	sales_assort	11	0.039	d_card
12	73.556	d_credits	12	88.316	d_credits	12	72.971	d_region	12	0.009	d_self_b
13	70.177	nbr_cust	13	86.139	d_card	13	72.587	d_risks	13	0	ST
14	67.816	med_income	14	85.948	med_income	14	69.856	d_SI_high_risk	14	0	sales_assort
15	62.856	d_self_b	15	85.570	d_curracc	15	69.758	nbr_cust	15	0	med_income
16	61.791	d_lifec_stage_2	16	72.341	d_lifec_stage_4	16	64.157	d_SI_low_risk	16	0	d_SI_low_risk
17	57.312	d_lifec_stage_3	17	66.582	d_SI_low_risk	17	59.016	ST	17	0	d_SI_high_risk
18	54.640	sales_assort	18	63.783	d_lifec_stage_3	18	54.052	med_income	18	0	d_SI_stepst
19	54.232	d_SI_stepst	19	59.480	d_soc_status_6	19	47.041	d_lifec_stage_3	19	0	d_gender
20	36.329	d_region	20	54.517	d_self_b	20	39.008	d_SI_stepst	20	0	d_soc_status_1
21	28.606	d_soc_status_1	21	52.649	d_soc_status_8	21	38.321	d_soc_status_8	21	0	d_soc_status_2
22	25.130	d_soc_status_7	22	39.375	d_soc_status_2	22	25.617	d_soc_status_7	22	0	d_soc_status_3
23	21.529	d_soc_status_2	23	37.604	d_SI_stepst	23	22.509	d_soc_status_3	23	0	d_soc_status_4
24	19.544	d_soc_status_6	24	35.199	d_soc_status_1	24	21.363	d_soc_status_6	24	0	d_soc_status_5
25	14.317	d_soc_status_3	25	32.463	d_soc_status_7	25	17.138	d_soc_status_2	25	0	d_soc_status_7
26	8.789	d_soc_status_8	26	30.958	d_soc_status_3	26	13.215	d_soc_status_1	26	0	d_soc_status_8
27	4.66	d_lifec_stage_1	27	12.634	d_soc_status_5	27	11.699	d_soc_status_4	27	0	d_lifec_stage_1
28	4.193	d_gender	28	9.913	d_gender	28	10.836	d_lifec_stage_1	28	0	d_lifec_stage_3
29	0	d_soc_status_4	29	9.059	d_lifec_stage_1	29	9.72	d_gender	29	0	d_lifec_stage_4
30	0	d_soc_status_5	30	6.666	d_soc_status_4	30	4.288	d_soc_status_5	30	0	d_region

[a] An importance measure of '0' represents no significant impact of the corresponding explanatory variable on the target variable of investigation.

4.3. Investigation of the direction of impact on the dependent variables

While Section 4.2 provides a clear understanding of the explanatory variables that have a strong impact on the four dependent variables of this study, the directions of these impacts are still unknown. For example, the variable 'd_region' plays an important role in the prediction of active partial-defection, but nevertheless we have no indication whether Flemish customers, in contrast to their Walloon counterparts, are less or more likely to defect. Hence, we decided to perform some additional descriptive analyses to gain insight into the direction of the most important explanatory variables. Analogous to Section 4.2, we focus on the top-10 most important predictors and we only investigate the binary target variables. Table 4 summarizes the descriptive statistics. In fact, we analyze two strata (e.g. next buyers or not) and we wonder whether we observe a statistically significant difference with respect to the 10 most important variables. For the binary explanatory variables, we apply simple chi-square statistics, whereas *T*-tests are performed for the other covariates.

While most explanatory variables have the excepted sign, some other findings deserve some further explanation. In the next paragraphs we briefly summarize the most intriguing findings of Table 5.

Section 4.2 revealed the importance of the past behavior variables, such as total product ownership, cross-buying, monetary value and specific product ownership; in this extended analysis we find that all these explanatory variables have a positive association with *all* the events under investigation: that is next buy, active partial-defection and profit drop. The latter finding implies that, for instance, customers with higher monetary value or individuals that possess more products from different categories (cross-buying) are not only more likely to buy new products in the future (next buy), they are also more vulnerable to cancel other products with a non-ending status (active partial-defection), which probably results in a negative profitability evolution (profit drop). In sum, our findings suggest the

