



# Improved marketing decision making in a customer churn prediction context using generalized additive models

Kristof Coussement<sup>a,b,\*</sup>, Dries F. Benoit<sup>b</sup>, Dirk Van den Poel<sup>b</sup>

<sup>a</sup> University College HUBrussel, Faculty of Economics and Management, Stormstraat 2, B-1000 Brussels, Belgium

<sup>b</sup> Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweeherkenstraat 2, B-9000 Ghent, Belgium

## ARTICLE INFO

### Keywords:

Customer Relationship Management (CRM)  
Churn modeling  
Marketing decision making  
Generalized Additive Models (GAM)

## ABSTRACT

Nowadays, companies are investing in a well-considered CRM strategy. One of the cornerstones in CRM is customer churn prediction, where one tries to predict whether or not a customer will leave the company. This study focuses on how to better support marketing decision makers in identifying risky customers by using Generalized Additive Models (GAM). Compared to Logistic Regression, GAM relaxes the linearity constraint which allows for complex non-linear fits to the data. The contributions to the literature are three-fold: (i) it is shown that GAM is able to improve marketing decision making by better identifying risky customers; (ii) it is shown that GAM increases the interpretability of the churn model by visualizing the non-linear relationships with customer churn identifying a quasi-exponential, a U, an inverted U or a complex trend and (iii) marketing managers are able to significantly increase business value by applying GAM in this churn prediction context.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

Today, the business environment is characterized by fierce competition and saturated markets. In this context, companies increasingly derive revenue from the creation and enhancement of long-term relationships with their customers. This move towards a customer-centric approach to marketing, coupled with the increasing availability of customer-transactional data has made Customer Relationship Management (CRM) the leading strategy for marketing decision makers and this is reflected in firms' significant investments in CRM (Reinartz & Kumar, 2002; Teo, Devadoss, & Pan, 2006). Companies realize that their existing customer database is their most valuable asset (Athanasopoulos, 2000; Jones, Mothersbaugh, & Beatty, 2000; Thomas, 2001). It has been shown that it is more profitable to keep and satisfy existing customers than to constantly attract new customers who are characterized by a high attrition rate (Reinartz & Kumar, 2003). It is even suggested that it costs 12 times more to gain a new customer than to retain an existing one (Torkzadeh, Chang, & Hansen, 2006). Moreover, retained customers produce higher revenues and margin than new customers (Reichheld & Sasser, 1990).

It is clear that customer retention rates are important metrics in CRM (Hoekstra, Leeflang, & Wittink, 1999), resulting in an increasing number of customer churn related papers (e.g. Burez & Van den Poel, 2009; Coussement & Van den Poel, 2009). Predicting churn enables the elaboration of targeted retention strategies to limit the losses and to improve marketing decisions (Shaffer & Zhang, 2002). For example, specific incentives may be offered to the most risky customer segments, i.e. the most inclined to leave the company, with the hope that they remain loyal (Burez & Van den Poel, 2007). Two possible outcomes are observed: the customer churned or he/she stayed with the company. So, technically spoken, customer churn modeling is a binary classification problem. Moreover, for the prediction of customer churn using the information stored in the data warehouse one needs tools that can handle large amounts of data, i.e. data-mining tools (Shaw, Subramaniam, Tan, & Welge, 2001).

A broad range of data mining techniques have been used in the past, but one of the most popular models in the churn context remains the (binary) Logit model (Lemmens & Croux, 2006). This model has been used extensively in marketing decision making to model churn problems (e.g. Buckinx & Van den Poel, 2005; Hwang & Euiho Suh, 2004; Kim & Yoon, 2004). The popularity of Logistic Regression is not surprising since the model has some outspoken advantages. First of all, the method is able to combine relative simplicity and good performance. The parameter estimates of the Logit model are interpretable in terms of odds ratios which facilitates full understanding of the results. Moreover, the technique is relatively robust (Buckinx & Van den Poel, 2005), while

\* Corresponding author. Address: University College HUBrussel, Faculty of Economics and Management, Stormstraat 2, B-1000 Brussels, Belgium. Tel.: +32 (0)2 608 14 19.

E-mail addresses: [Kristof.Coussement@HUBrussel.be](mailto:Kristof.Coussement@HUBrussel.be) (K. Coussement), [Dries.Benoit@UGent.be](mailto:Dries.Benoit@UGent.be) (D.F. Benoit), [Dirk.VandenPoel@UGent.be](mailto:Dirk.VandenPoel@UGent.be) (D. Van den Poel).

its popularity is reflected in the availability in almost all statistical software packages. However, as often, one of the advantages can also be considered as a disadvantage. The simplicity of the model is achieved by assuming that the functional form of the dependence on the explanatory variables is known. However, the functional form is seldom known in practice. If the functional form is misspecified then the estimates of the coefficients and the inferences based on them are misleading (Horowitz & Savin, 2001). In addition, with Logistic Regression, a linear relationship is modeled on the data, which is an oversimplification of the real relationship in the data (Allison, 1999). It is possible to relax these restrictive assumptions by applying the Generalized Additive Models (GAM) approach.

The GAM approach has at least two distinct advantages compared to Logistic Regression. First, by relaxing the usual assumptions it becomes possible to uncover (non-linear) structures in the data that otherwise might be missed. These structures often give substantial new insights into the effects of the covariates. Furthermore, since the GAM approach allows for more complex relationships between the dependent variable and the covariates, more accurate predictions are likely (Hastie & Tibshirani, 1986). The GAM approach has shown its relevance in a broad range of different research domains ranging from biology (e.g. Jowett, Parkyn, & Richardson, 2008) over cancer research (e.g. Lee et al., 2007) and supply-chain management (e.g. Cakir, 2009). However, only very few marketing-related studies have used a related non- or semi-parametric methodology (e.g. Kalyanam, 1998; Kumar, Scheer, & Steenkamp, 1998; Shively, 2000). Moreover, to the best of our knowledge this is the first study in a customer churn context that evaluates and uses GAM models.

This study contributes to the existing literature in several ways. Three distinct perspectives on GAM models in the current churn context are elaborated. First, a comparison in terms of predictive performance is made between the popular Logistic Regression model and the GAM approach. Second, it is shown that relaxing the linearity assumption results in a much richer insight into the effects of the covariates which helps decision makers in a thorough understanding of the churn problem of their company. Finally, it is shown how GAM models increase business value of the company. The added value of the GAM approach is therefore shown in monetary terms based on a real-life business setting.

The remainder of this paper is organized as follows. Section 2 focuses on the methodology behind both the logistic model and the GAM approach. The evaluation criteria used in the empirical part of this study are explained in the same section. Section 3 applies the proposed methodology to a real-world churn setting, i.e. the newspaper case. Section 4 treats the managerial implication of the proposed methodology and finally, Section 5 gives conclusions and directions for further research.

## 2. Methodology

### 2.1. Base learner

The methodological link between the Logistic Regression and the GAM approach is explained starting from the Generalized Linear Model (GLM) framework. The most common representation of the GLM is as follows (Tabachnick & Fidell, 1996):

$$E(Y) = \mu = g(X'\beta) \quad (1)$$

with  $E(Y)$  representing the expected value of  $Y$ ,  $X'\beta$  is a linear combination of the data with unknown parameters  $\beta$  and  $g$  is the link function. As Eq. (1) shows, each outcome of the dependent variable  $Y$  is assumed to be generated from a particular distribution function. Different choices of distribution function and link function

result in different statistical techniques. For example, the normal distribution in combination with the identity link function results in the normal multiple regression model. When the responses  $Y$  are binary, the distribution function is the binomial distribution and the interpretation of  $\mu_i$  is then the probability,  $p_i$ , of  $Y_i$  taking on the value one. A Logistic Regression model is then obtained by applying the logit link function by

$$\text{logit}\{P(X)\} \equiv \log \left\{ \frac{P(X)}{1 - P(X)} \right\} = \alpha + \sum_{j=1}^p \beta_j' X_j \quad (2)$$

with  $P(X) = \Pr(Y = 1|X)$ . The parameter vector  $\beta$  is estimated from observations  $(y, x)$  and this can be done using maximum likelihood among others. More technical details on Logistic Regression are found in Anderson (1983).

