UNIVERSITEIT GENT

FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Separating Financial From Commercial Customer Churn: A Modeling Step Towards Resolving The Conflict Between The Sales And Credit Department

Jonathan Burez [1]

Dirk Van den Poel [2]

August 2007

2007/476

[1] Jonathan Burez, researcher, Department of Marketing, Ghent University.
[2] Corresponding author: Dirk Van den Poel, Professor of Marketing Modeling, Analytical Customer Relationship Management, Ghent University, Department of Marketing, Tweekerkenstraat 2, B-9000 Gent, Belgium. Phone: +32 9 264 89 80, Fax: +32 9 264 42 79, Email: dirk.vandenpoel@UGent.be

For more papers about analytical CRM, please visit: www.crm.UGent.be or www.datamining.UGent.be

D/2007/7012/47

Separating financial from commercial customer churn:

A modeling step towards resolving the conflict between the sales and credit department.

Jonathan Burez & Dirk Van den Poel[3]

Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweekerkenstraat 2, B-9000 Gent, Belgium

Abstract

In subscription services, customers who leave the company can be divided into two groups: customers who do not renew their fixed-term contract at the end of that contract, and others who just stop paying during their contract to which they are legally bound. Those two separate processes are often modeled together in a so-called churn-prediction model, but are actually two different processes. The first type of churn can be considered *commercial* churn, i.e., customers making a studied choice not to renew their subscriptions. The second phenomenon is defined as *financial* churn, people who stop paying because they can no longer afford the service. The so-called marketing dilemma arises, as conflicting interests exist between the sales and marketing department on the one hand, and the legal and credit department on the other hand.

This paper shows that the two different processes mentioned can be separated by using information from the internal database of the company and that previous bad-payment behavior is more important as a driver for financial than for commercial churn. Finally, it is shown on real-life data that one can more accurately predict financial churn than commercial churn (increasing within period as well as out-of-period prediction performance). Conversely, when trying to persuade customers to stay with the company, the impact of 'loyalty' actions is far greater with potential *commercial* churners as compared to financial churners. Evidence comes from a real-life field experiment.

Keywords: Customer Intelligence, analytical customer relationship management (aCRM), customer churn, attrition research, commercial churn, financial churn, credit risk, out-of-period validation

[3] Corresponding author: Prof. Dr. Dirk Van den Poel, Professor of Marketing, Tweekerkenstraat 2, B-9000 Gent, Belgium, Phone: +32 9 264 89 80, Fax: +32 9 264 42 79, E-mail: Dirk.VandenPoel@UGent.be,

Visit http://www.crm.UGent.be to obtain some more papers about Analytical CRM/Customer Intelligence

Introduction

In this section, we first discuss the broad issue of conflicting interests of consumers, dealers and management, and within a company, of several departments (sales or marketing as opposed to credit and legal departments). The subject matter is introduced by a GM case study, followed by a description of subscription services.

GM's Marketing Dilemma

Among automakers, General Motors distinguished itself for its marketing techniques. GM led the field during the 1920s with their new focus on advertising, installment financing, and the art of styling low-priced automobiles (see e.g. Clarke, 1999). Sloan's autobiography became a classic in business management: Sloan (1963) presented a narrative for the efficient organization and performance of GM as a model for the modern corporation.

Sloan's account was accurate, but also incomplete: it overlooked the tensions between dealers and managers born out of the particular details in marketing automobiles – details that added up to what Clarke (2003) calls consumers' marketing dilemma. It was not just that management addressed the question of profitability in terms of efficiency – as consumers' design choices proliferated, for instance, GM faced a more complex job in producing vehicles; or inaccuracies in coordinating the production of cars with dealers' orders or consumers' demands resulted in unwanted cars (and costs) that either dealers or GM shouldered – but they also focused on profits in terms of market transactions – how dealers worked with the factory in ordering cars, and how they negotiated with consumers in selling cars. The three parties – consumers, dealers, and management – faced tension to the extent that one party's profit came at the expense of another party.

Subscription Services

Part of that same dilemma exists in the world of pay tv. Premium (cable) television content is a classic example of subscription services: you have the basic service (package of channels) with a few extra options. Pay-tv companies sell their service to customers typically via contracts with a fixed term[4]. The actual (initial) selling is done trough retailers. From the moment a customer signs a contract, every contact (for invoices, complaints, renewal, contract termination…) goes via the company itself.

The reminder of this study is organized as follows. In a literature review, we will first discuss behavioral loyalty and customer relationship management (CRM) in subscription services, and then return to the marketing dilemma faced in subscription services. Subsequently, we will build a case about, and hypothesize on churn definitions, variable importance, and the impact of marketing interventions. After a methodological part, the case study at the pay-TV company is described, and results are reported. These are used to analyze a field experiment, followed by conclusions, and directions for further research.

[4] We consider pay per view as a separate product.

Literature

Behavioral Loyalty and CRM

Broadly speaking, a subscription renewal decision is a type of repeat buying. Dick and Basu (1994, p. 99) state that "customer loyalty is viewed as the strength of the relationship between an individual's relative attitude and his/her repeat patronage". In a contractual setting and more specifically in subscription services, the behavioral loyalty measure is the renewal decision: does a customer renew his/her fixed term contract with the company.
Kumar and Ramani (2004) viewed customer relationship management (CRM) as the process of achieving and maintaining an ongoing relationship with customers across multiple customer touch points through differential and tailored treatment of individual customers based on their likely responses to alternative marketing programs, such that the contribution of each customer to the overall profitability of the firm is maximized. CRM in subscription services comes down to "attracting new customers", what Ryals (2005) calls offensive marketing, and "keeping the existing customers", known as defensive marketing (Ryals, 2005).
The importance of retaining customers should not be overlooked, as emphasized in articles published by Reichheld and Sasser (1990). Reinartz, Thomas and Kumar (2004) show that insufficient allocation to customer-retention efforts will have a greater impact on long-term customer profitability as compared to insufficient allocation to customer-acquisition efforts.

Marketing Dilemma at a Pay TV Company

For both offensive and defensive marketing in subscription services, the marketing dilemma as first described by Clarke (2003) manifests itself within the service provider. With regard to attracting new customers, retailers are motivated to sell as many new subscriptions as possible, as they are rewarded per subscription sold. Whether the customer is creditworthy or not, does not really bother them when selling a contract. But of course, both the legal department – that deals with illegal termination of a contract - and the credit department – that deals with arrears – are the victim of this overzealous selling. The credit department suffers because outstanding debts are a burden to the company. The legal department gets an overload of work, if customers terminate their contract midway their contractual term. Considering retaining the existing customer base, the marketing department tries to convince as many customers to renew their contract as possible. This can be done proactively and reactively. Reactively is the standard procedure. A customer who did not renew his or her contract is contacted in an attempt to convince him/her to do so anyway. Proactive targeting of defection-prone customers can be done using a churn-prediction model. But here again, the marketing dilemma emerges. As mentioned before the credit and the legal department only want creditworthy customers to renew contracts, whereas the sales department wants every existing customer to renew. Moreover, as proactive targeting is rather expensive, a company does not want to spend money on a customer who would renew anyway. On the contrary, it wants to convince those customers who can afford the service, but are not entirely sure they need the service.

Churn Definitions

In order to proactively target those customers who do not face financial distress, but are no longer absolutely convinced of the service itself, we redefine our churn definitions. Customers who died or moved abroad during the period of investigation are left out of the investigation in any case. This involuntary churn is detectable, but not predictable and absolutely not actionable marketing wise. Within the group of churners, a further distinction should be made between those customers who can no longer afford the service, and those who cancel the service because they do no longer want it. In the internal database of the service provider, we find no information on the reason why a customer defected; only about the way (and time) a customer terminated his/her contract. We hypothesize this is a good proxy for the difference between financial and commercial churn.

> H_1: Based on the way a customer terminates a contract – which we find in the internal database of the service provider – we will show that there exist three different types of churn: involuntary churn, financial churn and commercial churn.

Variable Importance

Occam's Razor, long admired, is usually interpreted to mean that simpler is better. Unfortunately, in prediction, accuracy and simplicity (interpretability) are in conflict (Breiman, 2001a, 2001b). When using Random Forests (Breiman, 2001 – see Methodology Section) as classification algorithm, no 'sign' information nor parameter estimate is returned for each of the independent variables. It does, however, provide a very good and unbiased importance measure. We will therefore only formulate hypotheses on the importance of some variables.

As this study focuses on the difference between customers defaulting their contracts, and customers deliberately terminating their contracts, we also include variables often used in credit scoring, but not (often) used in churn prediction. Predictor variables commonly used in credit scoring studies include various debt-ratio and other cash flow-oriented surrogates, employment time, home ownership, major credit card ownership, and representations of past payment history (e.g., Overstreet et al., 1992). Additional variables that can be added to the model include detailed credit bureau reports (e.g., Overstreet and Bradley, 1994) which give a better view on bad-payment behavior of a customer with other companies as well. Thomas (2000) compares credit-scoring systems to behavioral scoring systems, which allow lenders to make better decisions in managing existing clients by forecasting their future performance. The extra information is the repayment and ordering history of this customer. Desai et al. (1996) use credit bureau reports and the number of delinquent accounts over the last 12 months.

