A methodology based on profitability criteria for defining the partial defection of customers in non-contractual settings

M. Clemente[*], S. San Matías[1], V. Giner-Bosch

Department of Applied Statistics, Operations Research and Quality, Universitat Politècnica de València, Camino de Vera s/n, 46022 Spain.

Abstract

The defection or churn of customers represents an important concern for any company and a central matter of interest in customer base analysis. An additional complication arises in non-contractual settings, where the characteristics that should be observed to saying that a customer has totally or partially defected are not clearly defined. As a matter of fact, different definitions of the churn situation could be used in this context. Focusing on non-contractual settings, in this paper we propose a methodology for evaluating the short-time economic effects that using a certain definition of churn would have on a company. With this aim, we have defined two efficiency measures for the economic results of a marketing campaign implemented against churn, and these measures have been computed using a set of definitions of

---

[*]Corresponding author. Tel.: +34 96 387 74 90; Fax: +34 96 387 74 99.

Email addresses: `mclement@eio.upv.es` (M. Clemente), `vigibos@eio.upv.es` (V. Giner-Bosch)

[1]Prof. Susana San Matías sadly passed away in the Summer of 2012 while this paper was under review. We deeply regret the loss of such a brilliant researcher and good person and feel fortunate to have had the opportunity to work with her. Susana will be deeply missed.

partial defection. Our methodology finds that definition maximizing both efficiency measures and moreover, the monetary amount that the company should invest per customer in the campaign for achieving the optimal solution. This has been modelled as a multiobjective optimization problem that we solved using compromise programming. Numerical results using real data from a Spanish retailing company are presented and discussed in order to show the performance and validity of our proposal.

1. Introduction and background

1.1. Introduction

The fact of a customer ending his commercial relations with a company is usually known as *defection* or *churn.* For any firm, it has become a critical issue knowing beforehand which of their customers are more likely to become churners. Why is it so important to anticipate the defection of a customer? Reichheld & Sasser (1990) and Zeithaml et al. (1996) (among others) have analysed the negative impact of customer defections on a company profits. The loss of customers not only entails opportunity costs due to reduced sales, but also leads to an increased need to attract new customers, which is 5 to 6 times more expensive than selling to the existing customers (see also Bhattacharya (1998) or Dyché (2002)).

Churn has been extensively treated in the recent literature related to customer-base analysis. In this sense, a distinction between contractual and

non-contractual business settings is fundamental, because it is completely inappropriate to apply a model developed for a contractual setting in a non-contractual setting (see for example Reinartz & Kumar (2000)). Contractual environments are characterized by the existence of a contract between the firm and the customer; in such a case, the date of churn is clearly known and it matches up with the contract cancellation date. However, in a non-contractual setting there is no agreement or affiliation between the buyer and seller. In most cases, the customers can change their pattern of purchases without informing the seller. The difficulty lies in defining analytically when a customer becomes effectively inactive in an environment that is not governed by a contract between both parts.

1.2. Previous definitions of churn in non-contractual settings

There is relatively little literature dealing with churn definition in non-contractual settings. Most of the papers related to this case propose probability models; Fader & Hardie (2009) provide a review of such models that are well-suited for forward-looking projections in customer-base analysis. Probability models assume an underlying probability distribution for the buying behaviour of customers and this is used for estimating at any time T the probability of a customer being "alive" or active, the customer lifetime value or the expected future number of purchases. The negative binomial distribution (NBD) model by Ehrenberg (1959) was the starting point for modelling the repeated purchase behaviour. One of the first references proposing a NBD model is the work by Schmittlein et al. (1987); extensions and empirical validations of this model can be found in Schmittlein & Peterson (1994) or Fader et al. (2005), while applications can be found for example in Reinartz

3

& Kumar (2000, 2003) and Hopmann & Thede (2005). Also with a stochastic approach, Netzer et al. (2008) apply a Hidden Markov model in order to capture the dynamics of the behaviour of the customers (donors of several universities). Apart from these probability models, most definitions of churn use the product activity of a customer and a threshold fixed by a business rule. Following this line, Buckinx & Van den Poel (2005) used real data from a company in the retailing sector in order to propose an analytical definition of churners based basically on the mean frequency of purchases and on the interpurchase time. We will refer to the approach of Buckinx & Van den Poel in more detail later, when we formally define the class of churn definitions. Also in the retailing sector, Miguéis et al. (2012) and Miguéis et al. (2013) classify those customers who, from a certain period, did not buy anything else and those who in all subsequent periods spent less than a fixed threshold as churners. On the other hand, Glady et al. (2009) proposed instead to define churners as those customers with a decreasing future customer lifetime value, and for whom a retention action is profitable.

1.3. Aims and scope

As we can see, different definitions of churn in non-contractual settings have been (and could be) established. Which one is more appropriate? or more precisely, how to evaluate the adequacy of a churn definition? That is the main question that we address in this study. Because most companies rely on performance criteria in making their decisions, we recommend selecting a churn definition based on economic parameters.

In this paper, we propose a comprehensive methodology based on economic arguments in order to build a definition for the partial defection of

customers in a non-contractual setting. Despite being a vital aspect, there are very few references in the literature from this economical point of view. We claim that our objective is not exactly to establish a single definition of churn but to provide with a tool aimed to evaluate any churn definition and to decide if its implementation would be economically convenient for the company.

First of all, in Section 2 we propose a formal way to generate different definitions of partial defection, which we will use in this paper. Section 3 includes our proposal for evaluating economically a definition of churn; we propose to take into account the gains and losses that the execution of marketing actions to avoid customer's defection would mean for the company. Our methodology for choosing the most convenient (from the economic point of view) among several definitions is presented in Section 4; our proposal consists of maximizing both profits and return due to the campaign based on the churn definition, and this appears to be a multiobjective programming problem. In order to analyse the benefits of our methodology, in Sections 5 and 6 we present and discuss numerical results using real data provided by a company of commercial retail distribution in Spain. For comparison purposes, we have used the work of Buckinx & Van den Poel (2005), due to the similarities of our environments and data. Conclusions of the paper are shown in Section 7.

2. Defining churn

In this section, we present a new family of definitions of churn. Similarly to Buckinx & Van den Poel (2005), we will focus on partial defection and

we consider it as a change in the status of a customer. How do you define this change, is the first question. The fact of this definition being or not profitable, is a different one that we will face separately in a subsequent section.

In their paper, Buckinx & Van den Poel propose focusing only on the best customers of the company in order to study partial defection. At some specific point in time T, they define two segments of clients (loyal and unloyal customers) based on their behavioural attributes. A customer is classified into the loyal segment if he/she satisfies the following two conditions during a period of 5 months before T: (a) frequency of purchases is above average and; (b) coefficient of variation of the interpurchase time is below average. That is, loyal customers are those who shop frequently and exhibit a regular buying pattern; the rest of customers are considered to be unloyal. Their objective is to determine which behaviourally loyal clients in the database at time T may partially switch their purchases to another store after T. Thus, they classify a customer as partially defective if he has moved from the loyal segment before T to the unloyal segment during the 5 months after T.

2.1. The class of ad-hoc definitions of churn

We propose a generalization of the definition of churn of Buckinx & Van den Poel. In the following, at a time T we will call *period of observation* to the period of k months before T and *period of evolution* to the period of k months after T, for $k \geq 1$.

Let us consider a set $X(c, j)$ of behavioural variables in the database, each variable in this set measured monthly during a period j for each customer c. Moreover, consider a set of pairs $\{(v_i(X(c, j)), t_i), i = 1, \ldots, h\}$, $h \geq 1$,

where each $v_i(X(c, j))$ is an univariate function of the variables in $X(c, j)$. We will call *loyal segment* at the period j to the group of customers for which $v_i(X(c, j)) \geq t_i$ for all i, and *improvable segment* to the group formed by the rest of customers, for $j = 0$ (period of observation), 1 (period of evolution). Then, the partial defection of a customer can be formally defined as follows.

Definition 1. Given a temporal length k and a loyalty segmentation defined by $\{(v_i(X(c, j)), t_i), i = 1, \ldots, h\}$, we say that a *customer has partially defected* at the time $T + k$ if he belonged to the loyal segment during $j = 0$ and he has moved to the improvable segment in $j = 1$.