### 2.2. Generalized additive models

An attractive alternative to Logistic Regression is Generalized Additive Models (GAM) (Hastie & Tibshirani, 1986; Hastie & Tibshirani, 1987; Hastie & Tibshirani, 1990). Additive Logistic Regression relaxes the linearity constraint and suggests a nonparametric fit to the data. In other words, the regression function is modeled in a nonparametric way and the data itself decides on the functional form. However, Hastie and Tibshirani (1990) argue that nonparametric methods perform worse when the number of explicative variables increases, because the sparseness of the data inflates the variance of the estimates. This is often referred to as the curse of dimensionality (Bellman, 1961). It was Stone (1985) who proposed additive models to approximate the multivariate regression function. Consequently, the curse of dimensionality is avoided since each individual additive term is estimated using a univariate smoother. Hastie and Tibshirani (1990) pick up the additive model principle by extending the GLM framework.

This study gives a general overview on the GAM principle. For more details, we kindly refer to Hastie and Tibshirani (1986), Hastie and Tibshirani (1987, Hastie and Tibshirani (1990) or Hastie, Tibshirani, and Friedman (2001). Suppose that  $Y$  is the response variable with binary target labels and  $X_1, X_2, \dots, X_p$  is a set of independent variables, GAM generalize the Logistic Regression principle by replacing the linear predictor  $\sum_{j=1}^p \beta_j' X_j$  in Eq. (2) with an other additive component where

$$\text{logit}P(X) \equiv \log \left\{ \frac{P(X)}{1 - P(X)} \right\} = \alpha + \sum_{j=1}^p s_j(X_j) \quad (3)$$

with  $s_1(\cdot), s_2(\cdot), \dots, s_p(\cdot)$  as smooth functions. This study uses smoothing splines for  $s_1(\cdot), s_2(\cdot), \dots, s_p(\cdot)$  to estimate the nonparametric trend for the dependence of the logit on  $X_1, X_2, \dots, X_p$ . A smoothing spline solves the following optimization problem: among all functions  $\eta(x)$  with continuous second order derivatives, find the function that minimizes the penalized residual sum of squares

$$\sum_{i=1}^n (y_i - \eta(x_i))^2 + \lambda \int_a^b (\eta''(t))^2 dt \quad (4)$$

where  $\lambda$  is a fixed constant and  $a \leq x_1 \leq x_2 \leq \dots \leq x_n \leq b$ . The goodness-of-fit is measured by the first part of Eq. (4), while the second term is a penalty term that penalizes curvature in the function by the smoothing parameter  $\lambda$ . The complexity of  $\eta(x)$  is measured by  $\lambda$  and it is inversely related to the degrees of freedom ( $df$ ). If  $\lambda$  is small (i.e. the  $df$  are large),  $\eta(x)$  is any function that approaches an interpolation to the data, when  $\lambda$  is large (i.e. the  $df$  are small),  $\eta(x)$  is closely related to a simple least squares fit. It is shown that an explicit and unique minimizer for Eq. (4) exists, i.e. a natural

cubic spline with knots at the unique values of  $x_i$  (Hastie & Tibshirani, 1990).

In order to optimize the GAM, the local scoring algorithm (Hastie & Tibshirani, 1986) is applied, while the backfitting algorithm of (Friedman & Stuetzle, 1981) is used to estimate the additive models. The current study follows the recommendation of Hastie et al. (2001) to use a  $\lambda$  corresponding to four degrees of freedom, because it is shown that a small number of  $df$  fits most datasets very well.

### 2.3. Feature selection

Feature selection techniques are often used in the context of churn modeling (e.g. Coussement & Van den Poel, 2008). It is the process of choosing a subset of the original predictive variables by eliminating variables that are either redundant or possess little predictive information. This significantly improves the comprehensibility of resulting models and makes resulting models generalize better (Kim, 2006). In addition, selection techniques defy the curse of dimensionality in order to improve prediction performance (Guyon & Elisseeff, 2003). Moreover, extracting a comprehensive subset of relevant predictor variables is critical for decision makers who want to build marketing strategies based on key drivers of customer response. A forward selection is used within this research setting. The choice for this relatively standard selection technique makes it easy to replicate results with most statistical software packages (Kim, 2006).

### 2.4. Evaluation criteria

It is essential to evaluate the classifiers in terms of performance. A brief review of evaluation metrics for marketing is found in (Rosset, Neumann, Eick, Vatnik, & Idan, 2001). Two important and often used performance evaluation criteria in the context of churn prediction are the Area Under the receiver operating characteristic Curve (AUC) and the top-decile lift (e.g. Buckinx & Van den Poel, 2005; Coussement & Van den Poel, 2008; Lemmens & Croux, 2006).

The first measure used in this study is the AUC. In contrast to some other evaluation measurements (e.g. accuracy), the AUC is not influenced by any threshold value since it takes into account all possible thresholds on the predicted probabilities. For all these points, it considers the sensitivity (the number of true positives versus the total number of defectors) and one minus the specificity (the number of true negatives versus the total number of non-defectors) of the confusion matrix in a two-dimensional graph (Egan, 1975). Plotting the outcomes of these calculations, results in a receiver operating characteristics curve. The area under this curve (AUC) lies between 0.5 and 1 and is then used to evaluate the predictive performance of classification models. Again, the larger the AUC, the better is the performance of the model. Comparing and evaluating the predictive performance of two or more models using AUC is conducted using the non-parametric test proposed by DeLong, DeLong, and Clarke-Pearson (1988).

The second performance measure used in this study is the top-decile lift. The top-decile lift focuses exclusively on the most critical group of customers and their churn risk. The top 10% riskiest customers (i.e. the group of customers with the highest predicted churn probabilities) represents an ideal segment for targeting in a retention-marketing campaign (Lemmens & Croux, 2006). This performance measure is very appealing because it incorporates somewhat the notion that marketing budgets are limited and actions to reduce churn are therefore limited to a segment of customers that is at high risk. In practice one will order the customers on decreasing predicted churn probability. Next, the proportion of real churners in the top 10% is compared with the proportion of churners in the total dataset. It is clear that the higher the top-decile lift,

the better is the model. For example, a top-decile lift of 2 means that the model under investigation identifies twice as many churners in the top 10% as a random assignment would do.

It is important to see that the AUC and the top-decile lift are measuring different aspects of the predictive accuracy of the models. Both evaluation criteria provide complementary information. A model can be good at identifying the most risky segment but less effective at recognizing less risky customers. Combining the two metrics provides a thorough evaluation of the performance of the models under investigation.

## 3. Empirical validation

This Section applies the proposed methodology to a real-world churn setting. In a first subsection, an overview of the research setting is given. Section 3.2 defines the best subset of marketing variables by means of a Logistic Regression and a Generalized Additive Model using a forward selection procedure, while both models are compared in terms of the predictive performance. Finally, Section 3.3 clarifies the interpretability of both prediction models.

### 3.1. Data

The current study is performed on data supplied by the largest Belgian newspaper publishing company. Churn prevention is an important issue for the newspaper business (Lemmens & Croux, 2006), which makes this setting an ideal testbed for validating the proposed methodology. Customers subscribe during a certain period for a specific newspaper. This customer churn prediction context comes down to predicting whether or not a subscriber will renew his/her subscription during the 30-day period after the expiry date. The company still delivers the newspaper to the subscribers during this 30-day period in order to allow some flexibility in the renewal procedure. The churn response variable is coded as '1' when the subscription is not renewed within a 30-day period after the renewal date, and '0' otherwise. The explanatory variables contain information covering a 30-month period returning from every individual renewal point. The predictors include client/company interaction variables (e.g. the number of responses on direct marketing actions...), renewal-related variables (e.g. the number of times the customer churned in the past...), subscriber-describing variables (e.g. age...) and subscription-describing variables (e.g. the length of the current subscription...). Appendix A gives a detailed overview of the independent variables used within this research setting.