Somol et al. (2005) and Baesens et al. (2003) name credit history as one of the most important variables for classification of credit scoring in the *German credit* data set from the UCI repository. West (2000) uses two datasets, one of which is German credit data. Credit history emerges as the second most important variable when using both neural networks and mixture-of-experts networks. Piramuthu (1999, p. 259) states that "the past record of the borrower in meeting obligations is usually weighed heavily".

While the payment method itself is often used in churn prediction (e.g. Au et al, 2003), previous (bad) payment behavior is little used, and – to our knowledge – never mentioned as an important predictor. One study explicitly mentions the use of (bad) payment behavior: Yan, Wolniewicz and Dodier (2004) incorporate it in their customer behavior model, with which they attempt to prevent customer churn in telecommunications.

Mozer et al. (2000) state that the prediction of subscriber dissatisfaction should be used in combination with customer credit risk to decide on the retention campaign for that customer.

Drew et al. (2001) and Wei and Chiu (2002) do mention billing information as one of the sources for their churn model, but both studies do not mention bad-payment behavior specifically. Other studies did not include payment behavior at all: Au, Li and Ma (2003), Bhattacharya (1998), Bolton (1998), Weerahandi and Moitra (1995) to name a few.

H_2: Previous bad-payment behavior is more important in the financial churn prediction model than in the commercial churn prediction model.

Marketing Interventions

It has been noted that different types of interventions can have different impacts on customers, depending on their customer characteristics (e.g., De Wulf et al. 2001). CRM interventions can have different purposes. For example, the CRM literature generally distinguishes between interventions with a call for action focusing on cross-selling and retention (i.e., direct mailings) and more relationship-oriented instruments (i.e., relationship magazines) (Berry 1995, Bhattacharya and Bolton 1999, McDonald 1998).

Convincing a customer in financial distress to stay is both pointless and ineffective. While there is a great possibility that you can convince them to renew their contract, you enlarge the chance that they will default on their contract. Therefore, we assume that targeting financial defection-prone customers worsens customer retention. Enhancing the relationship with potential commercial churners, however, might improve customer retention. In the latter case, targeting with relationship-oriented marketing interventions will hence improve customer retention.

H_3: Targeting financial defection-prone customers worsens customer retention, whereas targeting commercial defection-prone customers with relationship-oriented marketing interventions enhances customer retention.

Methodology

Attrition modeling has been the topic of interest of a lot of studies recently. In a first part of this methodology section, two modeling techniques will be discussed. In a second part, the split sample-design will be discussed in more detail.

Modeling techniques

Two notable types of attrition models can be distinguished based on the time window of observation (Van den Poel and Lariviere, 2003). Static attrition models investigate churn behavior at a specific moment in time, while dynamic attrition models observe a set of individuals over a period of time. Models that make it possible to estimate the risk or hazard rates, which change over time are preferred, because they produce more accurate forecasts than models that do not allow variables to take another value over time (Weerahandi and Moitra, 1995). In this study, both a static attrition model, using random forests, and a dynamic one, using survival analysis, are constructed. Both techniques are explained briefly in the following paragraphs.

Random Forests

In a binary classification context, Decision Trees (DT) became very popular because of their ease and interpretability (Duda et al., 2001). Moreover, DTs have the ability to handle covariates measured at different measurement levels. One major problem with DTs is their high instability (Hastie et al. 2001). A small change in the data often results in very different series of splits, which is often suboptimal when validating the trained model. In the past, this problem was extensively researched.

It was Breiman (2001) who introduced a solution to the previously mentioned problem. The new classification technique is called: Random Forests. This technique uses a subset of *m* randomly chosen predictors to grow each tree on a bootstrap sample of the training data. Typically, this number of selected variables – i.e. *m* - is much lower than the total number of variables in the model. After a large number of trees is generated, each tree votes for the most popular class. By aggregating these votes over the different trees, each case is predicted a class label.

Random forests are already applied in several domains like bioinformatics, quantitative criminology, geology, pattern recognition, medicine, … . However, the applications in marketing are rare (Buckinx and Van den Poel, 2005; Larivière and Van den Poel, 2005). Random Forests is used in this study for five reasons: (1) Luo et al. (2004) stated that the predictive performance is among the best of the available techniques. (2) The outcomes of the classifier are very robust to outliers and noise (Breiman, 2001). (3) This classifier outputs useful internal estimates of error, strength, correlation and variable importance (Breiman, 2001). (4) Reasonable computation time is observed by Buckinx and Van den Poel (2005). (5) Random forests are easy to implement because there are only two free parameters to be set, namely *m*, the number of randomly chosen predictors, and the total number of trees to be grown. We follow Breiman's (2001) suggestions: *m* is set equal to the square root of the total number of variables - i.e. 13 because 171 explanatory variables are included in the model - and a large number of trees – in casu 300 – is chosen.

The Area Under the receiver operation Curve (AUC) is used as predictive performance indicator.

Survival analysis

Survival analysis is a class of statistical methods modeling the occurrence and timing of events (in casu: customer attrition) with the aim to establish descriptive or predictive models in which the risk of an event depends on covariates.

All of the standard approaches to survival analysis are probabilistic or stochastic. That is, the times at which the events occur are assumed to be realizations of some random process. It follows that *T*, the event time for some particular customer, is a random variable having a probability distribution.

Let us denote the probability density function (p.d.f.) of this variable by $f(t)$. The cumulative distribution function (c.d.f.) of variable *T*, denoted by $F(t)$. Hence,

$$F(t) = \Pr\{T \le t\}.$$

For some individuals the time to failure may be observed completely, whereas for others we only have partial observation until some specific censoring time *c*. In survival analysis, it is common to work with a closely related function called the survivor function, defined as

$$S(t) = \Pr\{T > t\} = 1 - F(t).$$

This leads to the following relationships:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}.$$

The hazard function is a central topic in the field of survival analysis and is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\{t \le T < t + \Delta t \mid T \ge t\}}{\Delta t}.$$

The aim of the definition is to quantify the instantaneous risk that the event will occur at time t.

The Kaplan-Meier estimator (also known as the Product Limit Estimator) provides an estimate of the survival function from life-time data (Kaplan and Meier, 1958). In customer churn prediction, one measures the length of time customers remain with the company.

Let $S(t)$ be the probability that an item from a given population will have a lifetime exceeding t. For a sample from this population of size N let the observed times until death of N sample members be

$$t_1 \leq t_2 \leq t_3 \leq \ldots \leq t_N \,.$$

Corresponding to each t_i is n_i, the number "at risk" just prior to time t_i, and d_i, the number of deaths at time t_i.The nonparametric maximum likelihood estimate of $S(t)$ is then a product of the form

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} \,.$$

When there is no censoring, n_i is just the number of survivors just prior to time t_i. With censoring, n_i is the number of survivors less the number of losses (censored cases). It is only those surviving cases that are still being observed (have not yet been censored) that are "at risk" of an (observed) death.

Split Sample Design

For the static churn model, a split sample design was used (see Figure 1). 60 percent of the data is used for training the model, 40 percent for validating the model. Churn is modeled on the training set, and that model is applied on the validation set and an out-of-period sample to check it’s lasting power.

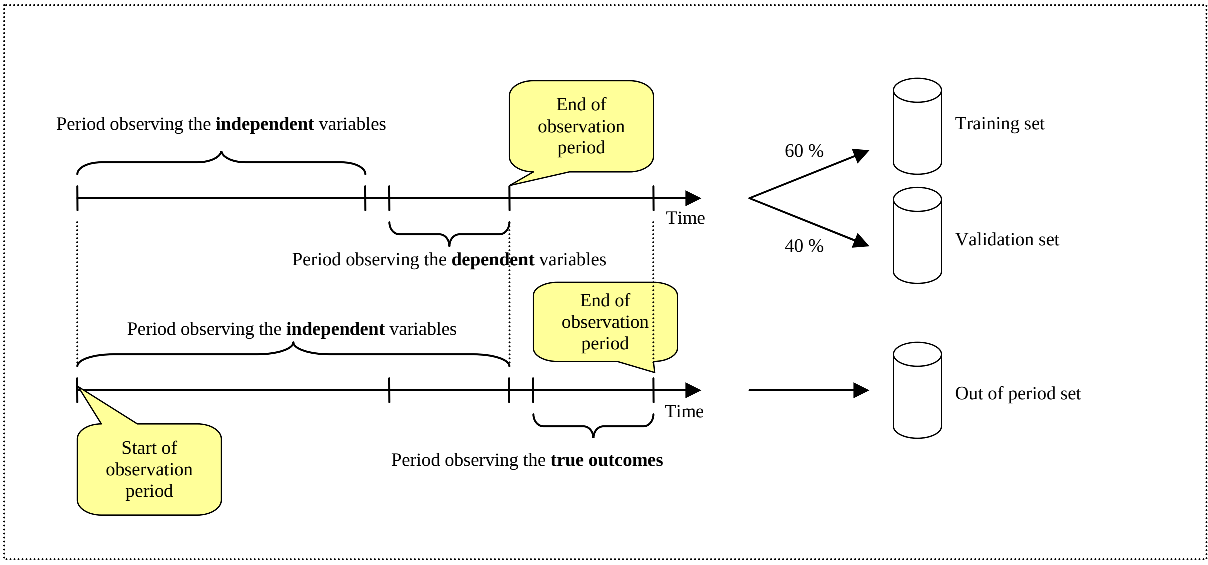

Figure 1: Graphical reflection of the time window used to build the static churn model

Three different types of independent variables will be modeled. They are defined later on. Furthermore, the impact on the churn prediction model of leaving some types of customers out of the training set will be investigated. This will become clear when defining the independent variables.