In the following, we will say that a churn definition belongs to the *ad-hoc class* of definitions, C_{Ad}, if a loyalty segmentation of the customers defined by $\{(v_i(X(c, j)), t_i), i = 1, \ldots, h\}$ is available and the partial defection of a customer is established using Definition 1. We point out that C_{Ad} includes a wide range of definitions, differing in the length k and in the set $\{(v_i(X(c, j)), t_i), i = 1, \ldots, h\}$. In fact, the definition of churn by Buckinx & Van den Poel belongs also to this general class.

Later in Subsection 4.1 we will propose a selection of loyalty segmentation variables v_i with their corresponding thresholds t_i to be considered for defining churn, that will be used to illustrate our methodology.

3. Measuring the economic impact attributable to a definition of churn

In our opinion, a definition of churn is well-defined if the transition of customers from the loyal segment to the improvable segment would result

in an economic loss for the company. Let D be a definition of churn, and suppose that a predictive model is developed at a time T in order to forecast which of the loyal customers during $j = 0$ will defect (according to definition D) at the end of the period of evolution. The predictive model will assign a propensity to defect to each of those loyal customers, and customers with a probability greater than 0.5 (although this cutoff value may vary) will be automatically targeted as future churners. In such a case, some marketing campaigns or retention actions could be carried out in order to prevent churn and, consequently, avoiding part of the economic loss derived from it.

Our proposal consists of analysing the short-term economic benefit that such a marketing action could report in order to assess the convenience of its implementation, and consequently, whether the definition of churn is or not useful. The idea of using economic measures to evaluate and compare different models has been applied in recent similar studies. For instance, Verbeke et al. (2012) use maximum profit to evaluate customer churn prediction models in the telecommunication sector.

In this context, it is appropriate to consider classical static decision-making attributes in a business investment environment, such as benefits, costs, invested capital, average net flow, payback period, return of investment (ROI), etc. As we mean, this approach is pertinent when dealing with relatively small and short-term investments, as it is our case, and usually requires quite a small amount of information and calculations (see Jacobson, 1987 and Phillips & Phillips, 2008, for example). Other dynamic and/or modern models are sometimes difficult to apply due to the huge amount of information they need.

More precisely, we are going to introduce two efficiency measures; the first one is based on the net profit reported by the marketing campaign against churn, and the other one is inspired on the ROI of such a campaign. As it is known, the ROI is a measure of performance used for evaluating the efficiency of an investment and it is computed as benefit divided by cost, expressing the result as a percentage or ratio. It is usually built using average values.

3.1. Notation

In the following, we will call:

$S_{L,0}$:= set of loyal customers during $j = 0$

$n_{L,0}$:= number of customers in $S_{L,0}$.

$S_{D,1}$:= set of loyal customers during $j = 0$ defecting at the end of $j = 1$.

$n_{D,1}$:= number of customers in $S_{D,1}$.

$S_{T,0}$:= set of loyal customers during $j = 0$ targeted as future churners by the predictive model.

$n_{T,0}$:= number of customers in $S_{T,0}$.

p := proportion of initial loyal customers targeted as likely future churners.

s_p := proportion of churners correctly identified as such by the predictive model (sensitivity).

pm := profit margin of the company, $0 \leq \text{pm} \leq 1$.

3.2. Efficiency measures of a marketing campaign against churn

The first measure that we propose is the POAC function (*Profit On an Against-Churn campaign*). We define the POAC function associated with the definition of chrun D as:

$$\text{POAC}_D(g) = \text{NAL}_D(g) - \text{Co}_D(g) \,, \tag{1}$$

the parameter g being the marginal action cost or monetary units per customer invested by the company in the marketing campaign against churn

(and/or other associated retention actions such as cross-selling, up-selling, promotions, etc.), and:

$\mathrm{NAL}_D(g)$ (*Net Avoidable Loss*) := estimation of the net income, the loss of which could be avoided due to the marketing campaign implemented over $S_{\mathrm{T},0}$, in short term.

$\mathrm{Co}_D(g)$:= cost of the marketing campaign implemented over $S_{\mathrm{T},0}$.

Thus, POAC is computed as the total profit that the company would obtain when the marketing campaign addressed to avoid churn is executed.

The second efficiency measure that we propose is the ROAC function (*Return On an Against-Churn campaign*). We define the ROAC function associated with D as:

$$\mathrm{ROAC}_D(g) = \frac{\mathrm{POAC}_D(g)}{\mathrm{Co}_D(g)}. \tag{2}$$

That is, ROAC is defined as the ratio of return of the campaign, or number of monetary units obtained for every monetary unit invested in the campaign against churn (the churn being defined by D). Let us note that the higher the values of POAC and ROAC for a certain value of g, the most convenient will be a definition of churn (from an economic point of view), because its implementation will lead to a higher profit for the company.

In order to compute the POAC and ROAC functions for each specific definition of churn, we should state explicit expressions for $\mathrm{NAL}_D(g)$ and $\mathrm{Co}_D(g)$. In the following subsections we present our proposals in this sense.

10

3.3. Estimation of the Churn Campaign Cost and the Net Avoidable Loss

The number of customers to be included in the against-churn campaign (target people of the retention action) is $n_{T,0}$, and let us note that:

$$n_{T,0} = n_{L,0} \cdot p.$$

Therefore, we will estimate the required cost to implementing the marketing campaign against churn as:

$$\mathrm{Co}_D(g) = n_{L,0} \cdot p \cdot g. \tag{3}$$

On the other hand, the estimation of the so-called Net Avoidable Loss is a complex issue, because it depends on several factors, all of them with a degree of uncertainty. We propose the following estimation for $\mathrm{NAL}_D(g)$:

$$\mathrm{NAL}_D(g) = \mathrm{pm} \cdot (M_E - M_R) \cdot s_p \cdot e(g), \tag{4}$$

where:

$M_E :=$ estimation of the total purchase monetary amount that the group of churners $S_{D,1}$ would have spent during the period of evolution if they had not defected.

$M_R :=$ real total purchase monetary amount spent by the group of churners $S_{D,1}$ during the period of evolution.

$e(g) :=$ estimated proportion of the predicted future churners that would react favourably when the marketing campaign against churn will be implemented (i.e., would not defect due to this campaign), if the company invested in the campaign up to g monetary units per customer.

Let us explain the derivation of the expression (4). The difference $M_{\mathrm{E}} - M_{\mathrm{R}}$ is an estimation of the amount due to the lost purchases of the churners, and therefore it is an upper bound for the amount to be recovered. This sum has to be weighted by the constant pm, the estimated profit margin, in order to obtain the lost net income of the company. However, from the estimated total lost income pm $\cdot (M_{\mathrm{E}} - M_{\mathrm{R}})$ only the percentage corresponding to the future churners who are captured by the predictive model (s_p) and who react favourably to the marketing campaign $(e(g))$ is subject to be recovered.

3.4. Numerical expressions for the POAC *and* ROAC *functions*

Replacing (3) and (4) into (1) and (2), the following expressions are obtained:

$$\mathrm{POAC}_D(g) \;=\; \mathrm{pm} \cdot (M_{\mathrm{E}} - M_{\mathrm{R}}) \cdot s_p \cdot e(g) - n_{\mathrm{L},0} \cdot p \cdot g \,, \tag{5}$$

$$\mathrm{ROAC}_D(g) \;=\; \frac{\mathrm{pm} \cdot (M_{\mathrm{E}} - M_{\mathrm{R}}) \cdot s_p \cdot e(g)}{n_{\mathrm{L},0} \cdot p \cdot g} - 1 \,. \tag{6}$$

For practical purposes, all the elements that make part of the expressions (5) and (6) should be numerically computed. Estimates for $M_{\mathrm{R}}, M_{\mathrm{E}}$ can be obtained as:

$$M_{\mathrm{R}} \;=\; \sum_{c \in S_{\mathrm{D},1}} \sum_{i=1}^{k} \mathrm{Monetary}(c, i) \,, \tag{7}$$

$$M_{\mathrm{E}} \;=\; \left(\frac{\displaystyle\sum_{c \in (S_{\mathrm{L},0} - S_{\mathrm{D},1})} \sum_{i=1}^{k} \mathrm{Monetary}(c, i)}{n_{\mathrm{L},0} - n_{\mathrm{D},1}} \right) \cdot n_{\mathrm{D},1} \,, \tag{8}$$

where $\mathrm{Monetary}(c, i)$ stands for the total amount of the purchases of customer c during the month i.