This database contains all renewals from the first of March 2005 till the end of June 2005. In order to make a correct prediction on whether or not a customer will leave the company, the total dataset is divided into a training and test set. The training set is used for model building, while the test set allows for valid assessment of performance (Lemmens & Croux, 2006) and as such, one is sure that the trained model generalizes well. Practically, we randomly assign 70% of the data to the training set in order to build the prediction model, while the other 30% of the data is used to validate the proposed models to unseen data. The training set contains 93,859 subscriptions with a churn incidence of 11.95%, while the test set contains 40,225 subscriptions of which an equal proportion of 11.95% are churners. More information on the data characteristics is found in Table 1.

In order to answer the research question whether or not it is beneficial to allow more flexibility in the prediction model by using nonparametric trends, two churn-prediction models are built: (i) MOD\_LOG containing the parametric terms of a forward Logistic Regression and (ii) MOD\_GAM which models customer churn via a forward GAM model which allows for nonlinear relationships.

**Table 1**  
Overview of the data characteristics.

	Number of subscriptions	Relative percentage
<i>Training set</i>		
Subscriptions not renewed	11,218	11.95
Subscriptions renewed	82,677	88.05
Total	93,895	100
<i>Test set</i>		
Subscriptions not renewed	4808	11.95
Subscriptions renewed	35,417	88.05
Total	40,225	100

### 3.2. Predictive performance

By employing MOD\_LOG and MOD\_GAM, one is able to obtain the best subset of predictors differentiating between churners and non-churners. The variable selection procedure results in a subset of 18 predictors for MOD\_LOG and 27 variables for MOD\_GAM, both at a significance level of  $p < 0.05$  (see Table 2).

Table 2 clearly indicates that MOD\_GAM selects more significant variables describing the churn behavior than MOD\_LOG due to added flexibility in the prediction model.

**Table 2**  
Best subset of predictors.

	Step	Variable name	Variable description
MOD_LOG	1	mon_val	Monetary value
	2	rec_complaint	Elapsed time since the last complaint
	3	avg_complaint_pos	the average positioning of the complaints in the current subscription
	4	nbr_renew_point	The number of renewal points
	5	term	Elapsed time since last renewal point
	6	age	Age of the subscriber
	7	days_renew_after	How many days after the expiry date, the previous subscriptions were renewed
	8	nbr_suspension	The number of suspensions
	9	nbr_change_edition	The number of conversions made in edition
	10	rec_change_pay	Elapsed time since last conversion in payment method
	11	lor	The length of relationship
	12	length_current_subsc	The length of the current subscription
	13	avg_renewaltime	The average of the subscription renewal time
	14	nbr_complaints	The number of complaints
	15	avg_suspension time	Average suspension time
	16	nbr_prior_churn	The number of prior churning events
	17	nbr_dm_action	The number of responses on direct marketing actions
	18	timelag_response	Elapsed time since response on direct mailing campaign
MOD_GAM	1	mon_val	Monetary value
	2	rec_complaint	Elapsed time since the last complaint
	3	rec_change_pay	Elapsed time since last conversion in payment method
	4	days_renew_before	How many days before the expiry date, the previous subscriptions were renewed
	5	term	Elapsed time since last renewal point
	6	nbr_renew_point	The number of renewal points
	7	rec_suspension	Elapsed time since last suspension
	8	timelag_response	Elapsed time since response on direct mailing campaign
	9	days_renew_after	How many days after the expiry date, the previous subscriptions were renewed
	10	age	Age of the subscriber
	11	avg_complaint_pos	the average positioning of the complaints in the current subscription
	12	nbr_change_pay	The number of conversions made in payment method
	13	avg_suspension time	Average suspension time
	14	lor	The length of relationship
	15	rec_dm_action	Elapsed time since last response on direct marketing action
	16	nbr_complaints	The number of complaints
	17	nbr_suspension	The number of suspensions
	18	length_current_subsc	The length of the current subscription
	19	nbr_change_channel	The number of conversions made in distribution channel
	20	rec_change_channel	Elapsed time since last conversion in distribution channel
	21	avg_renewaltime	The average of the subscription renewal time
	22	rec_change_edition	Elapsed time since last conversion in edition
	23	nbr_dm_action	The number of responses on direct marketing actions
	24	nbr_free copy	The number of free newspapers
	25	var_renewaltime	The variance on the subscription renewal time
	26	nbr_compensation	The number of compensation newsletters
	27	nbr_change_edition	The number of conversions made in edition

This finding results in a better predictive performance of MOD\_GAM compared to MOD\_LOG. Table 3 shows the predictive performance of MOD\_LOG and MOD\_GAM. It is clear that MOD\_GAM outperforms MOD\_LOG in terms of AUC. The AUC increases from 0.8298 to 0.8452 on the test set. This improvement is significant ( $\chi^2 = 187.40$ ,  $df = 1$ ,  $p < 0.001$ ) (Delong et al., 1988). Insightfully, MOD\_GAM is able to better distinguish churners from non-churners compared to MOD\_LOG.

Moreover, the beneficial effect of using an Additive Logistic Regression contrary to a traditional Logistic Regression is confirmed when having a look at the top-decile lift. MOD\_GAM is able to increase the lift in the first decile from 4.08 to 4.26. In other words, MOD\_LOG is able to increase the base line churn rate of 11.95% in the first decile to 48.75%, while an additional increase is obtained by MOD\_GAM to 50.91% churners.

These results suggest that applying an Additive Logistic Model instead of a traditional Logistic Regression is a valuable strategy to improve the prediction power of a customer churn model.

### 3.3. Interpretability of the proposed models

This subsection explores the impact of the best subset variables on customer attrition. Table 4 shows the univariate parameter



**Table 3**

Test set performance in terms of AUC and top-decile lift.

	AUC	Top-decile lift
MOD_LOG	0.8298 <sup>a</sup>	4.08
MOD_GAM	0.8452 <sup>a</sup>	4.26

<sup>a</sup> Significant with  $\chi^2 = 187.40$ ,  $df = 1$ ,  $p < 0.001$ .**Table 4**

Overview of the univariate standardized parameter estimates and odds ratios.

Variable name	Standardized parameter estimates <sup>a</sup>	Odds ratios
mon_val	−0.72	0.994
rec_complaint	0.26	1.046
avg_complaint_pos	−0.29	0.100
nbr_renew_point	−1.90	0.642
term	−0.03	0.988
age	−0.19	0.976
days_renew_after	0.05	1.010
nbr_suspension	−0.14	0.819
nbr_change_edition	−0.26	0.314
rec_change_pay	−0.09	0.960
lor	−0.75	0.986
length_current_subsc	−0.87	0.984
avg_renewaltime	−0.18	0.285
nbr_complaints	−0.42	0.815
avg_suspension time	−0.08	0.632
nbr_prior_churn	0.07	1.381
nbr_dm_action	0.11	1.229
timelag_response	0.12	1.951

<sup>a</sup> Significant with  $p < 0.05$ .

estimates from a traditional Logistic Regression model. As such, the marketing analyst is able to measure the true impact of a single variable on the churn behavior.

The relationship of the variables in Table 4 is linear to the logit of churning. However, this assumption can be relaxed by taking into account nonlinear effects between the independent and the logit. This is done by applying the GAM approach. Appendix B visualizes the effect of every single variable in the analysis on the churn probability. On all sub panels, the horizontal axis represents the values of the independent variable, while the vertical axis represents the fitted churn probabilities. The orange line represents the non-monotonic relationship between the independent and the churn probability as estimated using a GAM. Appendix B clearly shows that more flexible solutions are obtained when one relaxes the linearity constraint of the Logistic Regression. It gives the marketing analyst a much better insight into the way a certain independent variable contributes to the chance that a customer will leave the company. The non-parametric trends are visually

summarized by four main categories: a quasi-exponential, a U, an inverted U and a complex trend.