Case

Subscription services

Pay TV is a subscription service requiring the customer to pay a fixed monthly contribution only; there is no extra per-minute-of-use charge (as is the case for instance in the mobile phone industry). Pay-per-view is considered as a separate branch of the market, and is not considered here. At this European pay-TV company, all customers have a 12-month subscription. Cancelling within that 12-month period is not allowed, nor is prematurely reporting that a subscription will not be renewed (customers have to report that they will not renew their subscription during the last month of the 12-month contract). If nothing is reported, the subscription is automatically renewed for a period of 12 months. While in theory thus not possible (because it is being refused by the company), in practice premature contract termination does occur a lot. In almost all cases, this is due to bad payment behavior: a customer cannot or does not want to pay his subscription any more, and therefore does no longer receive the broadcasting signal. The customer will have to pay the full amount on his contract though!

Churn definitions

A customer can have many reasons for cancelling a subscription: children moved out of the house, the service became too expensive, not satisfied with the offered broadcast, not happy with complaint handling… The optimal situation would be to know the reason why somebody cancels his/her subscription. And thus be able to identify those customers who cancelled because they do no longer want the service versus those who do want the service but have *financial* limits. This information is not known to the company. However, information in the database enables us to make a distinction between those customers who came at the end of their annual subscription, but did not renew. Customers have to send a cancellation letter to do so. Not renewing is then a well thought through decision. Those customers are considered *commercial* churners. Secondly, you have those customers who stopped paying during their subscription. The company doesn't really know why they stop: do they really have too little money, or do they just hope to dispose of their subscription. A third type of churn is a combination of both financial and commercial churn, called *overall* churn.

Two remarks should be made. First of all, there will be customers who come at the end of their subscription, and do not renew because of a lack of money. Or the other way around, there will be customers who stop paying halfway through a subscription, for other reasons than being short of money. Secondly, note that there are some special cases: people who die, or move to another country… Those customers are what we call *involuntary* churners, and are handled separately: they are either considered non-churners or are left out of the model (see further).

Because of these three churn definitions (financial, commercial and overall churn), three versions of both the survival and the random forests model are created: one for commercial, one for financial and one for overall churn.

Churn Models

The dynamic model (using survival analysis) takes into account all customers who started since 1995 and, right censors those customers still active in 2002. The static model (using random forests) considers all customers active on 28/2/2002. The dependent variable is calculated on the subsequent year.

As explained in the *Split sample design* Section, different subsets of the training set will be used while modeling churn. This is done to investigate the effects of leaving some customers out of the training set, on the predictive power of the model. E.g. when modeling financial churn, does one have to leave the commercial churners (who in this case are considered non churners) in the training set, yes or no? Will one be able to better capture the behavior of financial churners when commercial churners are incorporated in the data? Four subsets are created (see Figure 2):

1. All = all customers
2. AC = active customers and commercial churners
3. AF = active customers and financial churners
4. ACF = active customers, commercial and financial churners

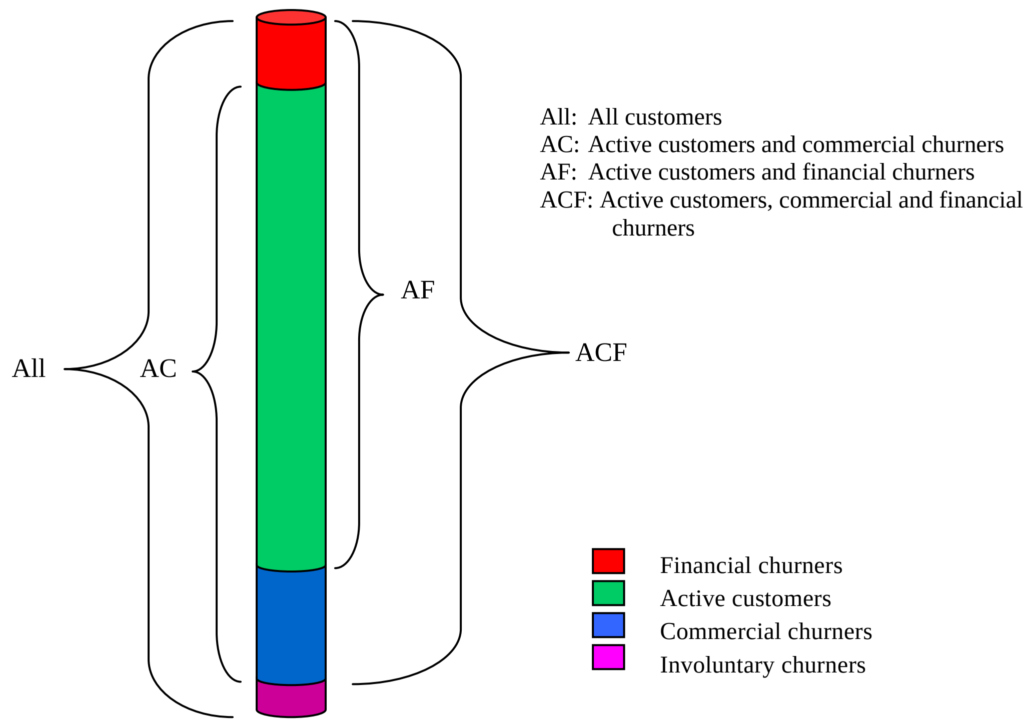

Figure 2: Subsetting the training set

Data

Data is extracted for this study from the data warehouse of the pay-TV company—a single integrated source of information combining data from all departments and services. It contains information about all customers that the pay-TV company has ever had. We used Oracle PL/SQL for data preparation and manipulation, and MATLAB and R to perform the statistical analyses.

For the dynamic model (using survival analysis), no explanatory variables are incorporated. All customers the service provider has ever had or still has were used (over 500,000). The static model had over 100,000 observations, of whom we collected 171 independent variables (Burez and Van den Poel, 2007). Those independent variables are listed in Appendix A and contain information about the current subscription (19 variables), socio-demographic information (12), previous bad-payment behavior (14), history information (32) and contact information (94).

Results

Survival Analysis

The survival curve of the customers of the pay-TV company (see Figure 3) clearly confirms the subscription renewal process: after one year, a huge drop can be noticed in the survival curve. A lot of newly attracted customers take a one-year subscription and afterwards leave the company. After two years, a smaller number attrites…

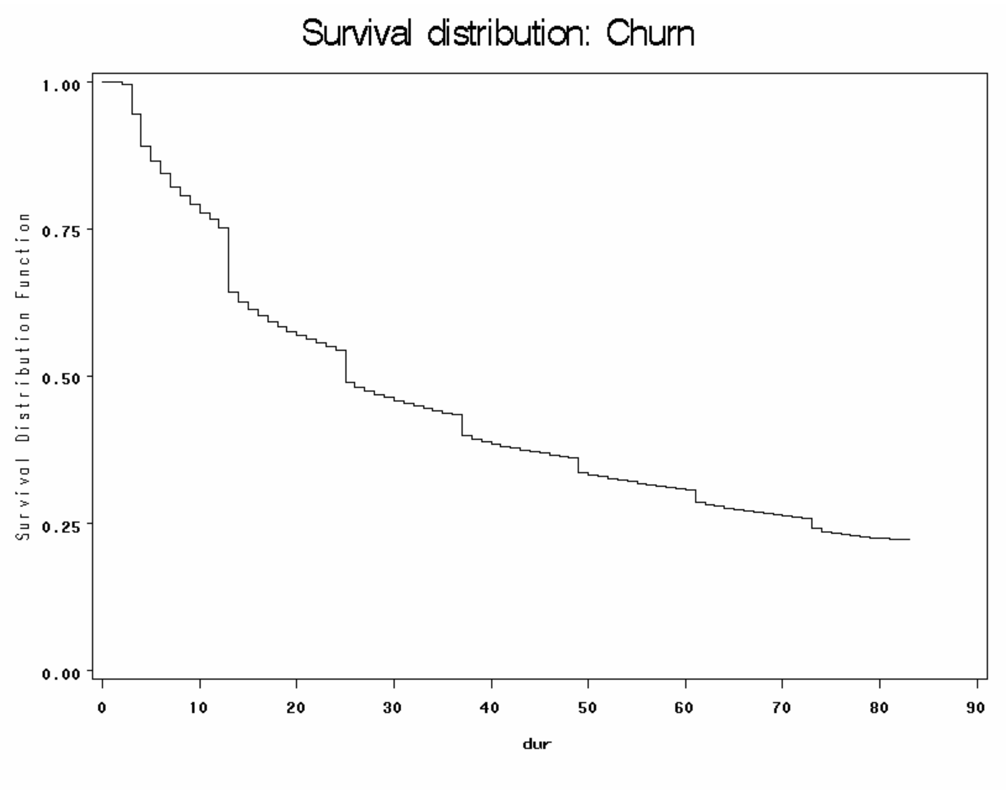

Figure 3: Survival distribution of overall churn

But another pattern emerges from the survival curve: those customers who stopped paying, for one reason or another, also represent a large proportion of churners. Especially in the first few years of the subscription, a lot of customers stop paying *during* a given subscription year, and thus stop receiving the signal.