The value for pm should be fixed by the user. In the next subsections we discuss how to determine p, s_p and $e(g)$.

3.4.1. Estimating p and s_p

We need to set a value for the parameter p, the proportion of customers to be included into the campaign. We are going to assume that this proportion will be equal to the real proportion of churners. That is:

$$p = \frac{n_{\mathrm{D},1}}{n_{\mathrm{L},0}} \, . \tag{9}$$

This seems to be a quite reasonable assumption; the predictive model will select the $100 \cdot p\%$ of customers with the highest probabilities of churn, with p as in (9). If the predictive model is accurate enough, this means that this group of customers should coincide practically with the real churners.

On the other hand, we have to select an appropriate value for the predictive model sensitivity s_p, but it depends on the performance of the future classificator, and this cannot be fixed a priori. We recommend using one of two possibilities; the first one is to select $s_p = 0.5$, a very conservative value. The second possibility is to determine a value for s_p satisfying a permissive requirement. A good choice could be setting a value guaranteeing that the maximum value of ROAC is nonnegative. In this case, for a certain churn definition D, we would obtain the value $g_{\max}$ where the maximum value for the ROAC function associated to D is reached, that is:

$$g_{\max} = \underset{g}{\mathrm{argmax}} \, \mathrm{ROAC}_D(g) \, .$$

Using (6), the maximum value of ROAC would be:

$$\mathrm{ROAC}_D(g_{\max}) = \frac{\mathrm{pm} \cdot (M_{\mathrm{E}} - M_{\mathrm{R}}) \cdot s_p}{n_{\mathrm{L},0} \cdot p \cdot g_{\max}} \cdot e(g_{\max}) - 1\,. \tag{10}$$

Finally, from expression (10), the values for s_p guaranteeing that the maximum value of ROAC is nonnegative would be:

$$s_p \;\geq\; \frac{n_{\mathrm{L},0} \cdot p \cdot g_{\max}}{\mathrm{pm} \cdot (M_{\mathrm{E}} - M_{\mathrm{R}}) \cdot e(g_{\max})}\,. \tag{11}$$

3.4.2. Estimating the a priori effectiveness of the marketing campaign against churn $e(g)$

The expected effectiveness on a particular customer included as target person of the campaign against churn has been modelled as a function $e(g)$, g being the marginal action cost, as we have already stated in Subsection 3.3. In this subsection we are going to explain the conditions for selecting the function $e(g)$. First, we assume the next hypothesis to be held.

Hypothesis 1. A marketing action addressed to prevent churn cannot worsen the purchasing behaviour of a customer.

More precisely, let Λ be a constant, with $0 < \Lambda < 1$, and let us define the random variable $G =$ *(theoretical) investment required in each customer for the campaign being effective.* Then, we define the function $e(g)$ as:

$$e(g) \;=\; \Lambda \cdot P(G \leq g) \;\;,\;\; \forall g \geq 0\,. \tag{12}$$

It is easy to show that $e(g)$ defined according to (12) is a nonnegative, nondecreasing function, and that $\lim_{g \to 0^+} e(g) = 0$ and $\lim_{g \to +\infty} e(g) = \Lambda$. Moreover, as a consequence we have that $\forall \epsilon > 0$, there exist $\delta_\epsilon^{\mathrm{L}}, \delta_\epsilon^{\mathrm{U}}$ such that:

$$
\begin{aligned}
e(g) &< \epsilon & \forall g \in [0, \delta_\epsilon^{\mathrm{L}})\,, \\
e(g) &> \Lambda - \epsilon & \forall g \in (\delta_\epsilon^{\mathrm{U}}, +\infty)\,.
\end{aligned}
$$

14

Figure 1 illustrates all the elements involved in the estimation of $e(g)$. The parameter Λ is a corrective factor. It represents the proportion of target people that would react (favourably) to the marketing retention action if we spent an infinity amount of money in each target person, that is, when g tends to be equal to $+\infty$. It is introduced in our model because, in practical terms, there is always a subset of target people that will never react to the marketing action, no matter the amount of money we invest on them. That means that we are assuming that the maximum possible effectiveness of the campaign would be $\Lambda < 100\%$. This factor has to be estimated according to previous knowledge.

The parameter ϵ is supposed to be a very small quantity (say, 1%, for instance), meaning that $(1 - \Lambda - 2\epsilon)100\%$ of the target people will react (favourabily) to the marketing retention action if the amount invested per person g is within $[\delta_\epsilon^{\mathrm{L}}, \delta_\epsilon^{\mathrm{U}}]$. This is a simple way to check whether our estimation of $e(g)$ makes sense or not, in a practical case.

The probability distribution of the variable G can be approached either by standard probability distributions or using the knowledge about previous marketing campaigns. A practical example of how these parameters and the distribution of G can be determined is shown in Section 5.

3.5. A numerical example

Next we show how to compute POAC and ROAC in practice, for a given definition of churn. A complete numerical experience showing how to compare different churn definitions based on a real situation is also presented and discussed in Sections 5 and 6.

Let us consider a given definition of churn D together with an appropriate

Proportion of target people that will react favourably to the marketing retention action

Λ

$\Lambda-\varepsilon$

$e(g)$

ε

0

$\delta_\varepsilon^{\mathrm{L}}$ $\delta_\varepsilon^{\mathrm{U}}$

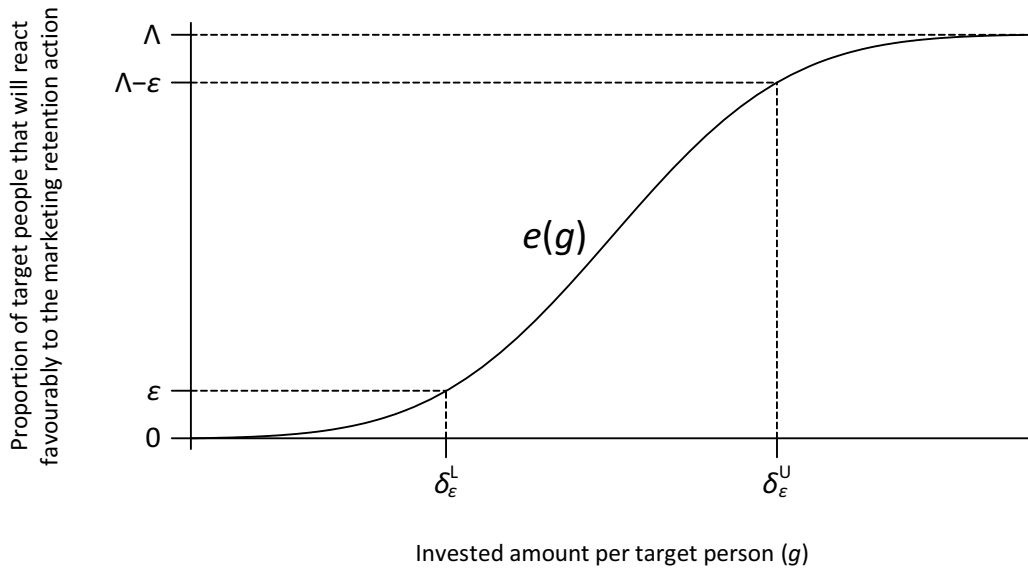Invested amount per target person (g)

Figure 1: Estimation of the proportion of target churners that will react favourably to a marketing retention action as a function of the amount invested per customer.

value for k, say $k = 5$ for instance. Also, let us take an instant of time T such that the behavioural information of our customers during the interval of time from $T - k$ to $T + k$ is available.

From that, we can identify the sets $S_{\mathrm{L},0}$ and $S_{\mathrm{D},1}$ in our database, and then obtain the values for $n_{\mathrm{L},0}$, $n_{\mathrm{D},1}$ and the total amount of money those customers spent in the evolution period —for example:

- $n_{\mathrm{L},0} = 37000$ customers.