A first category of variables follows a quasi-exponential trend. Hereby the effect on the churn probability is very pronounced for small values of the independent variable, while it decreases rapidly for middle to higher values. For example, in Appendix B, the variable 'length\_current\_subsc' clearly shows a strong effect for small values, while for values larger than 100, the effect is minimal and starts to stabilize. A much richer insight into the effect of 'length\_current\_subsc' on the churn probability is obtained compared to what is learned from the Logistic Regression coefficients (see Table 4), i.e. increasing values of the explicative variable leads to decreasing churn probabilities.

For variables following the U-trend, the churn probability decreases till it reaches a minimum. When the values of the independent variable become larger, the chance of not renewing the subscription gets bigger. The 'nbr\_suspension' exerts such an effect. Extreme values for this variable lead to higher churn probabilities, while more central values result in a lower churn probability. Interpretation based only on the Logistic Regression coefficient (see Table 4) overlook the insight of low values for 'nbr\_suspension' leading to higher churn probabilities.

The inverse happens when one has an inverted U trend. The churn probability curve is striving to a maximum, but when the independent variable gets even bigger, the churn probability becomes smaller. This is the case for the variable 'term'. Looking only at the linear effect from the Logistic Regression model (see Table 4), one concludes that higher values for this variable result in lower churn probabilities. Again, the added value of the GAM approach is shown in Appendix B. The graph shows a much richer insight into the effect of the covariate. From this graph we can learn that both high and low values lead to smaller churn probabilities, while more central values push the churn probability up.

The last category of variables is characterized by a complex trend. The effects on the churn probability are Gordian, what makes interpretation burdensome. One can say that the variables in this category enhance the predictive power of the model more than its interpretability. Moreover comparing Table 2 and Appendix B, this type of variables is often not selected in the MOD\_LOG model.

In the end, it should be clear that relaxing the linearity constraint by using GAMs gives data analysts a more profound insight into the explicative variables.

#### 4. Managerial implications

As the difference in predictive performance between MOD\_LOG and MOD\_GAM is significant, the impact on the company's marginal profitability will not fail to appear. Indeed, several studies

**Table 5**

Managerial implications.

RANDOM			MOD_LOG				MOD_GAM				
Top-decile		Discounted profit	Top-decile		Discounted profit	Marginal Profit	Top-decile		Discounted profit	Marginal profit	
Lift	#churners	RANDOM	Lift	#churners	MOD_LOG	MOD_LOG-RANDOM	Lift	#churners	MOD_GAM	MOD_GAM-RANDOM	MOD_GAM-MOD_LOG
1	1000	19,230.77	4.08	4080	78,461.54	59,230.77	4.26	4260	81,923.08	62,692.31	3461.54
1	2000	38,461.54	4.08	8160	156,923.08	118,461.54	4.26	8520	163,846.15	125,384.60	6923.07
1	3000	57,692.31	4.08	12,240	235,384.62	177,692.31	4.26	12,780	245,769.23	188,076.90	10,384.61
1	4000	76,923.08	4.08	16,320	313,846.15	236,923.07	4.26	17,040	327,692.31	250,769.20	13,846.16
1	5000	96,153.85	4.08	20,400	392,307.69	296,153.84	4.26	21,300	409,615.38	313,461.50	17,307.69
<b>1</b>	<b>4800</b>	<b>92,307.69</b>	<b>4.08</b>	<b>19,584</b>	<b>376,538.46</b>	<b>284,230.77</b>	<b>4.26</b>	<b>20,448</b>	<b>393,076.92</b>	<b>300,769.20</b>	<b>16,538.46</b>

Discounted and marginal profit in Euro.

showed that minor increases in customer attrition rates result in significant changes in contribution (e.g. Reichheld & Sasser, 1990; Van den Poel & Larivière, 2004). Table 5 gives an overview of the increase in profitability between the base model (i.e. random customer selection or RANDOM), MOD\_LOG and MOD\_GAM performance. Table 5 focuses on the top-decile performance to calculate the additional profits, because marketing analysts are often interested in just 10% of the customer base who are most likely to churn (Coussement & Van den Poel, 2008). The reason is twofold: marketing budgets for customer retention campaigns are often limited and actions to reduce churn typically involve only 10% of the entire list of customers in order to maintain a feasible operational solution.

Table 5 contains the profit implications for the random model, MOD\_LOG and MOD\_GAM. RANDOM has always a top-decile lift of 1, because theoretically selecting  $X$  random customers from the total customer base in the current case results in the selection of 11.95% churners, i.e. the base churn rate. A sensitivity analysis on the profitability is done based on the number of churners in the top-decile of RANDOM. The discounted profit is defined as today's value of the future one-year profit gained by the retained churners from the top-decile. The discounted profit for model  $X$  is then calculated by

$$\text{discounted profit}|X = \frac{\varphi * P}{1 + d} \quad (5)$$

with  $P$  equal to the average yearly profit per subscription,  $d$  equal to the discount rate and  $\varphi$  equal to the number of retained churners.  $\varphi$  is calculated by

$$\varphi = (\kappa * \tau) * \eta \quad (6)$$

with  $\kappa$  equal to the number of churners in the top-decile from RANDOM,  $\tau$  equal to the top-decile lift of model  $X$  and  $\eta$  equal to the conversion rate of the churners.

Within this research paper, the marginal (discounted) profit of model  $X$  over model  $Y$  is defined as the difference between the discounted profit of model  $X$  and the discounted profit of model  $Y$  or

$$\text{marginal profit}|X - Y = \text{discounted profit}|X - \text{discounted profit}|Y \quad (7)$$

For this real-life example, the company's average yearly profit per subscription  $P$  is close to 200 Euro, the discount rate  $d$  is set equal to 4%, the number of churners in the top-decile from RANDOM  $\kappa$  ranges arbitrarily from 1000 to 5000 by steps of 1000 or  $\kappa = \{1000; 2000; 3000; 4000; 5000\}$ , the churn conversion rate  $\eta$  is set to a realistic 10% meaning that one out of ten churners is retained by the company's retention program and the top-decile lift  $\tau$  is obtained from the results of MOD\_LOG and MOD\_GAM, while  $\tau$  is 1 for RANDOM meaning that the relative number of churners in the top-decile equals the base churn rate of 11.95%.

Table 5 indicates that churn prediction modeling is highly beneficial. The marginal profits of MOD\_LOG and MOD\_GAM over RANDOM are high. For instance, per 1000 churners in the top-decile of the base model, the company is able to make an additional profit of 59,230.77 Euro when predicting churn based on Logistic regression (MOD\_LOG) and 62,692.31 Euro when a Additive Logistic Model (MOD\_GAM) is used. Predictive models are able to better distinguish churners from non-churners. In other words, the density of churners in the top-decile is drastically increased which results in additional profits. Moreover, the impact of distinguishing churners from non-churner using MOD\_GAM

instead of MOD\_LOG is highly significant. The increase in top-decile from 4.08 to 4.26 results in an additional increase of 3,461.54 Euro per 1000 churners in the top-decile. Extending these results on a yearly basis to this real-life customer churn context has the following impact. The company has approximately 16,000 churners during a 4-month period (see Table 1), which results in about 48,000 subscriptions that are not renewed on a yearly basis. The real-world impact on profits for RANDOM, MOD\_LOG and MOD\_GAM for this setting is shown in bold and italic in Table 5. For instance, it is shown that the company is able to save a substantial amount of 16,538.46 Euro on a yearly basis by relaxing the linearity constraint in the customer churn prediction model. In summary choosing MOD\_GAM instead of MOD\_LOG for distinguishing churners from non-churners is a valuable strategy to improve the prediction power of a customer churn prediction model resulting in an additional increase in company's profit.

## 5. Conclusion and directions for further research

Nowadays, a lot of marketing analysts focus on predicting whether or not a customer will leave the company. Indeed, customer churn prediction is one of the facets in a well-considered CRM strategy within a lot of companies. Using the customer-transactional database as a starting point for their churn analyses, marketing analysts are using data mining techniques to set up a churn prediction system.