Table 5
Descriptive statistics for the most important explanatory variables

Explanatory variable		Strata		
		Next buyers	No next buyers	*p*-value
1 d_SI_high_risk	Percentage of customers that possess SI products characterized by high risks	26.78%	12.80%	<0.0001
2 d_curracc	Percentage of customers that own a current account	55.69%	37.74%	<0.0001
3 age	Mean age per strata	50.70	44.32	<0.0001
4 mon_val	Mean monetary value per strata	14,849	7380.9	<.0001
5 nbr_p	Mean number of products owned by the customer	11.08	4.58	<0.0001
6 d_card	Percentage of customers that possess cards	40.29%	23.88%	<0.0001
7 d_risks	Percentage of customers that own risk products	24.82%	8.28%	<0.0001
8 d_lifec_stage_4	Percentage of customers that belong to lifecycle stage 4	34.06%	27.70%	<0.0001
9 cross_b	Mean number of cross-buyings per strata	2.99	1.87	<0.0001
10 ST	Mean selling tendency of the customer's intermediary	34.70	31.24	<0.0001

Explanatory variable		Strata		
		Active partial-defectors	No active partial-defectors	*p*-value
1 ST	Mean selling tendency of the customer's intermediary	30.99	31.76	0.025
2 nbr_cust	Mean number of customers served by the salesperson	1367	1540.5	<0.0001
3 sales_assort	Mean number of different product categories sold by the intermediary	11.837	11.987	<0.0001
4 age	Mean age per strata	47.18	44.99	<0.0001
5 mon_val	Mean monetary value per strata	12,198	8104	<0.0001
6 Nbr_p	Mean number of products owned by the customer	12.70	4.91	<0.0001
7 d_region	Percentage of customers belonging to the Flemish part of Belgium	73.60%	70.54%	<0.0001
8 d_risks	Percentage of customers that own risk products	17.50%	9.96%	<0.0001
9 cross_b	Mean number of cross-buyings per strata	2.49	1.99	<0.0001
10 d_SI_high_risk	Percentage of customers that possess SI products characterized by high risks	24.61%	13.93%	<0.0001

Explanatory variable		Strata		
		Profit droppers	No profit droppers	*p*-value
1 mon_val	Mean monetary value per strata	9964.8	7731.1	<0.0001
2 d_curracc	Percentage of customers that own a current account	63.48%	30.81%	<0.0001
3 d_card	Percentage of customers that possess cards	43.94%	18.93%	<0.0001
4 nbr_p	Mean number of products owned by the customer	5.79	5.31	0.002
5 age	Mean age per strata	46.86	44.42	<0.0001
6 cross_b	Mean number of cross-buyings per strata	2.37	1.89	<0.0001
7 d_credits	Percentage of customers that possess credit products	24.40%	12.56%	<0.0001

(*continued on next page*)

Table 5 (*continued*)

Explanatory variable		Strata		
		Profit droppers	No profit droppers	*p*-value
8 d_lifec_stage_4	Percentage of customers that belong to lifecycle stage 4	27.21%	29.08%	<0.0001
9 d_self_b	Percentage of customers doing self-banking	16.74%	7.42%	<0.0001
10 d_lifec_stage_2	Percentage of customers that belong to lifecycle stage 2	31.96%	35.39%	<0.0001

existence of a typical group of active customers that are constantly buying and defecting on financial products.

Furthermore, with respect to the variables related to the salesperson in the active partial-defection case, Table 5 reveals rather small (but significant) differences when comparing defectors versus non-defectors. Given the fact that these variables nevertheless represent the top three of most important variables, we can certainly ascertain the need to consider the intermediaries' role when trying to understand typical customer behavior outcomes; since even small improvements in, for example, the intermediary's selling capabilities or the sales assortment are likely to result in favorable customer behaviors. With regard to the 'number of customers served' variable, we find the opposite effect of what was hypothesized; that is customers belonging to larger agencies show lower active partial-defection rates. A possible explanation might be found in the fact that serving fewer customers is just the result (instead of the 'cause') of customer defections in the past. Furthermore, it is also likely to assume that intermediaries who serve fewer customers, experience a heavier competition in their immediate vicinity, such that their customers have more alternatives to switch. Another explanation might be that customers perceive large agencies as more reliable, and as a consequence prefer them above smaller agencies. Further research on this issue is warranted.