- $n_{\mathrm{D},1} = 7000$ customers.

- Amount spent by customers in $S_{\mathrm{L},0} - S_{\mathrm{D},1}$ during the evolution period $= 39.00 \cdot 10^6$ euros.

- Amount spent by customers in $S_{\mathrm{D},1}$ during the evolution period $= 3.50\cdot$

10^6 euros.

Based on this, we can calculate M_R and M_E using (7) and (8), respectively:

$$
\begin{aligned}
M_\mathrm{R} &= 3.50 \cdot 10^6 \text{ euros,} \\
M_\mathrm{E} &= \frac{39.00 \cdot 10^6}{37000 - 7000}\, 7000 = 9.00 \cdot 10^6 \text{ euros.}
\end{aligned}
$$

There are also some parameters in our model that need to be estimated. Let us assign the following values to them:

- $p = n_{\mathrm{D},1}/n_{\mathrm{L},0} = 18.92\%$, according to (9).

- $s_p = 0.50$.

- pm $= 0.3$.

We need also to assume an expression for the effectiveness function $e(g)$ (12). As further shown in the numerical experience in Section 5, the shifted exponential distribution $\mathrm{Exp}(\mu, \theta)$ with $\mu = 30$ and $\theta = 5$ is a valid pattern for $e(g)$, with an upper bound of responders of $\Lambda = 0.75$.

With all the previous calculations and assumptions, concrete expressions for POAC and ROAC can be derived from equations (5) and (6), respectively:

$$
\begin{aligned}
\mathrm{POAC}_D(g) &= 0.30(9.10 - 3.50)10^6 \cdot 0.50 \cdot 0.75\, F(g) - 37000 \cdot 0.1892\, g \\
&= 945000\, F(g) - 7000\, g \text{ euros,}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{ROAC}_D(g) &= \frac{0.30(9.10 - 3.50)10^6 \cdot 0.50 \cdot 0.75\, F(g)}{37000 \cdot 0.1892\, g} - 1 \\
&= \frac{945000\, F(g)}{7000\, g} - 1\,,
\end{aligned}
$$

where F represents the cumulative distribution function of the shifted exponential distribution that was mentioned before, that is

$$F(g) = 1 - e^{-\frac{g-\theta}{\mu}} .$$

Finally, according to equations (4) and (3), the expressions for NAL and Co would be the following:

$$
\begin{aligned}
\text{NAL}_D(g) &= 0.30(9.10 - 3.50)10^6 \cdot 0.50 \cdot 0.75\, F(g) \\
&= 945000\, F(g) \text{ euros,} \\
\text{Co}_D(g) &= 37000 \cdot 0.1892\, g = 7000\, g \text{ euros.}
\end{aligned}
$$

3.6. Suitability and scope of the model

It is necessary to include here a small discussion about all the assumptions we have adopted to build the expressions for POAC and ROAC, especially those regarding the estimation of NAL.

We think that the proposed model is complete, general and flexible enough for being adapted to many different situations, but also simple enough to be put into practice, as we have shown in the previous subsection.

In the end, nearly all the involved parameters, such as $n_{D,1}$, $n_{L,0}$ pm, Λ, M_E, M_R, etc., need to be empirically estimated from existing knowledge, as it usually happens with this kind of models. According to our experience, this is not an impossible task in the context of companies that have a non-contractual relationship with their customers and that save records of their behaviour via loyalty cards, for instance (which is the same context as similar studies such as those by Buckinx & Van den Poel, 2005 and Miguéis et al., 2013). In fact, we evidence that with a real example in Sections 5 and 6.

18

Moreover, the way we propose to estimate the effectiveness $e(g)$ provides the model with enough flexibility to cover many different situations (see figure 1 and also figure 2 later in Section 5). The selection of an appropriate distribution for G, apart from its original meaning, allows to properly customise the way the expression of NAL (4) behaves as a function of g.

Anyway, the approach we have proposed in this section can be easily adapted to better meet some concrete requirements or situations, if necessary. That is, we have proposed a practical but formal way to estimate key decision attributes such as POAC and ROAC, but it obviously could inspire other (valid) ways to do it.

4. Methodology for the optimal selection of a churn definition by multiobjective programming

In this section we present our methodology for selecting the most valuable definition of churn for a company, among the different available options. First of all, we suggest using a specific subset of the class of churn definitions we have introduced in Subsection 2.1, involving the most used segmentation variables for characterising customer churn. Then we show how to determine which of those churn definitions would be the most appropriate for a given concrete context.

4.1. New ad-hoc churn definitions

In this subsection we select a family of new ad-hoc definitions of churn in the way introduced in Subsection 2.1, in order to later test the goodness of our methodology of evaluation. The segmentations for these definitions are based in the *recency-frequency-monetary* (RFM) concepts and in the length

of relationship, as an indicator of the level of commitment of a client. All of those concepts are widely used when studying the behaviour of customers toward the businesses they interact with (see Bhattacharya, 1998; Keaveney & Parthasarathy, 2001; Buckinx & Van den Poel, 2005 for example). We suggest using the following loyalty segmentation variables v_i:

L Length of relationship (days)

r $-$average of the k monthly interpurchase times

f average of the k monthly number of purchases

m average of the k monthly purchase amounts

r_c $-$coefficient of variation of the k monthly interpurchase times

f_c $-$coefficient of variation of the k monthly number of purchases

m_c $-$coefficient of variation of the k monthly purchase amounts

For these variables, the thresholds for defining the loyal segment are:

$$
\begin{aligned}
t &= \text{ average of } v \text{ in the database,} && \text{for } v = L, f, m\,, \\
t &= -\text{average of } v \text{ in the database,} && \text{for } v = r, r_c, f_c, m_c\,.
\end{aligned}
$$

We have fixed $k = 5$ and we have used 16 combinations of these variables in order to build different loyalty segmentations which generate different definitions of churn. They are shown in Table 1. Notice that D3 corresponds to the definition of churn proposed by Buckinx & Van den Poel.

4.2. Determining the most valuable definition of churn

Our proposal consists of selecting the definition of churn maximizing both POAC and ROAC, as defined in Section 3. This can be modelled as a mul-

| | Definition | | | | | | | | | | | | | | | |
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
L									x	x	x	x	x	x	x	x
r	x	x			x		x		x	x			x		x	
f	x	x	x					x	x	x	x					x
m	x	x		x		x			x	x		x		x		
r_{c}		x	x	x					x	x	x					
f_{c}		x			x	x			x				x	x		
m_{c}		x					x	x	x						x	x

Table 1: Segmentation variables used in each of the 16 definitions of churn.

tiobjective optimization problem, in the following way:

$$(\text{MP}) \quad \text{Max} \qquad \{\text{POAC}_D(g), \ \text{ROAC}_D(g)\}$$
$$\text{subject to} \quad g \geq 0$$
$$D \in \{\textit{set of available definitions of churn}\}$$

The solution of this problem will be a pair (D^*, g^*) formed by the churn definition and an associated marginal action cost that make POAC and ROAC as high as possible at the same time.

4.2.1. Selecting a multiobjective optimization method

There are several different ways to address a multiobjective optimization problem such as (MP).

As it is known, in most cases it is not possible to find a solution in the feasible region that simultaneously optimises (in our case, maximises) the value of all the objective functions being considered (POAC and ROAC, in our case). The usual approach is determining the so-called efficient set, that is, the set of all solutions x in the feasible region such that there is no feasible

solution being at least as good as x with regards to *all* the objectives being considered, and strictly better than x regarding at least one of those objectives (see, for example, Marler & Arora, 2004; Ehrgott, 2005 for a more formal definition). Then, one or several solutions belonging to the efficient set can be selected as *optimal* according to external subjective preferences. Those preferences, if any, are expressed by a decision-maker, normally by ordering or assigning relative importances to objectives *before* solving the problem (*a priori methods*), but also *during* or *after* generating the efficient set (*interactive* or *a posteriori methods*, respectively). There exist also some solving strategies that produce solutions not necessarily belonging to the efficient set. A comprehensive survey of classic and state-of-the-art multiobjective optimization methods can be found at Marler & Arora (2004). Miettinen (1999) and Ehrgott (2005) provide a complete and updated view of these and other concepts regarding multicriteria optimization.