This study contributes to the current churn literature by applying Generalized Additive Models. GAM relaxes the linearity constraint in contrast to a traditional Logistic Regression model. As such, it is possible to allow for complex non-linear relationships between the explanatory variables and the churn probability. First, it is shown that one is able to significantly increase the prediction performance of the churn model by applying GAM. Allowing for more complex relationships between the covariates and the logit, results in better identification of the customers at risk. Second, one gains better insights into the effects of the covariates which are modeled using a non-parametric relationship. Visually, one is able to distinguish four different patterns between the covariates and the logit: a quasi-exponential, a U, an inverted U or a complex trend. Finally, from a managerial point of view it is highly beneficial in monetary terms to use GAM for churn prediction in this research context.

While we strongly believe that this research paper fills a large gap in today's literature, there are still some directions for future research. This study introduces the use of the GAM approach in a churn prediction context. In order to externally validate the obtained results, one can apply GAM in other than customer churn contexts, e.g. response modeling, cross- and up-sell applications... or by exploiting this principle in other than non-binary classification domains of CRM like for instance customer lifetime value modeling. Moreover, this study compares Logistic Regression with a relaxed version of it, namely Additive Logistic Regression which allows for nonlinear fits to the independent variables. Future research could be done in comparing other nonlinear prediction techniques like decision trees, neural networks, random forests...

## Acknowledgments

We would like to thank the anonymous Belgian Publishing Company for making available their data. Next, we also like to thank Ilse Bellinck and Weijie Cai for their fruitful comments on this research paper.

## Appendix A. Overview of the explicative variables

### Client/company-interaction variables

- The number of complaints,
- Elapsed time since the last complaint,
- The average cost of a complaint (in terms of compensation newspapers),
- The average positioning of the complaints in the current subscription,
- The conversions made in distribution channel, payment method & edition,
- Elapsed time since last conversion in distribution channel, payment method & edition,
- The number of responses on direct marketing actions,
- The number of suspensions,
- The average suspension length (in number of days),
- Elapsed time since last suspension,
- Elapsed time since last response on a direct marketing action,
- The number of free newspapers.

### Renewal-related variable

- How many days before the expiry date, the previous subscription was renewed,
- The average number of days the previous subscriptions are renewed before expiry date,
- The variance in the number of days the previous subscriptions are renewed before expiry date,
- Elapsed time since last step in renewal procedure,
- The number of times the churning did not renew a subscription.

### Subscriber-describing variables

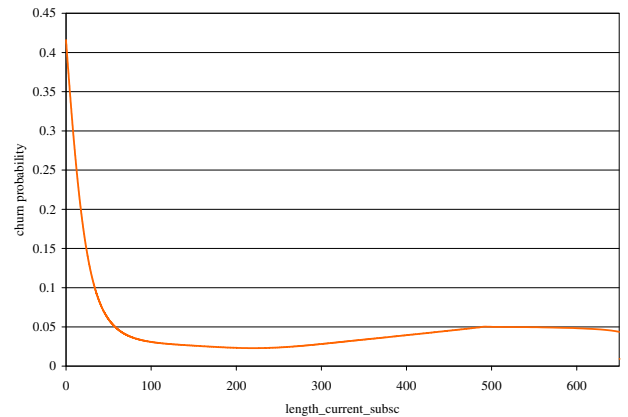
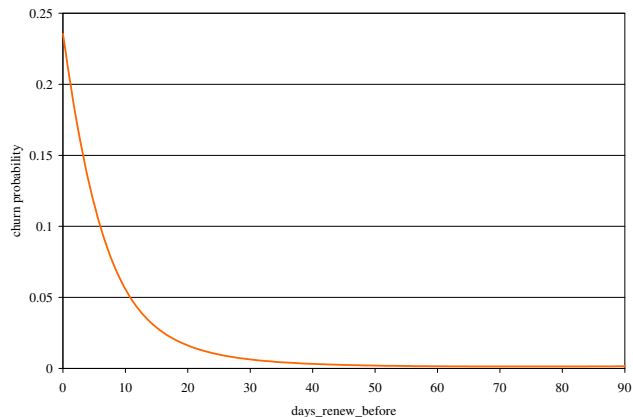
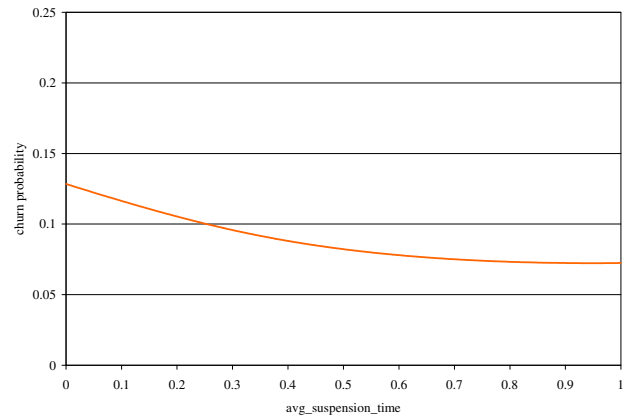
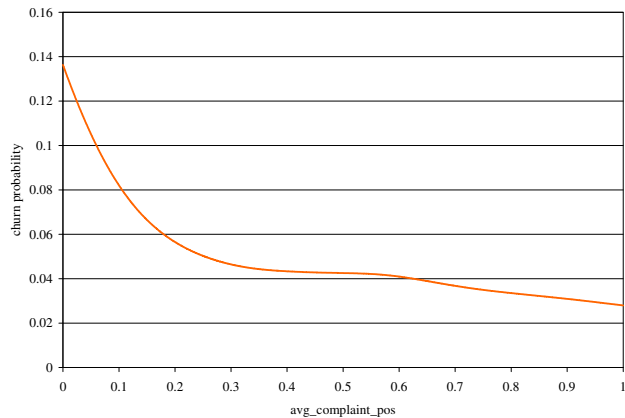
- Age,
- The length of relationship.