Both phenomena together – churning after a year of subscription, and churning during the year – make that one out of three customers do not stay at the company more than one year, and more than one out of two customers leaves the company within two years.

The fact that there might be two different processes active was the reason to define three types of churn instead of just one (see Churn definitions). The survival curves for both commercial and financial churn are clearly different (see Figure 4). Commercial churners do not churn during their one-year subscription; they do it at the end of their subscription[5]. The biggest part of commercial churn happens after one year of subscription: almost 20% of the customers leave at that point.

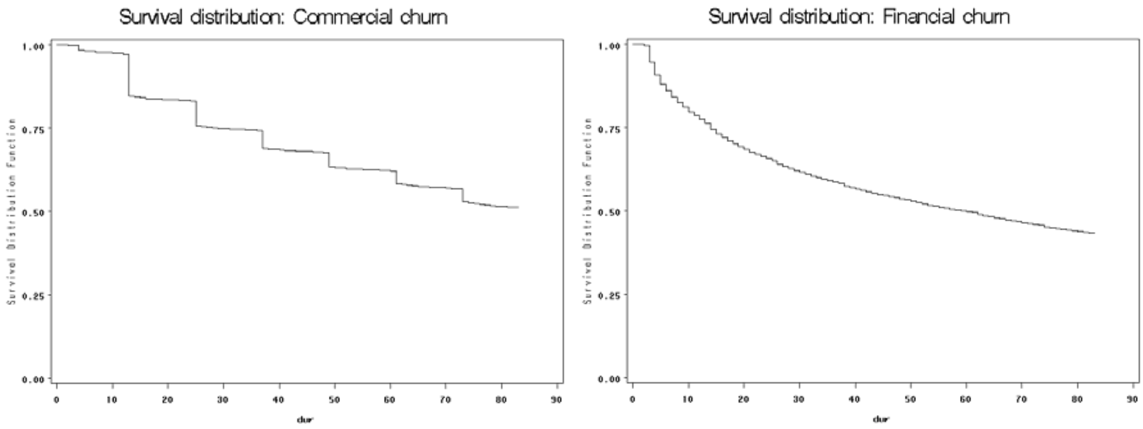

Figure 4: Survival distribution for commercial (left graph) and financial (right graph) churn

Financial churners leave at any point in their relationship with the company, in contrast to the yearly behavior for commercial churn. The curve decreases less and less, meaning that a customer is more at risk in the first few years of the subscription. The strange horizontal part of the curve in the first three months comes from the fact that customers have to pay a deposit plus more than two months subscription fees at the start of their first subscription. Hypothesis 1 is hereby confirmed, i.e., we are clearly able to visually distinguish the different types of churn behavior.

Churn Prediction

Now that we have shown that commercial and financial churn are two different processes, this information is used to make better churn predictions. From this point on, only the static models are used, and elaborated on.

The results of both the split sample design (training set, validation set and out-of-period sample), the subset creation within the training set, and the three churn definitions can be found in Table 1. In the left column, the dependent variable that is used for modeling is reported. Per dependent variable, the different subsets of the training set (as explained in Figure 2) are shown in the second column. Remember that *All* means that all observations in the training set are used when estimating a model, whereas *ACF* leaves out the involuntary churners, *AC* leaves out both involuntary and financial churners, and *AF* leaves out both involuntary and commercial churners. The predictive performance of all those models is shown for both the validation set (column 3) and the out-of-period sample (column 4).

An example from Table 1 will make this clearer: 90.3% is the crossing of *financial churn* and *ACF* for the *validation set*. *ACF* means that involuntary churners were left out of the training set. Because the dependent variable is financial churn, the commercial churners – who are left in the training set in this example – have zero as value on the dependent variable. The estimated model is then applied on the validation set, which yields an AUC performance of 90.3%.

[5] The horizontal parts are not perfectly horizontal because of individual exceptions.

Dependent variable used in the model	Subset of the training set used for estimating the model	Predictive performance on the validation set	Predictive performance on the out-of-period sample
Overall churn	All	80.5	74.6
	AC	69.4	61.1
	AF	75.7	71.0
	ACF	80.4	74.0
Commercial churn	All	75.7	68.6
	AC	74.3	67.9
	ACF	75.7	68.6
Financial churn	All	90.4	86.0
	AF	90.2	86.2
	ACF	90.3	86.2

Table 1: AUC performance results (shown in %) from the different models: 3 different dependent variables, and 4 different subsets of the training set. Predictive performance is reported on both the validation set and on an out-of-period set.

The following conclusions can be drawn from this table:

1. Financial churn is a lot easier to predict than commercial churn (and hence than overall churn) because on average all financial churn models score about 86% on the out-of-period dataset compared to 68% for commercial churn models. Predicting financial churn leans towards credit scoring, which is known for its very high classification accuracy, and where previous (bad) payment behavior is classically the key predictor (Verstraeten & Van den Poel, 2005).

2. There is no evidence that leaving the involuntary churners out of the training set influences predictive performance. Considering involuntary churners non-churners (i.e. keeping them in the training set, but with value zero on the dependent) yields similar performance for both financial and commercial churn.

3. Out-of-period forecasts are a lot more difficult than validation-set predictions since we observe a substantial drop-off between the predictive performance on the validation set (column 3) and the predictive performance on the out-of-period sample (column 4). It is thus very important to keep the models up to date. The performance of financial churn seems to be less affected than that of commercial churn, however.

It should be noted that any combination of both a financial and a commercial churn model does not come up with improved predictions for overall churn compared to the overall churn model itself.

Variable Importance

A nice feature of Random Forests is that it returns variable importance, by using internal out-of-bag estimates. For the financial churn prediction model, 136 variables are significant out of 171, for the commercial churn prediction model, 143 variables are significant (see Appendices B and C for the variable importance of all variables).

Table 2 shows the most important variables for the financial churn prediction model. The last three columns represent respectively the z-value, the ranking, and the significance. For those variables, we also include in columns 2-5 the importance of those variables in the commercial churn prediction model.

The table clearly shows that we can confirm hypothesis 2. Variables like *canc_99P*, *nbr_canc*, *canc_78P* and *rap_rec*, all variables describing the previous bad-payment behavior of a customer, are far more important in the financial compared to the commercial churn prediction model.

	Commercial Churn			Financial Churn		
Variable Name	**Z score**	**Importance ranking**	**p**	**Z score**	**Importance ranking**	**p**
canc_99P	28,443	6	0	52,275	1	0
nbr_canc	16,232	29	0	52,046	2	0
length_subs	31,786	4	0	32,944	3	0
lor	29,396	5	0	31,135	4	0
noc2	26,218	7	0	29,33	5	0
monetary	25,968	8	0	29,294	6	0
tel_fix	12,14	50	0	26,722	7	0
nov2	23,86	10	0	26,457	8	0
canc_78P	7,875	74	0	25,438	9	0
rec_og	38,233	2	0	23,967	10	0
og_3	22,042	12	0	22,054	11	0
cash_free	20,817	15	0	19,613	12	0
call_positive	14,008	38	0	18,406	13	0
no_renew	17,431	22	0	18,277	14	0
at_ttcmens	14,306	36	0	18,228	15	0
rap_rec	13,197	43	0	17,817	16	0
in_call_A	17,487	21	0	17,525	17	0
rap_amount_c	24,908	9	0	17,212	18	0
markov2	17,325	23	0	17,149	19	0
og_19	15,981	30	0	15,868	20	0

Table 2: Variable Importance for Financial versus Commercial Churn

Churn Prevention: A Field Experiment

A large-scale field experiment was conducted to test the influence of three customer retention actions on relational behavior: free movie tickets, an invitation to a unique event, and a satisfaction questionnaire. All Pay-TV customers who had to renew their subscription in December 2003, and of whom the company had a telephone number, were candidates for one of the actions. 24,185 customers satisfied this condition. We selected 7,350 customers with the highest churn probability (labeled top 30%).