We propose addressing the multiobjective problem (MP) with an *a priori* strategy. As it is said before, in this way, the decision-maker is given one single (efficient) *optimal* solution as a result of his/her preferences concerning the relative importance of each objective goal. According to our experience, we consider that, in the context of marketing action design, this option is generally better than offering several possibilities to choose. Moreover, if required, it is possible to generate different *optimal* solutions by adjusting the relative preferences and solving the problem again.

More precisely, we have adopted the weighted global criterion or compromise programming method (originally proposed by Yu, 1973 and Zeleny, 1973), because it is suitably accurate and flexible enough to incorporate

decision-marker's preferences. We will describe the method in the next subsection.

Other *a priori* methods could be also considered. According to the classification and scores given by Marler & Arora (2004), compromise programming happens to be a strategy with a quite low technical complexity, which makes it very suitable for the context we are dealing with.

4.2.2. Procedure for determining the optimal definition of churn

Compromise programming identifies the optimal solution of (MP) as the one that has the shortest distance to an ideal point (F_P^M, F_R^M) where the multiple objectives simultaneously reach their optimal values; in general, the ideal point is not attainable. In our case, let us call:

$$
\begin{aligned}
(D_P, g_{\max,P}) &:= \underset{D,g}{\arg\max}\, \mathrm{POAC}_D(g)\,, & (13)\\
(D_R, g_{\max,R}) &:= \underset{D,g}{\arg\max}\, \mathrm{ROAC}_D(g)\,. & (14)
\end{aligned}
$$

That is, (13) and (14) are the optimal solutions that individually maximize each of the two objective functions. Then, the ideal point is:

$$
\left(F_P^M, F_R^M\right) = (\mathrm{POAC}_{D_P}(g_{\max,P}), \mathrm{ROAC}_{D_R}(g_{\max,R}))\,.
$$

On the other hand, let us call:

$$
\begin{aligned}
F_P^m &:= \min\left(\mathrm{POAC}_{D_P}(g_{\max,P}), \mathrm{POAC}_{D_R}(g_{\max,R})\right)\,,\\
F_R^m &:= \min\left(\mathrm{ROAC}_{D_P}(g_{\max,P}), \mathrm{ROAC}_{D_R}(g_{\max,R})\right)\,.
\end{aligned}
$$

It is not difficult to see that:

$$
\begin{aligned}
F_P^m &= \mathrm{POAC}_{D_R}(g_{\max,R})\,, & (15)\\
F_R^m &= \mathrm{ROAC}_{D_P}(g_{\max,P})\,. & (16)
\end{aligned}
$$

The compromise solution for (MP) will be obtained by solving the problem of minimizing the distance to the ideal point, with a metric $h > 0$:

$$\text{(P)} \quad \text{Min} \quad \left[\lambda_P{}^h \cdot \left(\frac{F_P^M - \text{POAC}_D(g)}{F_P^M - F_P^m} \right)^h + \lambda_R{}^h \cdot \left(\frac{F_R^M - \text{ROAC}_D(g)}{F_R^M - F_R^m} \right)^h \right]^{1/h}$$

$$\text{subject to} \quad g \geq 0$$

$$D \in \{set\ of\ available\ definitions\ of\ churn\}$$

where the parameters λ_P, λ_R represent the relative importance of each of the objectives, with $\lambda_P + \lambda_R = 1$, $\lambda_P, \lambda_R \geq 0$ (generally, $\lambda_P, \lambda_R > 0$). Notice that in (P), each of the original objective functions (POAC and ROAC) have been normalized in order to get comparable terms to be added up. We have used the euclidean metric with $h = 2$ for (P).

To sum up, two steps are involved in the compromise programming. The first step is to find the best values (13) and (14) of each objective within the computation domain, and the second step is to seek the optimal solution (D^*, g^*) of (P), which will be the solution for our problem (MP). The solution of (P) represents a compromise between net profit (POAC) and marginal profit with respect to the costs (ROAC). The user is able to control the importance of each of these two objectives by choosing appropriate values for λ_P, λ_R. Let us note that by solving (P), we will obtain not only the most adequate definition of churn D, but also the amount g that we should spend per customer in order to achieve the compromise solution.

Algorithm 1 shows the whole procedure that we have used for finding the optimal solution of (MP).

Algorithm 1

Init. Select values for λ_P, λ_R such that $\lambda_P + \lambda_R = 1$, $\lambda_P, \lambda_R \geq 0$. Set $i = 1$, $i_P = 0$, $i_R = 0$, $g_P = 0$, $g_R = 0$, $\text{POAC}_{\text{ref}} = 0$, $\text{ROAC}_{\text{ref}} = 0$, $\mathbb{D} = \{D_j \text{ available definitions of churn}, j = 1, \ldots, S\}$

Step 1. Obtain $g_P^i = \underset{g}{\text{argmax}}\, \text{POAC}_{D_i}(g)$ and $g_R^i = \underset{g}{\text{argmax}}\, \text{ROAC}_{D_i}(g)$. Go to the Step 2.

Step 2. If $\text{POAC}_{D_i}(g_P^i) > \text{POAC}_{\text{ref}}$, then do $\text{POAC}_{\text{ref}} = \text{POAC}_{D_i}(g_P^i)$, $g_P = g_P^i$, $i_P = i$.

If $\text{ROAC}_{D_i}(g_R^i) > \text{ROAC}_{\text{ref}}$, then do $\text{ROAC}_{\text{ref}} = \text{ROAC}_{D_i}(g_R^i)$, $g_R = g_R^i$, $i_R = i$.

Go to the Step 3.

Step 3. If $i < S$, then do $i = i + 1$ and go to the Step 1. Otherwise, go to the Step 4.

Step 4. If $i_P = 0$ and $i_R = 0$, it is not possible (for any definition in $\mathbb{D}$) to find positive POAC nor ROAC, then STOP. Otherwise, do:

$$g_{\text{max},P} = g_P, \quad D_P = D_{i_P}, \quad F_P^M = \text{POAC}_{\text{ref}},$$
$$g_{\text{max},R} = g_R, \quad D_R = D_{i_R}, \quad F_R^M = \text{ROAC}_{\text{ref}}.$$

Go to the Step 5.

Step 5. Compute F_P^m, F_R^m as in (15) and (16) and go to the Step 6.

Step 6. For $i = 1, \ldots, S$, obtain the optimal solution g_i of the univariate

subproblem (Pi):

(Pi) Min z_i

subject to

$$z_i = \sqrt{{\lambda_\mathrm{P}}^2 \cdot \left(\frac{F_\mathrm{P}^\mathrm{M} - \mathrm{POAC}_{D_i}(g)}{F_\mathrm{P}^\mathrm{M} - F_\mathrm{P}^\mathrm{m}} \right)^2 + {\lambda_\mathrm{R}}^2 \cdot \left(\frac{F_\mathrm{R}^\mathrm{M} - \mathrm{ROAC}_{D_i}(g)}{F_\mathrm{R}^\mathrm{M} - F_\mathrm{R}^\mathrm{m}} \right)^2}$$

$$g \geq 0$$

Let $i*$ be the index such that $z_{i*} = \min\{z_i : i = 1, \ldots, S\}$. Then do $D^* = D_{i*}$, $g^* = g_{i*}$ and STOP.

5. Numerical experiences

We have used real data from a fast moving consumer goods (FMCG) retail distribution company in Spain in order to assess the validity of our proposed methodology. The data for the analysis were obtained from the available information from customers using the loyalty card of the company. From the whole database, including transactional and socio-demographic information, we extracted a random sample with records of 140202 customers from June 2006 to March 2007.