### Subscription-describing variables

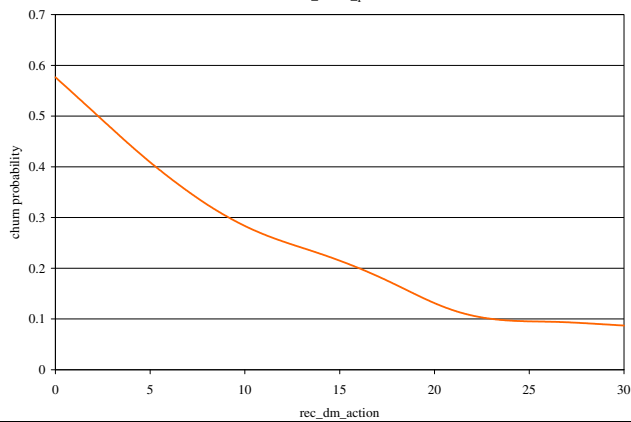
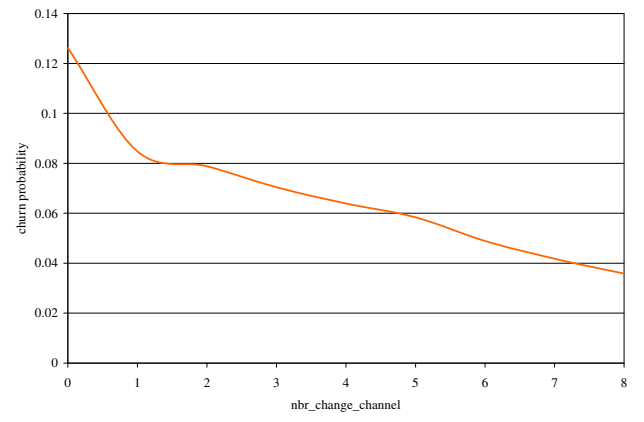
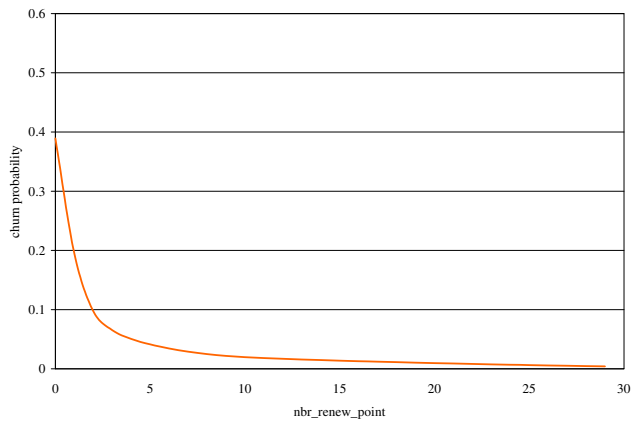
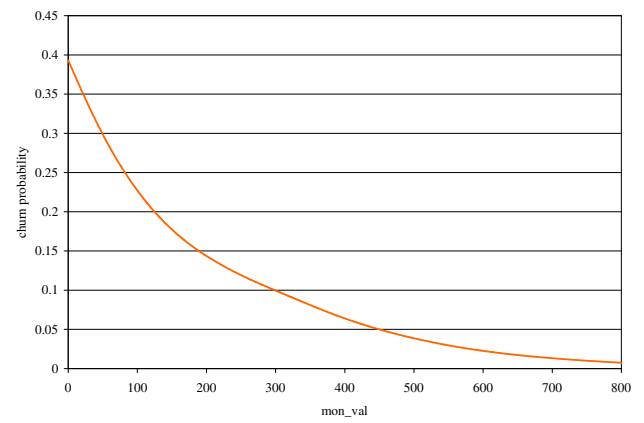
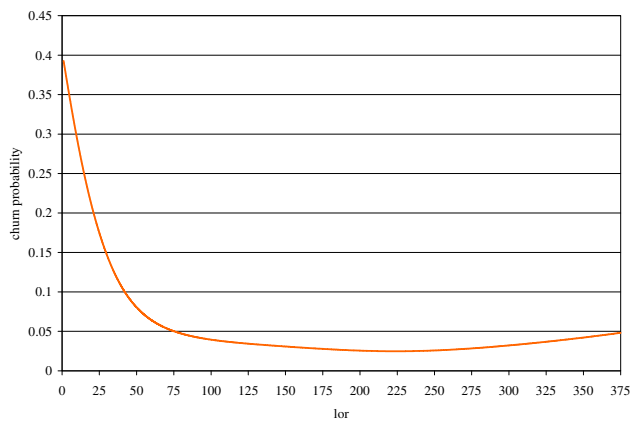
- Elapsed time since last renewal,
- Monetary value,
- The number of renewal points,
- The length of the current subscription,
- The number of days a week the newspaper is delivered (intensity indication).

## Appendix B. Visualization of MOD\_GAM

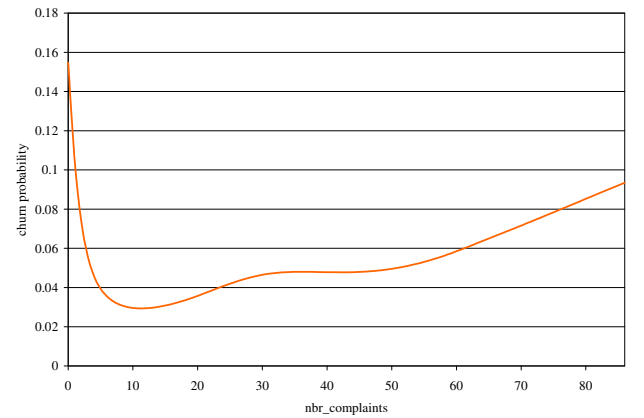
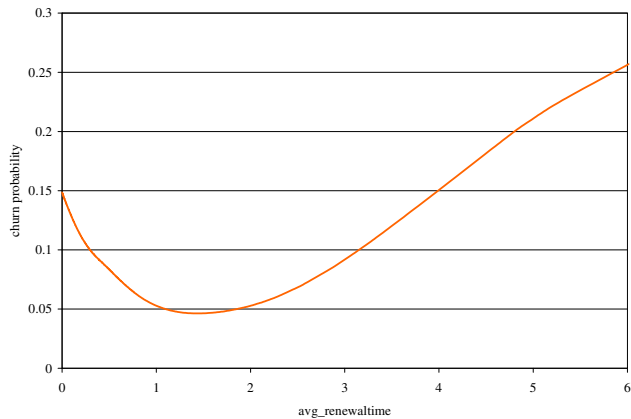
### Quasi exponential trend



## Appendix B (continued)



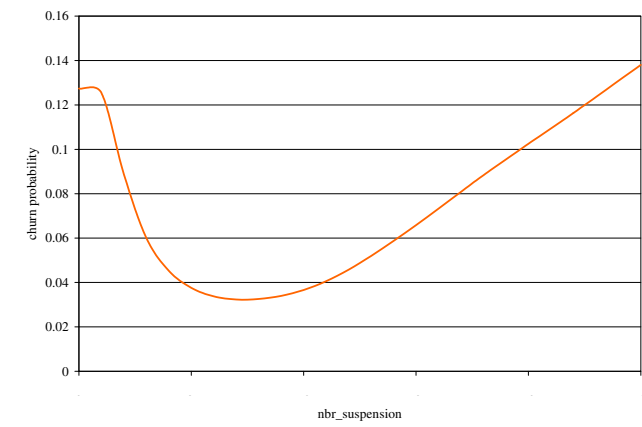
## U-trend



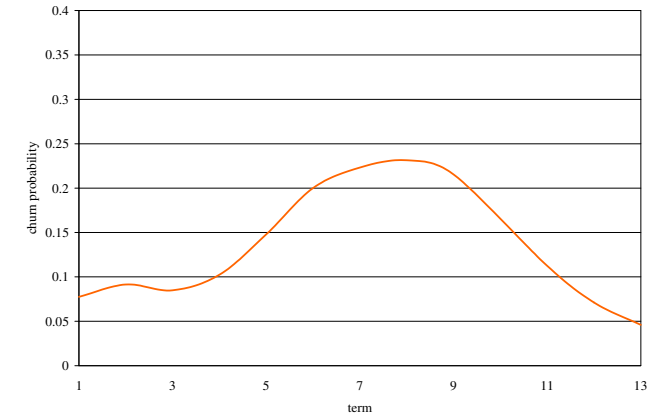
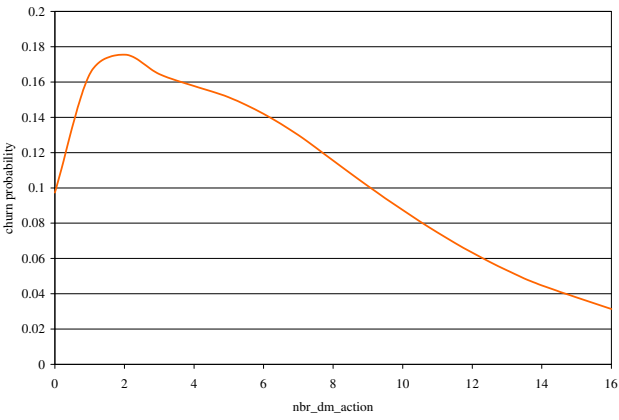
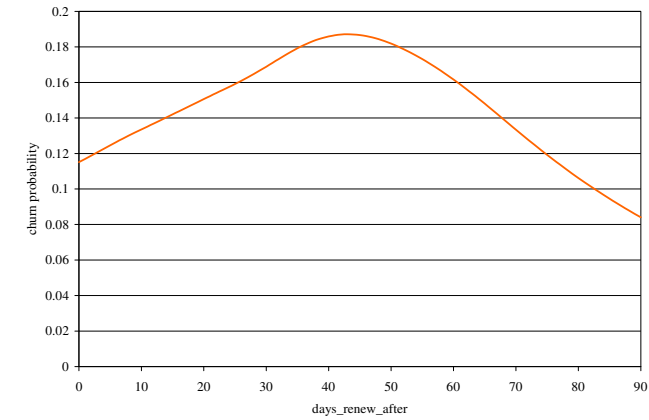
(continued on next page)