Those customers were divided into four groups via systematic sampling (Churchill and Iacobucci, 2002, p.484): One group of 1,050 customers would receive two free movie tickets, a second group of 2,100 customers would receive an invitation for two to a unique event, a third group of 2,100 customers was asked to respond to a satisfaction questionnaire by telephone; the fourth group (2,100 customers) served as a control group.

The questionnaire was conducted by telephone by a specialized firm, two and a half months before the moment in time at which customers had to renew their subscription (October 17, 2003). The interviewer requested participation from the “person in your household who was the main promoter of subscribing to Pay-TV”. The survey took about 7 minutes to complete on average, and inquired into a number of motivations, usage and satisfaction questions eliciting respondents’ evaluation of specific features (e.g. price, programming, movies broadcasted, sports programs…) followed by the general question, “Overall, how satisfied are you with the Pay-TV company?”

The results of the field experiment itself are described in Burez and Van den Poel (2007).

			Relationship Marketing Interventions		
Dependent variable	Predicted probability	Control group	Avant-premiere	Telephonic satisfaction questionnaire	Movie tickets
Commercial churn	All	13.84	11.03	9.65	11.00
	High	18.23	13.99	11.27	12.95
	Low	6.50	6.62	6.79	7.52
Financial churn	All	3.63	4.35	3.95	3.70
	High	4.13	4.73	5.11	4.15
	Low	3.10	3.93	2.70	3.28
Overall churn		17.99	15.86	14.00	15.00

Table 3: Results of the field experiment - a comparison between financial and commercial churn

While at the time of the field experiment, only an overall churn model was used, it can now be investigated whether a commercial and/or financial churn model would have improved customer retention. Table 3 gives the churn rates of the different groups, targeted with different relationship marketing interventions (RMIs). The leftmost column shows the dependent variable used for evaluation of the impact of the different RMIs. The second column indicates the predicted probability for the type of churn in column 1. E.g. for commercial churn, predicted probability *high* are those customers who were selected for the field experiment (because they were among the top 30% of the customers with the highest chance of overall churning) and are in the top 30% of the customers with the highest probability of commercial churning. *Low* customers were among the top 30% with highest overall churning propensity, but not among the top 30% with highest commercial churning chances. The same goes for financial churning. Note that for overall churn, because of the setup of the experiment, the predicted probability is high for all customers in the experiment.

Table 3 shows that the effect of all three actions was much higher on the at-risk commercial churners whereas the effect on the financial churners was opposite. Clearly the churn reduction was much bigger for commercial churn compared to overall churn. Financial churn could not be reduced; on the contrary, the churn rate slightly increased compared to the control group. These findings confirm our third hypothesis.

Conclusion

This paper investigated churn at a pay-TV company, using both static and dynamic churn prediction models. While the exact reasons for customer attrition are unknown, the way a customer ends a subscription turns out to be a good proxy for it. Based on this, two distinct types of churn were defined: commercial churn (not renewing a subscription) and financial churn (no longer paying invoices of a current subscription). Previous bad-payment behavior is far more important in financial churn prediction compared to commercial churn prediction. Financial churn prediction, which is similar to credit scoring, is easier to predict than commercial churn.

While financial churn is easier to predict, commercial churn is much easier to prevent. A field experiment pointed out that you can convince commercially defection-prone customers to stay at your company, while this is not possible for financially defection-prone customers. This also diminishes the conflict between the sales and the credit department: instead of targeting all defection-prone customers, only the commercially defection-prone (and allegedly creditworthy) customers are now targeted. This may reduce the inherent conflict between those two departments. It is very important for marketers to know that different types of churn exist, and that they should be prevented with different actions.

Directions for further research

In this study, we differentiate between commercial and financial churn, based on the way a customer terminates his/her contract. We do so because this information is available in the internal database of the company. Questionnaire data could learn us more on the reasons for contract termination, and thus enable us to make a better distinction between customers who churn for financial reasons, and others.

A second opportunity for further research lays with the proportional hazards models. The estimation of the baseline hazard (as for example proposed by Dekimpe et al.) makes it possible for proportional hazards models to make individual churn predictions (e.g. in R). While thus theoretically possible, considering the large potential number of time-varying variables, it would take a lot of computation time and huge memory requirements to come up with the predictions. Hence it will be very hard to make this operational with current technology.

The a posteriori interpretation of the field experiment results with regard to financial and commercial churn should be handled carefully. The setup of a new field experiment, with a selection based on a commercial churn prediction model is to confirm our results. Further research could inquire into the cost-effectiveness of different incentives for different types of churn.

Acknowledgements

We are indebted to the late Professor Leo Breiman for making available the fortran code of his Random Forests method. Moreover, this research is supported by the IAP research network grant number P6/03 of the Belgian government (Belgian Science Policy).

References

Allison, P.D. (1995). Survival Analysis Using The SAS System, SAS Institute Inc., North Carolina.

Au, T., Li, S. and Ma, G. (2003). Applying and Evaluating Models to Predict Customer Attrition Using Data Mining Techniques. *Journal of Comparative International Management, 6* (1).

Au, W., Chan, K.C.C. and Yao, X. (2003). A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction. *IEEE Transactions On Evolutionary Computation,7* (6), 532-545.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J. and Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society, 54*(5): 627-35.

Berry, L.L. (1995). Relationship marketing of services – growing interest, emerging perspectives. *Journal of the Academy of Marketing Science, 23* (4), 236-45.

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.

Bhattacharya, C. B. and Bolton, R.N. (1999). Relationship marketing in mass markets. Sheth, J.N. and Parvatiyar, A., eds. Handbook of Relationship Marketing. Sage Publications, Thousand Oaks, CA, 327-354.

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. (2001) Statistical Modeling: The Two Cultures. *Statistical Science, 16* (3), 199-231.

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

Brill, J. (1998). The importance of credit scoring models in improving cash flow and collections. *Business Credit, 100*(1), 16-17.

Buckinx, W. and 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.

Burez, J. and Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications, 32* (2), 277-288.

Churchill, G. A., Jr., & Iacobucci, D. (2002). Marketing research methodological foundations (8th ed.). Mason, Ohio: South-Western Thompson Learning.

Clarke, S. (1999). Managing Design: The Art and Colour Section at General Motors, 1927–1941. *Journal of Design History, 12*, 65–79.

Clarke, S. (2003). Closing the Deal: GM's Marketing Dilemma and its Franchised Dealers, 1921-41. *Business History, 45* (1), 60-79.

Cox, D.R. & Snell E.J. (1989). Analysis of Binary Data, 2nd Edition, Chapman and Hall, London.

Dekimpe, M., Van de Gucht, L., Hanssens, D. & Powers, K. (1998). Long-Run Abstinence After Narcotics Abuse: What Are the Odds? *Management Science, 44*(11), 1478-1492.

Desai, V.S., Crook, J.N. and Overstreet, G.A.Jr. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research, 95*, 24-37.

De Wulf, K., Odekerken-Schröder, G., Iacobucci, D. (2001). Investments in consumer relationships: a cross-country and cross-industry exploration. *Journal of Marketing, 65* (4), 33-50.

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

Drew, J.H., Mani, D.R., Betz, A.L., and Datta, P. (2001). Targeting Customers with Statistical and Data-Mining Techniques. *Journal of Service Research, 3* (3), 205 - 219.

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

Hastie, T., Tibshirani, R. and Friedman, J. (2001). The elements of statistical learning: data mining, inference and prediction. Springer-Verlag.

Kaplan, E.L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*, 457-481.

Kumar, V. and Ramani, G. (2004). Taking Customer Lifetime Value Analysis to the Next Level. *Journal of Integrated Communications*, 27-33.

Larivière B. and Van den Poel D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems With Applications, 29* (2), 472-484.

Lucas, A. (2001). Statistical challenges in credit card issuing. *Applied Stochastic Models in Business and Industry, 17*(1), 83-92.

Luo, T., Kramer K., Goldgof, D.B., Hall, L.O., Samson, S., Remsen, A. and Hopkins, T. (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.

Magee, L. (1990). R2 measures based on Wald and likelihood ratio joint significance tests. *American Statistician, 44*, 250-253.

McDonald, William J. (1998). Direct Marketing: An Integrated Approach. Irwin/McGraw-Hill, Boston, MA.

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E. and Kaushansky, H. (2000). Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry. *IEEE Trans. Neural Networks, 11* (3), 690-696.

Overstreet, Jr., G.A., and Bradley, Jr., E.L. (1994). Applicability of generic linear scoring models in the USA Credit Union environment: Further analysis, *Working Paper.* University of Virginia.

Overstreet, Jr., G.A., Bradley, Jr., E.L., and Kemp, R.S. (1992). The flat-maximum effect and generic linear scoring model: A test. *IMA Journal of Mathematics Applied in Business and Industry 4*, 97-109.

Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research, 112* (2), 310–321.

Reichheld, F. and Sasser, W. (1990). Zero defects: quality comes to services. *Harvard Business Review, 68* (5), 105-111.