The first objective of our empirical study was to obtain the estimated POAC and ROAC functions for each of the new definitions of churn proposed in Subsection 4.1. Using them, we were able to solve the problem (MP) which provided us with the most profitable definition of churn to be used by the company, our final objective. The numerical computation of the POAC and ROAC functions using (5) and (6) requires selecting a probability distribution for modelling $e(g)$, and fixing the values of several undetermined parameters, namely k, pm and s_p.
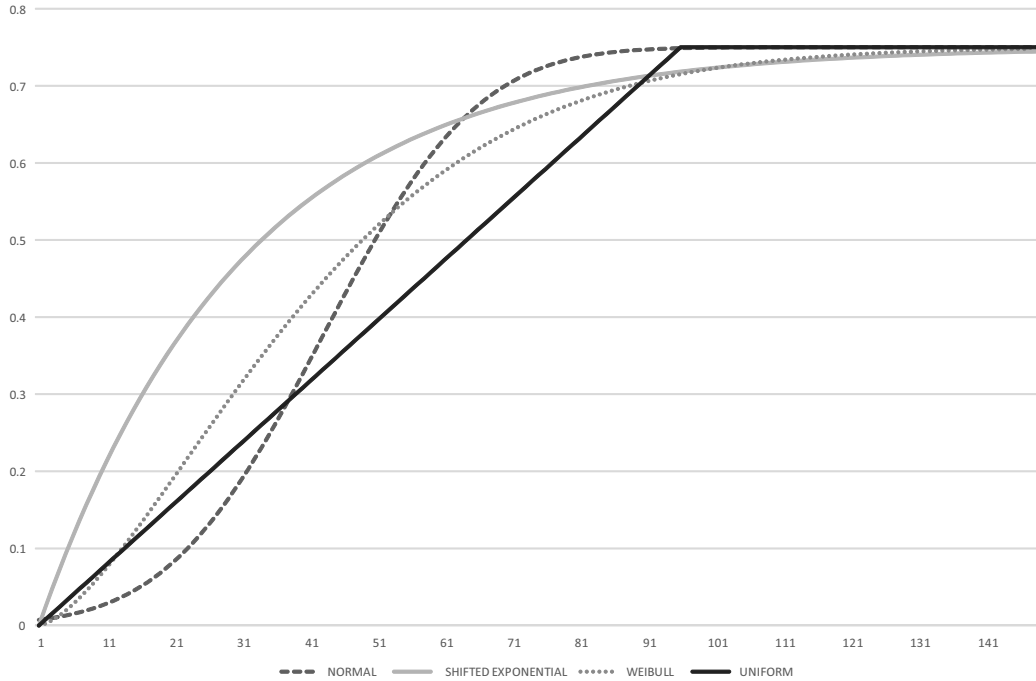
Figure 2: Different distributions for $e(g)$ that have been tested in our numerical experience.

Concerning to $e(g)$, we tested the following four probability distributions (see figure 2): (a) shifted exponential distribution $\mathrm{Exp}(\mu, \theta)$, with $\mu = 30, \theta = 5$; (b) Weibull distribution $\mathrm{We}(t_0, \eta, \beta)$, with $t_0 = 5, \eta = 45, \beta = 1.5$; (c) Normal distribution $\mathrm{N}(\mu, \sigma)$, with $\mu = 47, \sigma = 18$; and (d) Uniform distribution $\mathrm{U}(a, b)$, with $a = 5, b = 100$. Each of these distributions is associated to a different evolution pattern of the customer response to a marketing campaign. For modelling a quick response pattern, we would consider the shifted exponential distribution; for modelling a response with a constant growth pattern, we would consider the uniform distribution.

The distribution parameters have been chosen in order to lead to the values $\Lambda = 0.75$, $\epsilon = 0.01$, $\delta_\epsilon^{\mathrm{L}} = 5$ and $\delta_\epsilon^{\mathrm{U}} = 150$. With this choice, we are

assuming that there is barely effectiveness on the marketing campaign below an expenditure of €5 per customer; on the other hand, the maximum number of customers who could not defect due to the campaign effects would be at the most 75% of the total future churners who participate into the marketing campaign, and this maximum would be practically reached when expending €150. We think these are reasonable values for the context in which we are, but of course they can be adjusted depending on the characteristics of the company.

Concerning the rest of parameters, we set $k = 5$ months, pm $= 0.3$ (this is the most frequent value for the profit margin in the literature). As for the sensitivity, and in order to get comparable results, we selected the same value for s_p to be used with all the definitions of churn and all distributions, and therefore we decided to use $s_p = 0.5$.

On the other hand, we used $\lambda_P = \lambda_R = 0.5$ in Algorithm 1, giving in this way the same importance to both objective functions.

6. Results and discussion

First of all, with the goal of evaluating the performance of the different distributions, we obtained the maximum value of the ROAC functions for each churn definition and for each of the probability distributions mentioned in Section 5. Figure 3 shows these results in a chart. Each graph line in the plot represents a probability distribution, and each individual point in the line stands for the maximum ROAC of a definition model. We point out that each of these points represents the best expected return (ROAC) that we could expect using a particular definition and distribution. In the
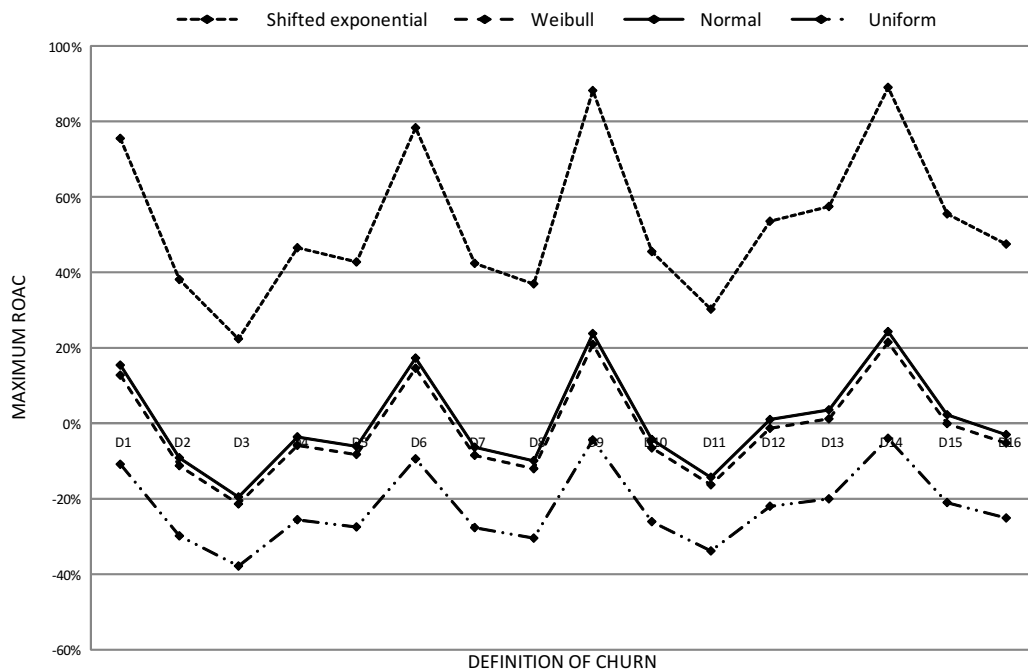
Figure 3: Maximum ROAC of the ad-hoc definitions of churn, for the different probability distributions, sensitivity 0.5.

horizontal axis, the first 8 points correspond to the definitions not using the variable L (Length of relationship), while the last 8 points are associated with the maxima for definitions using L.

From Figure 4, we can see that the exponential pattern shows a wide gap with the rest of distributions. Interestingly, there is little difference in the maxima ROAC using Weibull and Normal distributions (although responding to different response patterns). None of the definitions lead to a positive ROAC with the uniform distribution; therefore, if the uniform was our response pattern, we maybe should think about either increasing the value of k (number of months in each period) or looking for a new set of

churn definitions.

In the following, we will present only the results using the shifted exponential distribution which is the one that we think more appropriate in our context. In this case, we point out that the sensitivity value $s_p = 0.5$ satisfies condition (11) for all definitions of churn; i.e. this value guarantees that the maximum ROAC is nonnegative for all definitions.

Figure 5 shows the POAC curves as functions of g for all the churn definitions, while Figure 6 shows their ROAC curves. From Figure 4 it can be seen that profits achieved with definitions that use L (D9 to D16) are much lower than profits for definitions without L. Nevertheless, in Figure 5 we can see that, in general, definitions using L provide better ROAC curves, that is, for these definitions there is a higher return for each monetary unit invested in the campaign. The explanation for this behaviour is that definitions using L determine smaller groups of churners than definitions not using L; consequently, although the corresponding campaigns would provide good returns, the total profits would be considerably lower as they would be implemented over less customers. Definitions D1, D6 and their corresponding definitions using the variable L (D9 and D14) show the highest values for the maximum ROAC, while Buckinx & Van den Poel model (definition D3) is not as competitive as the rest.