Appendix B (continued)

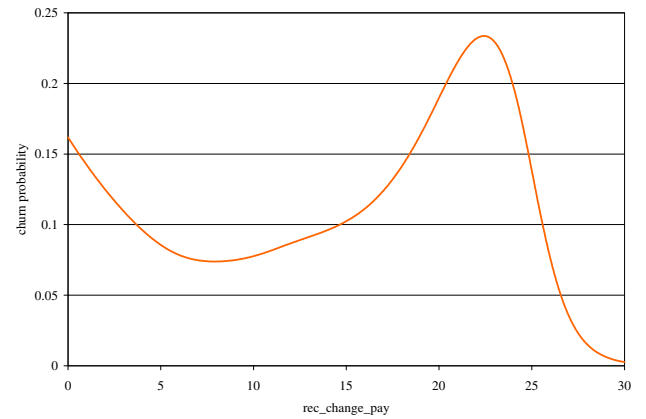
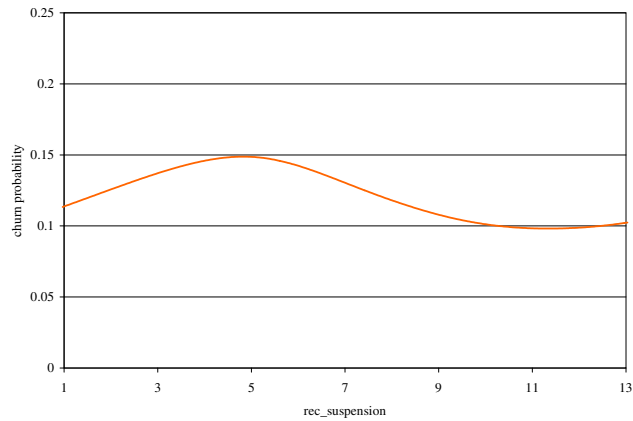
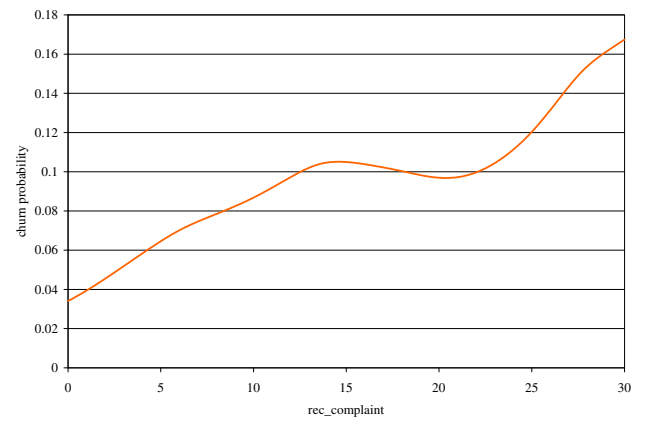
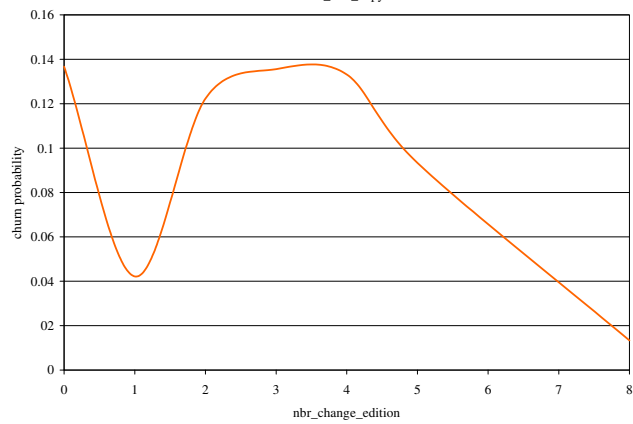
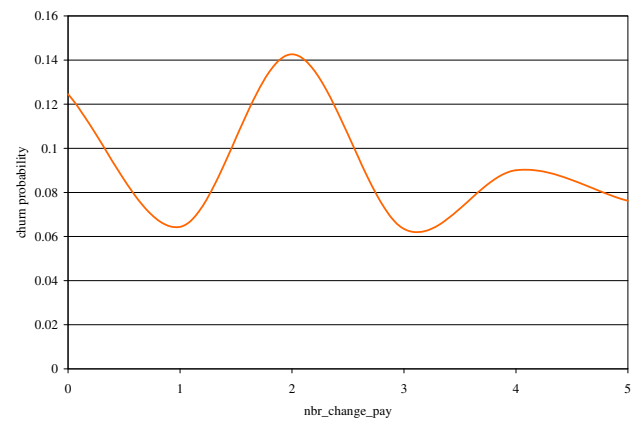
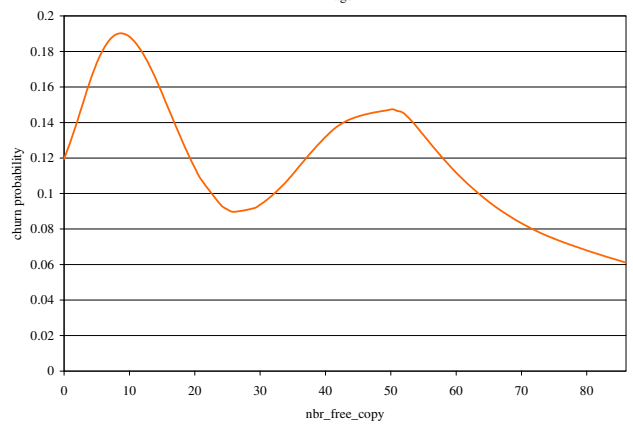
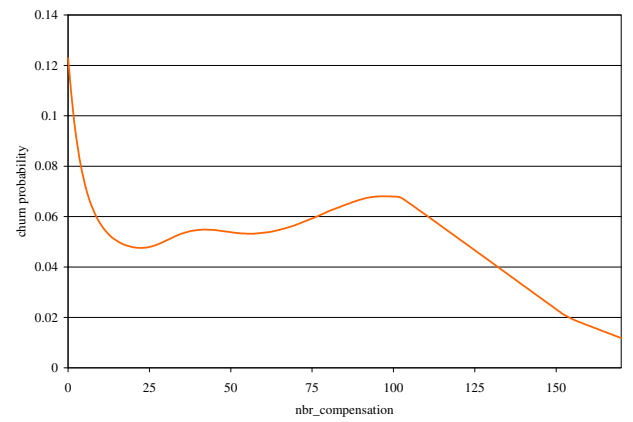
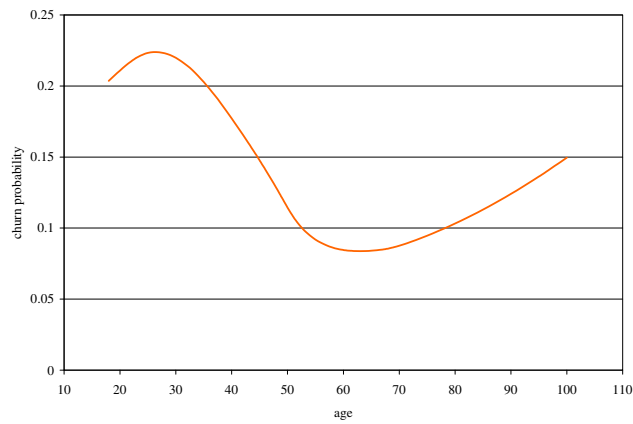