Reinartz, W., Thomas, J. and Kumar, V. (2005). Balancing Acquisition and Retention Resources to Maximize Profitability. *Journal of Marketing, 69* (1), 63-79.

Ryals, L. (2005). Making Customer Relationship Management Work: The Measurement and Profitable Management of Customer Relationships. *Journal of Marketing, 69* (4), 252-261.

Sloan, Alfred P. Jr. (1963), My Years with General Motors. DoubleDay.

Somol, P., Baessens, B., Pudil, P. and Vanthienen, J. (2005) Filter- versus Wrapper-based Feature Selection for Credit Scoring. *International Journal of Intelligent Systems, 20*, 985-999.

Therneau, T.M. & Grambsch, P.M. (2000). Modeling Survival Data: Extending the Cox Model, Springer, New York.

Thomas, L.C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *Internation Journal of Forecasting*, *16*(2), 149-172.

Thomas, L.C., Oliver, R.W. & Hand D.J. (2005). A survey of the issues in consumer credit modeling research. *Journal of the Operational Research Society, 56* (9). 1006-1015.

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

Venkatesan, R. & Kumar V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing. 68*(4): 106-125.

Verstraeten, G. & Van den Poel, D. (2005). The impact of sample bias on consumer credit scoring performance and profitability. *Journal of the operational research society. 56*(8): 981-992.

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

Wei, C.P. and Chiu, I.T. (2002). Tuning telecommunications call detail to churn prediction: A data mining approach. *Expert Systems with Applications, 23* (2), 103–112.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research, 27*, 1131-1152.

Yan, L., Wolniewicz, R. H., and Dodier, R. (2004). Predicting Customer Behavior in Telecommunications. *IEEE Intelligent Systems, 19* (2), 50-58.

Appendix A
List of variables in the static churn prediction model

The ID-variable		At_numabo	ID	Customer number

Dependent variable		Churn	Binary	Whether the customer churned between 28/2/2002 and 28/2/2003

Current subscription	1.	At_ttcmens	Continuous	Monthly payment of the current subscription
	2.	At_duree	Continuous	Length in months of the current subscription
	3.	Product1	Binary	Type of product the customer has now: product1=1➔basic service of the pay-TV company only, product2=1➔premium channels only, both=1➔ basic service as well as premium channels
	4.	Product2	Binary	
	5.	Tech	Binary	The technology used: 0 = analog, 1 = digital
	6.	Nbr_products	Continuous	Number of subscriptions the customer has on 28/2/2002 (based on the linked numabo's)
	7.	Domi_monthly	Binary	Does the customer pay monthly with direct debit? (1 = yes, 0 = no)
	8.	Exp_month_1	Binary	Month in which the current subscription will expire.
	9.	Exp_month_2	Binary	
	10.	Exp_month_3	Binary	
	11.	Exp_month_4	Binary	
	12.	Exp_month_5	Binary	
	13.	Exp_month_6	Binary	
	14.	Exp_month_7	Binary	
	15.	Exp_month_8	Binary	
	16.	Exp_month_9	Binary	
	17.	Exp_month_10	Binary	
	18.	Exp_month_11	Binary	
	19.	Opt	Binary	Does the customer have any options? 1 = yes, 0 = no

Socio-demographic variables	20.	Ccivil	Binary	Civil status of the customer (0 = Me, Ml, SM; 1 = rest)
	21.	Age	Continuous	Age of the customer in years at start numabo
	22.	Age_mv	Binary	Is the date of birth missing? 1 = yes, 0 = no
	23.	Tel_mob	Binary	Do we have a telephone number of the customer (tel_fix = 0 and tel_mob = 0)? If yes, is it a fixed (tel_fix = 1) or a mobile (tel_mob = 1) number?
	24.	Tel_fix	Binary	
	25.	Info_mv	Binary	Is the information field filled in? 1 = not filled in, 0 = filled in
	26.	Prov1	Binary	Variables indicating what province the customer comes from. If all = 0, then the customer lives in Flanders or abroad.
	27.	Prov4	Binary	
	28.	Prov5	Binary	
	29.	Prov6	Binary	
	30.	Prov7	Binary	
	31.	Business	Binary	Is the customer a company, based on the fields information and name? 1 = yes, 0 = no

Previous bad payment behavior variables	32.	No_rappels	Continuous	Total number of reminders
	33.	No_cycles	Continuous	Number of cycles of bad payments
	34.	Rap_smooth	Continuous	Number of monthly newsletter (etat = 2)
	35.	Rap_cut_adv	Continuous	Number of cutting advertisements (etat = 3 or 4)
	36.	Rap_deactiv	Continuous	Number of deactivations of the decoder (etat = 5)
	37.	Rap_to_pay	Continuous	Number of notices to pay (etat = 6)
	38.	Rap_debt_rec	Continuous	Number of debt recoveries (etat = 7 or 9)
	39.	Rap_not_recouvr	Continuous	Number of irrecoverable (etat = 8)
	40.	Rap_outb	Continuous	Number of outbounds (etat = 10)
	41.	Rap_amount	Continuous	Amount of all reminders
	42.	Rap_amount_c	Continuous	Amount of reminder times the number of cycle stage
	43.	Rap_proc_open	Binary	Is there a reminder process running on 28/2/2002 (1 = yes, 0 = no)
	44.	Rap_rec	Continuous	Recency since last reminder
	45.	Rap_mv	Binary	If there were any reminders in the past (1 = no, 0 = yes)

History variables	46.	Lor	Continuous	Length of relationship at 28/2/2002
	47.	Length_subs	Continuous	Length of actual subscriptions (lor – gaps)
	48.	Ind_subs	Continuous	Length_subs / lor
	49.	Markov	Continuous	Markov value of the fourth order (no distinction is made between analogue or digital)
	50.	Markov_dummy	Binary	Indicating that there were less then 50 observations for this Markov chain (Hence, Markov value of a lower order is used)
	51.	Markov2	Continuous	Markov value of the third order (distinction between analogue and digital)
	52.	Markov2_dummy	Continuous	Indicating that there were less then 50 observations for this Markov chain (Hence, Markov value of a lower order is used)
	53.	Noc	Continuous	Absolute number of contracts
	54.	Noc2	Continuous	Number of contracts relative to the lor (i.e. per year)
	55.	Nov	Continuous	Absolute number of versions of a contract
	56.	Nov2	Continuous	Number of versions of a contract relative to the lor
	57.	Lc	Binary	Are there 2 linked contracts for a customer, one for a smartcard and one for a setup box?
	58.	Monetary	Continuous	The monetary value of the customer.
	59.	Dist_display	Binary	If a decoder was bought at ‘display’ distributor
	60.	Nbr_int	Continuous	The number of times this customer has been without subscription within a numabo or between linked numabo’s.
	61.	Cash_free	Continuous	Summation of what the customer got for free in cash over his lor.
	62.	Terme	Continuous	Summation of what the customer got for free in monthly terms over his lor.
	63.	Parrain	Binary	Does the customer have a godfather (1=yes, 0=no)
	64.	No_godchild	Continuous	Of how many customers is this customer the godfather?
	65.	Camp_c	Continuous	Was the customer recruited with a campaign ? ➔if no, all 5 variables are 0 ➔if yes, one of the variables equals 1, indicating the type of campaign c=coupons, d=distributor, e=leaflet, p=member-gets-member, x=else
	66.	Camp_d	Continuous	
	67.	Camp_e	Continuous	
	68.	Camp_p	Continuous	
	69.	Camp_x	Continuous	
	70.	Payeur	Continuous	Number of subscriptions this customer pays for.
	71.	Num_numabo	Continuous	The number of linked numabo’s the customer has (had) up to 28/2/2002
	72.	Canc_78P	Continuous	Number of cancellations ‘Annul ancien abo (Transf)’
	73.	Canc_99P	Continuous	Number of cancellations ‘Coupure inform-mauvais payeur’
	74.	Canc_991	Continuous	Number of cancellations ‘Reprise inform-mauvais payeur’
	75.	Nbr_canc	Continuous	Number of cancellations (all types)
	76.	No_renew	Continuous	Number of times that customer renewed his subscription
	77.	Canc_tot	Continuous	Number of total cancellations (csupr ending on ‘T’)