Having into account these results, if our criterion for choosing a definition was the expected profit, we should use definition D6, while if the criterion was maximizing the expected return, we should use definition D14. This highlights the fact that we need an unified criterion in order to select the best definition, as our multiobjective optimization approach provides with.
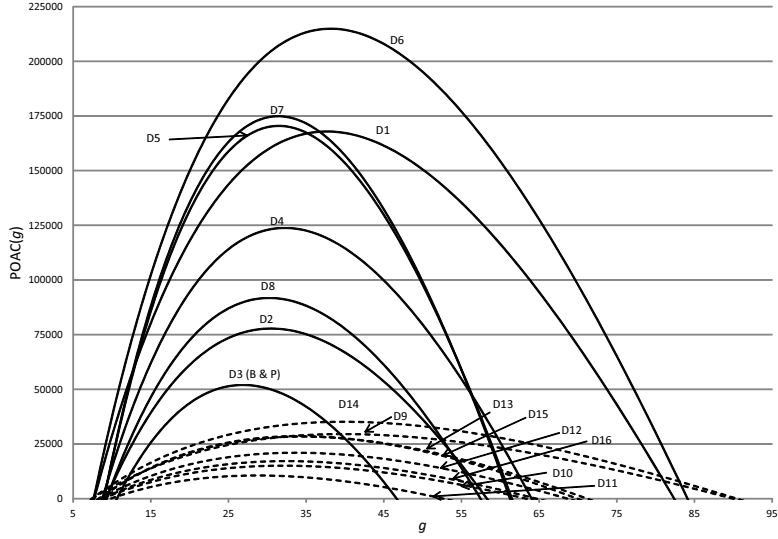
30

Figure 4: POAC of the ad-hoc definitions of churn, for the shifted exponential distribution, sensitivity $s_p = 0.5$.
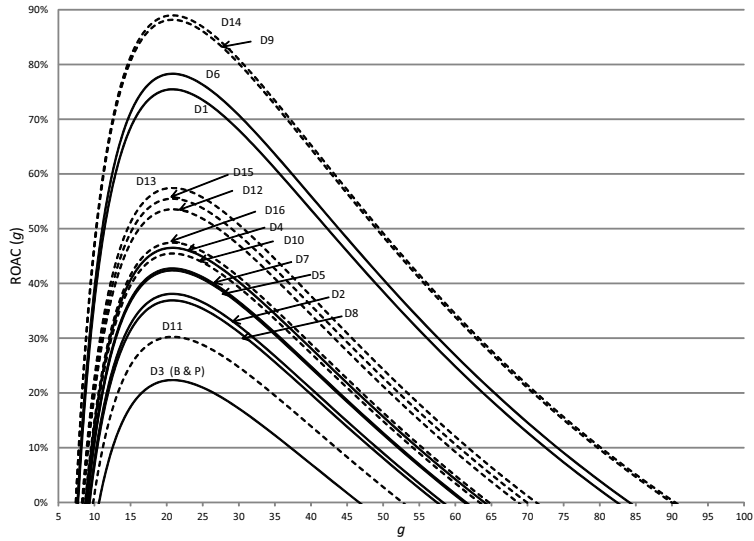

Figure 5: ROAC of the ad-hoc definitions of churn, for the shifted exponential distribution, sensitivity $s_p = 0.5$.

Applying Algorithm 1, we have found the optimal solution for the problem (MP), that is, the optimal churn definition D^* and the action cost g^* required to achieve the best solution. Before showing the results of the optimization process, let us discuss some intermediate results. In Table 2 we show the maximum value of POAC for each definition D_i, together with the values of the marginal action cost $g_{\mathrm{P}}^i = \mathrm{argmax}_g \, \mathrm{POAC}_{D_i}(g)$ where the maxima of POAC are reached and the values of ROAC for each g_{P}^i. These g_{P}^i have to be computed in the step 1 of the algorithm. Let us note that, if the only objective of the company was maximizing POAC, then the best definition of churn would be that one maximizing the column $\mathrm{POAC}(g_{\mathrm{P}}^i)$ in Table 2 when investing the quantity g_{P}^i; in our case, this would be definition $D6$, investing a marginal action cost of 38.15 monetary units per customer, and providing an estimated profit of 214919.43 monetary units. However, this solution would provide a poor value of ROAC (58.80%).

On the other hand, Table 2 shows also the maximum value of ROAC for each definition D_i, together with the values of the marginal action cost $g_{\mathrm{R}}^i = \mathrm{argmax}_g \, \mathrm{ROAC}_{D_i}(g)$ where the maxima of ROAC are reached and the values of POAC for each g_{R}^i. Now, if the company intended maximizing their campaign return (ROAC), the best definition of churn would be that one maximizing the column $\mathrm{ROAC}(g_{\mathrm{R}}^i)$ when investing the quantity g_{R}^i; the results indicate that this would be definition $D14$ when investing 20.80 monetary units per customer, obtaining a ROAC of 88.98% but with a poor estimated profit of 24932.32 monetary units.

We can see also that in order to achieve the maximum POAC it would be necessary to invest much more money per customer in the marketing
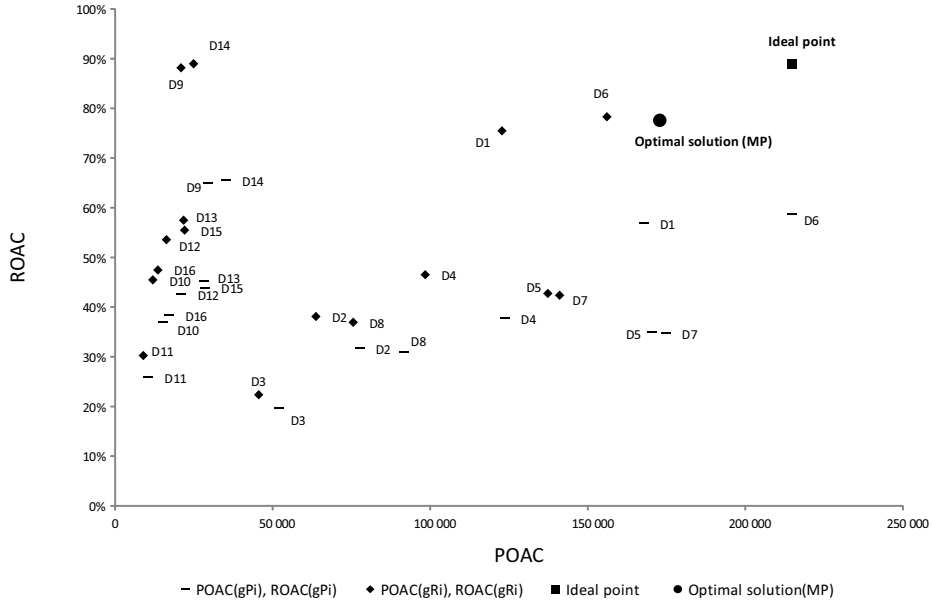
Figure 6: Map of the pairs $(\text{POAC}(g_\text{P}^i),\ \text{ROAC}(g_\text{P}^i))$ and $(\text{POAC}(g_\text{R}^i),\ \text{ROAC}(g_\text{R}^i))$, for each churn definition, sensitivity 0.5, shifted exponential distribution.

campaign than if we intended to achieve the maximum ROAC, whatever the definition we would use.

Figure 6 displays a two-dimensional plot of Table 2. Each pair $(\text{POAC}(g_\text{P}^i),$ $\text{ROAC}(g_\text{P}^i))$ and $(\text{POAC}(g_\text{R}^i),\ \text{ROAC}(g_\text{R}^i))$ have been plotted into the graph. The ideal point and the optimal solution obtained by Algorithm 1 have been drawn also. We remind that the optimal solution minimizes the euclidean distance from the normalized solution to the ideal point. In this case, the ideal point as explained in Section 4 would be the pair $(214919.43\,,\ 88.98\%)$.