Inverted U-trend



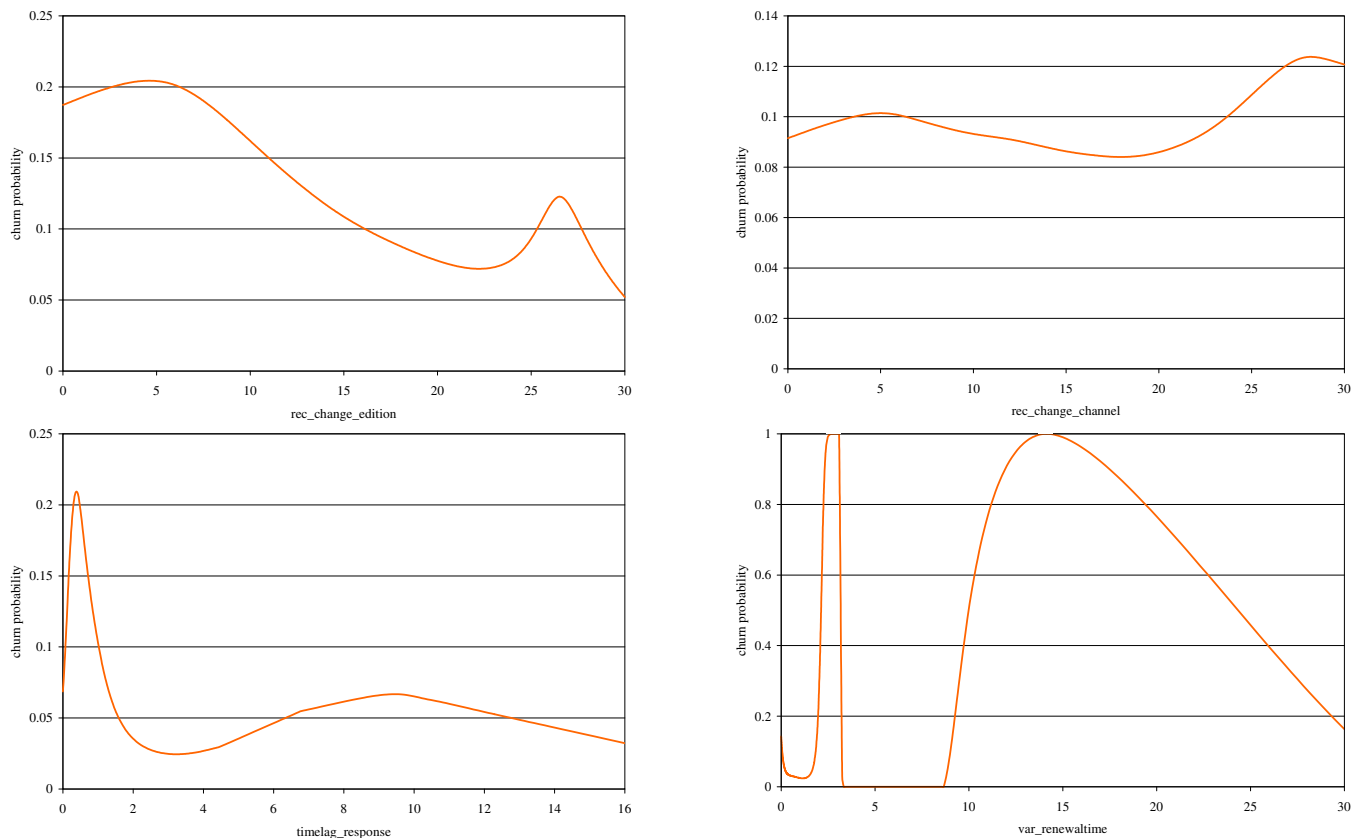
## Appendix B (continued)

## Complex trend



(continued on next page)

## Appendix B (continued)



## References

- Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and application*. Cary: SAS Institute Inc.
- Anderson, J. A. (1983). Logistic discrimination. In P. R. Krishnaiah & L. N. Kanal (Eds.), *Handbook of statistics* (Vol. 2). New York: Elsevier Science Publishing Co.
- Athanassopoulos, A. D. (2000). Customer satisfaction cues to support market segmentation and explain switching behaviour. *Journal of Business Research*, 47(3), 277–288.
- Bellman, R. (1961). *Adaptive control processes: A guided tour*. Princeton University Press.
- 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), 277–288.
- Burez, J., & Van den Poel, D. (2007). CRM at Canal + Belgique: Reducing customer attrition through targeted marketing. *Expert Systems with Applications*, 32(2), 277–288.
- Burez, J., & Van den Poel, D. (2009). Handling class imbalances in customer churn prediction. *Expert Systems with Applications*, 36(3).
- Cakir, O. (2009). Benders decomposition applied to multi-commodity, multi-mode distribution planning. *Expert Systems with Applications*, 36(4).
- Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1), 313–327.
- Coussement, K., & Van den Poel, D. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications*, 36(3).
- 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.
- Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.
- Friedman, J. H., & Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, 76, 817–823.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar).
- Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–318.
- Hastie, T., & Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, 82(398), 371–386.
- Hastie, T., & Tibshirani, R. (1990). *Generalized additive models. Monographs on statistics and applied probability* (vol. 43). Chapman and Hall. 335.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference and prediction*. Springer-Verlag Series in Statistics.
- Hoekstra, J. C., Leeflang, P. S., & Wittink, D. R. (1999). The customer concept: The basis for a new marketing paradigm. *Journal of Market Focused Management*, 4(1).
- Horowitz, J. L., & Savin, N. E. (2001). Binary response models: Logits, probits and semiparametrics. *Journal of Economic Perspectives*, 15(4).
- Hwang, H., & Euiho Suh, T. (2004). An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunications industry. *Expert Systems with Applications*, 26(2).
- Jones, M. A., Mothersbaugh, D. L., & Beatty, S. E. (2000). Switching barriers and repurchase intentions in services. *Journal of Retailing*, 76(2).
- Jowett, I. G., Parkyn, S. M., & Richardson, J. (2008). Habitat characteristics of crayfish (*Paraneophris planifrons*) in New Zealand streams using generalised additive models (GAMs). *Hydrobiologia*, 596(1).
- Kalyanam, K., & Shively, T. S. (1998). Estimating irregular pricing effects: A stochastic spline regression approach. *Journal of Marketing Research*, 35(1).
- Kim, Y. S. (2006). Toward a successful CRM: Variable selection, sampling and ensemble. *Decision Support Systems*, 41(2).
- Kim, H., & Yoon, C. (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28(9–10).
- Kumar, N., Scheer, L. K., & Steenkamp, J. B. (1998). Interdependence, punitive capability and the reciprocation of punitive actions in channel relationships. *Journal of Marketing Research*, 35(2).
- Lee, C. H., Wu, D. C., Lee, J. M., Wu, I. C., Goan, Y. G., Kao, E. L., et al. (2007). Carcinogenic impact of alcohol intake on squamous cell carcinoma risk of the oesophagus in relation to tobacco smoking. *European Journal of Cancer*, 43(7).
- Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2).
- Reichheld, F. F., & Sasser, W. E. Jr., (1990). Zero defections: quality comes to service. *Harvard Business Review*, 68(5).
- Reinartz, W. J., & Kumar, V. (2002). The mismanagement of customer loyalty. *Harvard Business Review*, 80(July).
- Reinartz, W. J., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, 67(1).
- Rosset, S., Neumann, E., Eick, U., Vatnik, N., & Idan, I., (2001). Evaluating campaign benefits and models for marketing campaigns. In Proceedings of the 7th international conference on knowledge discovery and data mining (KDD 01).
- Shaffer, G., & Zhang, J. Z. (2002). Competitive one-to-one promotions. *Management Science*, 48(9).

- Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1).
- Shively, T. S., Allenby, G. M., & Kohn, R. (2000). A nonparametric approach to identifying latent relationships in hierarchical models. *Marketing Science*, 19(2).
- Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, 13, 689–705.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper Collings Publishers.
- Teo, T. S. H., Devadoss, P., & Pan, S. L. (2006). Towards a holistic perspective of customer relationship management (CRM) implementation: A case study of the housing and development board, Singapore. *Decision Support Systems*, 42(3).
- Thomas, J. S. (2001). A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research*, 38(2).
- Torkzadeh, G., Chang, J. C.-J., & Hansen, G. W. (2006). Identifying issues in customer relationship management at Merck-Medco. *Decision Support Systems*, 42(2).
- 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.