	78.	Nbr_son	Continuous	Number of questionnaires the customer filled in

	79.	Rec_son	Continuous	The number of days that elapsed since filling in the last questionnaire
	80.	Son_mv	Continuous	If nbr_son = 0, son_mv = 1 ; else son_mv = 0.
	81.	Eg_2	Continuous	Number of generic letters of a certain type the pay-TV company received from a certain customer. ➔2=entrants, 6=sondages, 8=marketing
	82.	Eg_6	Continuous	
	83.	Eg_8	Continuous	
	84.	Eg_else	Continuous	
	85.	Ep_1	Continuous	Number of personal letters of a certain type the pay-TV company received from a certain customer. ➔1=bons restitution, 2=entrants, 5=messages cat
	86.	Ep_2	Continuous	
	87.	Ep_5	Continuous	
	88.	Ep_else	Continuous	
	89.	Og_3	Continuous	Number of generic letters of a certain type the pay-TV company sent to a certain customer. ➔3=sortants circul., 6=sondages, 7=divers, 8=marketing, 9=cadeaux autres, 10=cadeaux parrainage 11=cinema, 13=évenements, 14=sports, 15=jeux concours, 16=transformation, 19=parrainage offre, 21=reabo-routeur, 22=reabo-circulaires, 25=sortants magazine/gazette
	90.	Og_6	Continuous	
	91.	Og_7	Continuous	
	92.	Og_8	Continuous	
	93.	Og_9	Continuous	
	94.	Og_10	Continuous	
	95.	Og_11	Continuous	
	96.	Og_13	Continuous	
	97.	Og_14	Continuous	
	98.	Og_15	Continuous	
	99.	Og_16	Continuous	
	100.	Og_19	Continuous	
	101.	Og_21	Continuous	
	102.	Og_22	Continuous	
	103.	Og_25	Continuous	
	104.	Og_else	Continuous	
	105.	Op	Continuous	Number of personal letters the pay-TV company sent to a certain customer.
	106.	Rec_eg	Continuous	How long ago (on 31/3/2002) did the pay-TV company receive a letter from a certain customer?
	107.	Eg_mv	Binary	
	108.	Rec_ep	Continuous	
	109.	Ep_mv	Binary	
	110.	Rec_og	Continuous	How long ago (on 31/3/2002) did the pay-TV company send a letter to a certain customer?
	111.	Og_mv	Binary	
	112.	Rec_op	Continuous	
	113.	Op_mv	Binary	
	114.	In_call_A	Continuous	Number of times customer called the pay-TV company, (relative to length of relationship for last 3,5 years) concerning: A – his subscription, C – his contract D – to ask for additional information , E – other changes, F – payments G – call interrupted by subsciber O – games, gifts, member-gets-member P – program R – reabo T – technical problems
	115.	In_call_C	Continuous	
	116.	In_call_D	Continuous	
	117.	In_call_E	Continuous	
	118.	In_call_F	Continuous	
	119.	In_call_G	Continuous	
	120.	In_call_O	Continuous	
	121.	In_call_P	Continuous	
	122.	In_call_R	Continuous	
	123.	In_call_T	Continuous	
	124.	Distr_call_C	Continuous	Number of times distributor called concerning subscriber. (explanation of types as above)
	125.	Distr_call_D	Continuous	
	126.	Distr_call_E	Continuous	
	127.	Distr_call_R	Continuous	

	128.	After_call_C_	Continuous	“After call work” on the subscriber (types as above, X – else)
	129.	After_call_E	Continuous	
	130.	After_call_F	Continuous	
	131.	After_call_R	Continuous	
	132.	After_call_T	Continuous	
	133.	After_call_X	Continuous	
	134.	Inbound_mail	Continuous	Number of inbound mails
	135.	Stand_contact	Continuous	Number of contacts between customer and Stand
	136.	Outbound_calls_pos	Continuous	Number of positive outbound calls
	137.	Outbound_calls_neg	Continuous	Number of negative outbound calls
	138.	Rec_inbound_A	Continuous	Recency of the last inbound call of each type
	139.	Rec_inbound_C	Continuous	
	140.	Rec_inbound_D	Continuous	
	141.	Rec_inbound_E	Continuous	
	142.	Rec_inbound_F	Continuous	
	143.	Rec_inbound_G	Continuous	
	144.	Rec_inbound_O	Continuous	
	145.	Rec_inbound_P	Continuous	
	146.	Rec_inbound_R	Continuous	
	147.	Rec_inbound_T	Continuous	
	148.	Rec_inbound_mail	Continuous	Recency of last inbound mail
	149.	Rec_stand_contact	Continuous	Recency of last Stand contact
	150.	Rec_out_pos	Continuous	Recency of last positive outbound call
	151.	Rec_out_neg	Continuous	Recency of last negative outbound call
	152.	Rec_inbound_A_mv	Binary	If there was no call of certain type then = 1, else 0
	153.	Rec_inbound_C_mv	Binary	
	154.	Rec_inbound_D_mv	Binary	
	155.	Rec_inbound_E_mv	Binary	
	156.	Rec_inbound_F_mv	Binary	
	157.	Rec_inbound_G_mv	Binary	
	158.	Rec_inbound_O_mv	Binary	
	159.	Rec_inbound_P_mv	Binary	
	160.	Rec_inbound_R_mv	Binary	
	161.	Rec_inbound_T_mv	Binary	
	162.	Rec_inbound_mail_mv	Binary	
	163.	Rec_stand_contact_mv	Binary	
	164.	Rec_out_pos_mv	Binary	
	165.	Rec_out_neg_mv	Binary	
	166.	Call_positive	Continuous	Number of positive and negative inbound calls
	167.	Call_negative	Continuous	
	168.	Rec_positive	Continuous	Recency of last positive outbound call
	169.	Rec_negative	Continuous	Recency of last negative outbound call
	170.	Rec_pos_mv	Continuous	If there was no call of certain type then = 1, else 0
	171.	Rec_neg_mv	Continuous	

Appendix B
Variable importance in commercial churn model

Variable Name	Z score	Importance ranking	p
info_mv	42,518	1	0
rec_og	38,233	2	0
eg_2	38,201	3	0
length_subs	31,786	4	0
lor	29,396	5	0
canc_99P	28,443	6	0
noc2	26,218	7	0
monetary	25,968	8	0
rap_amount_c	24,908	9	0
nov2	23,86	10	0
rap_amount	23,471	11	0
og_3	22,042	12	0
rec_inbound_A	21,392	13	0
call_negative	20,932	14	0
cash_free	20,817	15	0
rec_positive	20,488	16	0
rec_ep	20,136	17	0
no_rappels	19,713	18	0
in_call_T	18,865	19	0
rec_negative	18,721	20	0
in_call_A	17,487	21	0
no_renew	17,431	22	0
markov2	17,325	23	0
in_call_F	17,32	24	0
in_call_R	16,958	25	0
rec_inbound_R	16,889	26	0
no_cycles	16,237	27	0
rec_inbound_R_mv	16,232	28	0
nbr_canc	16,232	29	0
og_19	15,981	30	0
og_22	15,803	31	0
rap_cut_adv	15,71	32	0
rec_inbound_F	15,624	33	0
canc_991	15,286	34	0
og_8	14,62	35	0
at_ttcmens	14,306	36	0
rec_inbound_A_mv	14,218	37	0
call_positive	14,008	38	0
rec_inbound_D	13,983	39	0
exp_month_4	13,753	40	0
rap_deactiv	13,426	41	0
rec_inbound_T	13,337	42	0
rap_rec	13,197	43	0
terme	13,127	44	0
markov	13,113	45	0
rec_inbound_F_mv	12,918	46	0
rap_smooth	12,901	47	0
ep_2	12,867	48	0
tel_mob	12,783	49	0
tel_fix	12,14	50	0

Variable Name	Z score	Importance ranking	p
ind_subs	11,926	51	0
ep_mv	11,406	52	0
nov	11,351	53	0
in_call_D	11,221	54	0
og_16	10,95	55	0
og_6	10,897	56	0
rec_eg	10,877	57	0
rec_inbound_D_mv	9,614	58	0
rec_neg_mv	9,586	59	0
rec_inbound_T_mv	9,561	60	0
rec_inbound_E	9,27	61	0
rec_inbound_C	9,255	62	0
age	8,959	63	0
age_mv	8,833	64	0
og_mv	8,633	65	0
rap_mv	8,432	66	0
in_call_C	8,408	67	0
ep_5	8,322	68	0
rap_to_pay	8,199	69	0
at_duree	8,161	70	0
og_11	8,077	71	0
product2	8,049	72	0
noc	7,993	73	0
canc_78P	7,875	74	0
in_call_E	7,746	75	0
no_godchild	7,735	76	0
eg_mv	7,714	77	0
og_7	7,395	78	0
inbound_mail	7,089	79	0
exp_month_1	6,974	80	0
tech	6,921	81	0
rec_inbound_E_mv	6,793	82	0
rec_inbound_C_mv	6,764	83	0
rec_inbound_mail_mv	6,757	84	0
rec_pos_mv	6,718	85	0
rec_inbound_mail	6,623	86	0
rec_stand_contact_mv	6,547	87	0
after_call_X	6,534	88	0
og_10	6,477	89	0
nbr_int	6,473	90	0
after_call_C	6,413	91	0
og_14	6,34	92	0
in_call_G	6,329	93	0
exp_month_3	5,879	94	0
og_15	5,78	95	0
after_call_R	5,715	96	0
domi_monthly	5,707	97	0
num_numabo	5,581	98	0
parrain	5,428	99	0
product1	5,384	100	0
eg_6	5,377	101	0
lc	5,194	102	0
eg_8	5,091	103	0
og_25	5,051	104	0
markov_dummy	4,869	105	0
exp_month_5	4,841	106	0
after_call_E	4,638	107	0
exp_month_2	4,456	108	0
camp_d	4,429	109	0
ep_1	4,275	110	0