Finally, when we consider the customer lifecycle stage, we observe that seniors (d_lifec_stage_4) are more likely to repurchase, but less vulnerable to decrease their profitability. Also, families with young children show evidence of positive profitability evolutions. As a consequence, the other categories, such as the youngsters and the midlife category are mainly responsible for the negative profit evolutions. Hence, it is crucial for financial services companies to gain a better understanding of these typical lifecycle stages such that the appropriate and proactive actions can be taken to guarantee the company's future profits.

5. Discussion

This study investigates two typical and major outcomes of customer relationship management (CRM): customer retention and profitability. For the first outcome, we analyze two different measures: the opening of a new product (next buy) and the decision to cancel a product with a non-ending status (active partial-defection). With respect to the latter outcome, we investigate how customers evolve in terms of the profitability they represent for the company by means of a binary (profit drop) and a linear (profit evolution) dependent variable. More specifically, the first three measures involve a binary classification problem and are analyzed by using random forests; for the latter target variable (profit evolution), we applied regression forests.

Our research findings support previous studies that favor the use of random forests techniques. In this study, we observe significant improvements in terms of prediction accuracy when benchmarking the random and regression forests against the conventional logistic and linear regression models.

Another interesting feature of the random forests technique concerns the produced importance measures which indicate the variables that have the greatest impact on the dependent variable of investigation. In this study, we find evidence that past customer behavior variables play an important role in predicting future customer behavior and profitability. Another important finding of the study is the relative importance of the variables related to intermediaries with respect to the active partial-defection classification. It is clear that good selling agents not only generate more repeat purchases, they also indirectly prevent customers from (partial) defection. The same logics apply for the sales assortment of the salesperson. For the company of investigation, it offers a viable opportunity to encourage its salesforce to supply the whole range of financial products and services, since a limited sales assortment is likely to stimulate customer-switching behavior. With respect to the customer demographic variables, our findings reveal the importance of the customer's age and the stage of his lifecycle. On the other hand, the customer's gender and the geo-graphical data gathered at the place of residence level are less powerful in terms of predicting customer retention and profitability, although they report significant associations with the binary dependent variables.

Furthermore, we comparing the three binary classification outcomes and its most important predictors, it is striking that four of the top-10 variables are the same: total product ownership, monetary value, cross-buying and the customer's age. Moreover, when exploring their impact of the dependent variable by means of descriptive statistics, we observe the same positive impact on next buy, active partial-defection and profit drop. As such, we find evidence that the same set of variables is likely to generate both next-buy and defection behavior in terms of profits and products. These intriguing findings suggest the existence of a highly active customer segment, that is buying new products while it is switching on other financial products and invite us to perform some extra analyses that relate the next buy with the active partial-defection and customer profitability variables.

In Appendix A, we present the statistics for the next buy versus active partial-defection versus customer profitability triad. The statistics indicate that more than 25% of the active partial-defectors also bought a new product within the same period of observation, compared to a 12.36 buying percentage for the customers who did not cancel a non-ending product. As such, we find support for our theory that the company contains a typical segment of active customers that constantly replace old products by newer ones. Another striking finding concerns the link between next buy and customers' profit evolutions. It is clear from Appendix A that more than 35% of the customers who bought a new product also experienced a profit drop, whereas their counterparts who did not repurchase report lower percentages for the profit drop variable (that is 27.44%). Fortunately, in terms of absolute profitability shifts (cf. profit evolution), we do not observe a statistically significant difference whether customers purchased a next product. With respect to the relationship between active partial-defection and customers' profit evolution we observe the dramatic impact of customers' decision to cancel a non-ending product on their profitability evolution. Appendix A reveals that almost 70% of the active partial-defectors experienced a profit drop, while approximately one quarter (25.56%) of the people who did not defect on products showed a negative evolution with regard to the revenues they represent for the company. Similar conclusions can be derived for the profit evolution variable. Summarized, the latter findings are in line with previous research studies that underscore the impact of customer retention on a company's profitability: 'It is important to retain existing customers'. Finally, when linking the two profit evolution variables with each other, we confirm our descriptive findings resulting from Table 1 (cf. Section 3.1): on average the extent to which customers experience profit drops (in terms of absolute profitability points) is more intense than the extent to which other customers are able to grow in profits (that is −8.72 versus +1.60 profitability points, respectively). In sum, just as the well-known claim that 'it is important to retain existing customers', our research findings extend the same analogy with regard to customers' profitability: 'It is more profitable to retain the most profitable customers of the company'.

Acknowledgements

The authors would like to thank the anonymous company that supplied the data to perform this research study. Moreover, we are grateful to Leo Breiman for the public availability of the random forests and regression forests software.