Variable Name	Z score	Importance ranking	p
op	4,268	111	0
opt	4,257	112	0
camp_p	4,066	113	0
rec_out_pos	4,027	114	0
og_21	3,841	115	0
og_13	3,812	116	0
after_call_T	3,777	117	0
markov2_dummy	3,73	118	0
nbr_products	3,603	119	0
after_call_F	3,465	120	0
og_else	3,199	121	0,001
rec_inbound_O	3,007	122	0,001
payeur	2,929	123	0,002
in_call_P	2,902	124	0,002
canc_tot	2,873	125	0,002
distr_call_E	2,873	126	0,002
distr_call_C	2,859	127	0,002
stand_contact	2,785	128	0,003
camp_x	2,752	129	0,003
exp_month_8	2,447	130	0,007
exp_month_11	2,446	131	0,007
rap_debt_rec	2,43	132	0,008
rec_inbound_G_mv	2,417	133	0,008
ccivil	2,235	134	0,013
exp_month_9	2,204	135	0,014
exp_month_6	2,195	136	0,014
eg_else	2,16	137	0,015
distr_call_D	2,146	138	0,016
op_mv	2,016	139	0,022
rap_not_recouvr	1,87	140	0,031
prov6	1,849	141	0,032
rec_inbound_O_mv	1,766	142	0,039
rec_stand_contact	1,688	143	0,046
rec_inbound_G	1,631	144	0,051
og_9	1,504	145	0,066
rap_proc_open	1,502	146	0,067
distr_call_R	1,479	147	0,07
rap_outb	1,397	148	0,081
prov1	1,328	149	0,092
dist_display	1,321	150	0,093
rec_out_pos_mv	1,278	151	0,101
rec_out_neg_mv	1,117	152	0,132
rec_op	1,073	153	0,142
business	1,071	154	0,142
in_call_O	0,221	155	0,413
prov5	0,07	156	0,472
exp_month_10	-0,332	157	1
outbound_calls_pos	-0,547	158	1
exp_month_7	-0,672	159	1
nbr_son	-0,917	160	1
rec_out_neg	-0,929	161	1
prov7	-0,975	162	1
camp_e	-1,068	163	1
rec_inbound_P_mv	-1,132	164	1
rec_inbound_P	-1,443	165	1
outbound_calls_neg	-1,723	166	1
rec_son	-1,809	167	1
prov4	-1,971	168	1
camp_c	-2,017	169	1
son_mv	-2,23	170	1
ep_else	-2,991	171	1

Appendix C
Variable importance in commercial churn model

Variable Name	Z score	Importance ranking	p
canc_99P	52,275	1	0
nbr_canc	52,046	2	0
length_subs	32,944	3	0
lor	31,135	4	0
noc2	29,33	5	0
monetary	29,294	6	0
tel_fix	26,722	7	0
nov2	26,457	8	0
canc_78P	25,438	9	0
rec_og	23,967	10	0
og_3	22,054	11	0
cash_free	19,613	12	0
call_positive	18,406	13	0
no_renew	18,277	14	0
at_ttcmens	18,228	15	0
rap_rec	17,817	16	0
in_call_A	17,525	17	0
rap_amount_c	17,212	18	0
markov2	17,149	19	0
og_19	15,868	20	0
rec_inbound_A	15,696	21	0
info_mv	15,565	22	0
og_8	15,55	23	0
rec_negative	15,305	24	0
terme	15,2	25	0
canc_tot	14,509	26	0
rap_smooth	14,235	27	0
tel_mob	14,215	28	0
in_call_T	13,972	29	0
rap_amount	13,562	30	0
ind_subs	13,198	31	0
rec_positive	12,723	32	0
call_negative	12,349	33	0
domi_monthly	12,216	34	0
rec_inbound_F	12,103	35	0
rec_inbound_T	12,081	36	0
no_rappels	12,003	37	0
age	11,864	38	0
markov	11,25	39	0
tech	11,233	40	0
nov	11,218	41	0
canc_991	11,158	42	0
noc	11,129	43	0
in_call_C	11,071	44	0
rap_cut_adv	10,692	45	0
rec_ep	10,671	46	0
no_cycles	10,044	47	0
rec_inbound_C	9,829	48	0
rec_inbound_E	9,734	49	0
eg_2	9,709	50	0

Variable Name	Z score	Importance ranking	p
ep_2	9,571	51	0
after_call_X	9,403	52	0
after_call_F	9,173	53	0
in_call_E	8,925	54	0
ep_5	8,916	55	0
rec_inbound_R	8,903	56	0
lc	8,805	57	0
in_call_F	8,796	58	0
og_6	8,757	59	0
camp_p	8,718	60	0
age_mv	8,707	61	0
num_numabo	8,555	62	0
rec_out_neg_mv	8,24	63	0
ep_1	8,164	64	0
rec_inbound_A_mv	8,154	65	0
nbr_int	8,038	66	0
rec_inbound_D	8,021	67	0
in_call_D	7,748	68	0
rec_inbound_D_mv	7,469	69	0
ccivil	7,375	70	0
dist_display	7,182	71	0
rec_eg	7,093	72	0
og_11	7,046	73	0
rap_deactiv	7,035	74	0
product2	6,994	75	0
opt	6,615	76	0
camp_d	6,562	77	0
ep_mv	6,509	78	0
after_call_C	6,485	79	0
rec_out_neg	6,479	80	0
in_call_R	6,479	81	0
rec_inbound_F_mv	6,165	82	0
og_15	6,161	83	0
rap_mv	6,078	84	0
og_16	5,939	85	0
eg_mv	5,93	86	0
parrain	5,918	87	0
rec_neg_mv	5,904	88	0
outbound_calls_neg	5,834	89	0
stand_contact	5,804	90	0
distr_call_R	5,771	91	0
no_godchild	5,744	92	0
rec_inbound_R_mv	5,741	93	0
at_duree	5,708	94	0
og_7	5,588	95	0
product1	5,57	96	0
rec_inbound_T_mv	5,555	97	0
og_10	5,43	98	0
eg_6	5,161	99	0
camp_x	4,984	100	0
og_25	4,935	101	0
rec_inbound_C_mv	4,904	102	0
rec_pos_mv	4,568	103	0
og_13	4,553	104	0
rap_to_pay	4,53	105	0
markov_dummy	4,436	106	0
after_call_E	4,235	107	0
rec_inbound_G_mv	4,209	108	0
rec_inbound_E_mv	4,124	109	0
markov2_dummy	4,1	110	0

Variable Name	Z score	Importance ranking	p
distr_call_D	4,069	111	0
eg_else	4,023	112	0
exp_month_4	3,99	113	0
exp_month_3	3,932	114	0
rec_stand_contact	3,765	115	0
rec_out_pos	3,623	116	0
rap_not_recouvr	3,516	117	0
exp_month_5	3,412	118	0
rap_outb	3,317	119	0
exp_month_11	3,146	120	0,001
eg_8	3,126	121	0,001
nbr_products	2,987	122	0,001
exp_month_8	2,911	123	0,002
outbound_calls_pos	2,843	124	0,002
og_14	2,701	125	0,003
rec_stand_contact_mv	2,683	126	0,004
after_call_R	2,642	127	0,004
payeur	2,586	128	0,005
rec_inbound_G	2,543	129	0,005
prov1	2,442	130	0,007
rap_debt_rec	2,424	131	0,008
og_9	2,389	132	0,008
exp_month_6	2,322	133	0,01
rec_out_pos_mv	2,273	134	0,012
op_mv	2,007	135	0,022
og_22	1,78	136	0,038
exp_month_2	1,571	137	0,058
og_mv	1,51	138	0,066
rec_inbound_O_mv	1,47	139	0,071
exp_month_1	1,468	140	0,071
distr_call_C	1,431	141	0,076
rec_son	1,328	142	0,092
prov6	1,051	143	0,147
rap_proc_open	0,954	144	0,17
exp_month_9	0,903	145	0,183
rec_op	0,845	146	0,199
in_call_G	0,839	147	0,201
after_call_T	0,725	148	0,234
prov4	0,653	149	0,257
ep_else	0,599	150	0,275
son_mv	0,584	151	0,28
inbound_mail	0,504	152	0,307
distr_call_E	0,481	153	0,315
nbr_son	0,453	154	0,325
op	0,351	155	0,363
rec_inbound_P	0,312	156	0,377
prov5	0,101	157	0,46
in_call_P	-0,137	158	1
business	-0,339	159	1
camp_e	-0,399	160	1
og_21	-0,431	161	1
exp_month_7	-0,447	162	1
rec_inbound_mail	-0,483	163	1
in_call_O	-0,627	164	1
rec_inbound_mail_mv	-0,721	165	1
rec_inbound_P_mv	-0,824	166	1
og_else	-1,096	167	1
rec_inbound_O	-1,244	168	1
prov7	-1,809	169	1
exp_month_10	-2,044	170	1
camp_c	-2,989	171	1