Appendix A. Investigation of the next buy–partial-defection–customer profitability triad

Frequency table of next buy × active partial-defection

	Frequency			
	Row%	Active partial-defection		
	Column%	No	Yes	Total
Next buy	No	81,671	5043	86,714
		94.18%	5.82%	
		87.64%	74.10%	
	Yes	11,523	1763	13,286
		86.73%	13.27%	
		12.36%	25.90%	
	Total	93,194	6806	100,000

p-value = <0.0001.

Frequency table of next buy × profit drop

	Frequency			
	Row%	Profit drop		
	Column%	No	Yes	Total
Next buy	No	62,923	23,791	86,714
		72.56%	27.44%	
		88.02%	83.43%	
	Yes	8561	4725	13,286
		64.44%	35.56%	
		11.98%	16.57%	
	Total	71,484	28,516	100,000

p-value = <0.0001.

Frequency table of active partial-defection × profit drop

	Frequency			
	Row%	Profit drop		
	Column%	No	Yes	Total
Active partial-defection	No	69,372	23,822	93,194
		74.44%	25.56%	
		97.05%	83.54%	
	Yes	2112	4694	6806
		31.03%	68.97%	
		2.95%	16.46%	
	Total	71,484	28,516	100,000

p-value = <0.0001.

T-tests for the profit evolution variables

Strata	Mean profit evolution
Next buyers	−1.83 ns
No next buyers	−1.31 ns
Active partial-defectors	−4.45[a]
No active partial-defectors	−1.11[a]
Profit droppers	−8.72[a]
No profit droppers	1.60[a]

ns, not statistically significant.
[a] Statistically significant at <0.0001.

References

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

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.

Bhattacharya, C. B. (1998). When customers are members: Customer retention in paid membership contexts. *Journal of the Academy of Marketing Science*, *26*(1), 31–44.

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

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.

Colgate, M. R., & Danaher, P. J. (2000). Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. *Journal of the Academy of Marketing Science*, *28*(3), 375–387.

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*(3), 837–845.

Deng, Y. P., Chen, H. S., Tao, L., Sha, Q. Y., Chen, J., Tsai, C. J., et al. (2004). Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC Bioinformatics*, *5*(81), 1–12.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: Wiley.

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.

Ganesan, S. (1994). Determinants of long-term orientation in buyer–seller relationships. *Journal of Marketing*, *58*(2), 1–19.

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

Guenzi, P., & Pelloni, O. (2004). The impact of interpersonal relationships on customer satisfaction and loyalty to the service provider. *International Journal of Industry Management*, *15*(3/4), 365–384.

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.

Hoch, S. J., Bradlow, E. T., & Wansink, B. (1999). The variety of an assortment. *Marketing Science*, *18*(4), 527–546.

Hsieh, N.-C. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, *27*(4), 623–633.

Huber, C. P., Lane, K. R., & Pofcher, S. (1998). Format renewal in banks—it's not that easy. *McKinsey Quarterly*, *1998*(2), 148–156.

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.

Ishwaran, H., Blackstone, E. H., Pothier, C. E., & Lauer, M. S. (2004). Relative risk forests for exercise heart rate recovery as a predictor of mortality. *Journal of the Amercian Statistical Association*, *99*(467), 591–600.

Kamakura, W. A., Ramaswami, S. N., & Srivastava, R. K. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Reseach in Marketing*, *8*(4), 329–349.

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.

Lewis, B. R., & Bingham, G. H. (1991). The youth market for financial services. *International Journal of Bank Marketing*, *9*(2), 3–11.

Luo, T., Kramer, K., Goldgof, D. B., Hall, L. O., Samson, S., Remsen, A., et al. (2004). Recognizing plankton images from the shadow image particle profiling evaluation recorder. *IEEE Transactions on Systems Man and Cybernetics Part B—Cybernetics*, *34*(4), 1753–1762.

McNeal, J. U. (1999). *The kids market: Myths and realities*. Ithaca, New York: Paramount Market Publishing.

Patterson, P. G., & Smith, T. (2003). A cross-cultural study of switching barriers and propensity to stay with service providers. *Journal of Retailing*, *79*(2), 107–120.

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. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, *67*(1), 77–99.

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 Dolen, W., Lemmink, J., de Ruyter, K., & de Jong, A. (2002). Customer-sales employee encounters: A dyadic perspective. *Journal of Retailing*, *78*(4), 265–279.