In the step 6 of Algorithm 1, we obtained the optima for the subproblems (Pi); each of these optima can be considered as the optimal solution for problem (P) when only a definition of churn is considered, what we call local optima. Table 3 shows all the local optima together with their corresponding

D_i	Maximizing POAC				Maximizing ROAC		
	g_{P}^i	$\mathrm{POAC}(g_{\mathrm{P}}^i)$	$\mathrm{ROAC}(g_{\mathrm{P}}^i)$		g_{R}^i	$\mathrm{POAC}(g_{\mathrm{R}}^i)$	$\mathrm{ROAC}(g_{\mathrm{R}}^i)$
D1	37.67	167917.87	56.99%		20.80	122779.00	75.47%
D2	30.49	77792.66	31.73%		20.80	63749.03	38.11%
D3	26.86	52012.94	19.75%		20.80	45601.23	22.36%
D4	32.26	123828.20	37.72%		20.80	98440.58	46.51%
D5	31.48	170479.00	35.07%		20.80	137324.85	42.74%
D6	**38.15**	**214919.43**	**58.80%**		20.80	156072.44	78.31%
D7	31.40	174917.34	34.81%		20.80	141082.16	42.39%
D8	30.23	91796.77	30.86%		20.80	75570.96	36.92%
D9	39.77	29544.54	64.94%		20.80	20981.57	88.17%
D10	32.05	15093.33	37.00%		20.80	12041.68	45.47%
D11	28.73	10601.64	25.89%		20.80	8967.76	30.25%
D12	33.67	21005.03	42.59%		20.80	16318.56	53.55%
D13	34.42	28397.47	45.23%		20.80	21799.70	57.45%
D14	39.90	35168.95	65.44%		**20.80**	**24932.32**	**88.98%**
D15	34.04	28589.83	43.90%		20.80	22078.58	55.48%
D16	32.45	17180.54	38.39%		20.80	13614.12	47.46%

Table 2: Maxima of POAC and maxima of ROAC (with its corresponding values of g and POAC), for each churn definition, sensitivity 0.5, shifted exponential distribution.

POAC and ROAC. The definitions have been ordered by increasing values of z_i, the objective function of (Pi); thus, the best definitions are those at the top of the table. D6 is the best one, followed by D1, D14 and D9. Notice that in a previous paragraph, we commented that when maximizing POAC, the best option was D6 but when maximizing ROAC the best was D14. Nevertheless, according to the results of Algorithm 1, definition D1 is better than D14, because it achieves a better balance between POAC and ROAC. D3 (corresponding to the Buckinx & Van den Poel definition) and its corresponding definition using L (D11) presented the worst results.

We solved Algorithm 1 using $s_p = 0.5$ but we repeated the analysis also with $s_p = 0.75$, which is a more demanding value but it is a quite realistic level of sensitivity in a reasonably accurate classificator. Table 4 shows the optimal solution (D^*, g^*) of (MP) in both cases. We can see that when the sensitivity grows (and consequently, the classificator accuracy), the estimates for the profit POAC and the return ROAC obtained by the marketing campaign dramatically increase. When $s_p = 0.5$, the optimal solution has achieved good values for both POAC and ROAC, being $\text{POAC}_{D^*}(g^*) = 172873.78$ and $\text{ROAC}_{D^*}(g^*) = 77.60\%$ (we remind that the ideal point was $(214919.43, 88.98\%)$, which is unattainable in any case), while the optimal marginal action cost g^* to get this optimality is just 23.27 monetary units. D6 resulted to be the optimal definition of churn for both levels of sensitivity, for the shifted exponential distribution.

Concerning the segmentation variables used for each churn definition, the best 4 definitions use variable m (average of the k monthly purchase amounts); in fact, the combination of the variable m with a coefficient of

35

Definition	Optimal solution of (Pi)			
D_i	g_i	z_i	POAC(g_i)	ROAC(g_i)
D6	23.27	0.22	172873.78	77.55%
D1	23.63	0.31	137716.82	74.50%
D14	26.27	0.49	30289.23	85.60%
D9	25.53	0.50	24990.77	85.56%
D13	21.44	0.73	22447.76	57.40%
D15	21.42	0.75	22710.97	55.44%
D4	22.08	0.76	104105.05	46.33%
D12	21.25	0.79	16662.68	53.52%
D5	21.88	0.79	144024.77	42.62%
D7	21.85	0.79	147761.10	42.27%
D16	21.14	0.87	13830.56	47.45%
D10	21.09	0.90	12207.13	45.46%
D2	21.78	0.93	66572.85	38.01%
D8	21.84	0.94	79109.53	36.80%
D11	20.98	1.11	9044.75	30.25%
D3	21.47	1.19	46967.90	22.32%

Table 3: Optimal solutions of (Pi) for all the definitions of churn, sensitivity 0.5, shifted exponential distribution.

		Optimal solution of (MP)			
Sensitivity	Definition D^*		g^*	POAC(g^*)	ROAC(g^*)
$s_p = 0.5$	D6		23.27	172873.78	77.60%
$s_p = 0.75$	D6		29.10	440074.28	157.81%

Table 4: Optimal solutions of (MP) for different values of the sensitivity, shifted exponential distribution.

variation variable seems to be the kind of segmentation producing the most promising results. Variable L (length of relationship) is not involved in the best definitions only, so we cannot say that its inclusion is determinant to guarantee a good definition of churn.

7. Conclusions and future work

In this paper, we propose a methodology providing a systematic and objective tool for comparing a set of definitions of churn. This system is based on the expected economic benefits that a marketing campaign based on each definition would report to the company. Specifically, our method finds the definition of churn by maximizing both profit and return of such a campaign, from a multiobjective optimization approach. The proposed algorithm obtains a compromise solution and it achieves very satisfactory results with the real data that we have used for our experiments, as we have shown in Section 6.

As well as determining the optimal churn definition, our method finds out the optimal marginal action cost g that the company should invest per customer when implementing the campaign using the optimal definition. Let us note that this is a critical issue also, as we have seen on the numerical

results that the values of g leading to the best value of the profit POAC and to the best value of ROAC do not match up with the optimal value obtained by the proposed algorithm, which leads always to a more balanced solution than optimizing only either POAC or ROAC.

Among the studied ad-hoc definitions with $k = 5$, the best performance was achieved by churn definitions using the variable segmentation m (average of the k monthly purchase amounts), particularly when combined with a coefficient of variation variable; the Buckinx & Van den Poel definition is not among the best options, according to the proposed multiobjective approach. This conclusion meets also with other distributions than the shifted exponential. In general, the inclusion of the variable L (length of the relationship) as a segmentation variable does not improve specially the economic behaviour of a definition of churn.

In our opinion, the proposed methodology can be a useful decision tool when a company is considering to undertake some kind of marketing actions for preventing the partial defection of customers. By means of its application, the company would be able to select the definition of churn which guarantees the most satisfactory economic results of the marketing campaign, and at the same time to estimate the required total investment in order to reach these results. Of course, this methodology is not restricted to the ad-hoc class of definitions that we used here for numerical purposes.

Actually, one of the ways to extend our approach we would like to explore is the use of longitudinal data for defining churn in non-contractual settings —that is, dealing with behavioural data not as aggregated values but as customer-focused multivariate time series of fixed length. As Chen et al.

(2012) point out in their paper using longitudinal data for churn prediction, there is not much research done in this field.

Acknowledgements

References

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

Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: Partial defection of behaviorally-loyal clients in a non–contractual fmcg retail setting. *European Journal of Operational Research*, *164*, 252–268.

Chen, Z.-Y., Fan, Z.-P., & Sun, M. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, *223*, 461–472.

Dyché, J. (2002). *The CRM Handbook: A Business Guide to Customer Relationship Management*. Boston: Addison-Wesley.

Ehrenberg, A. S. C. (1959). The pattern of consumer purchases. *Applied Statistics*, *8*, 26–41.

Ehrgott, M. (2005). *Multicriteria Optimization*. Berlin: Springer.

Fader, P., & Hardie, B. (2009). Probability models for customer-base analysis. *Journal of Interactive Marketing*, *23*, 61–69.

Fader, P., Hardie, B., & Lee, K. (2005). Counting your customers the easy way: an alternative to the pareto/nbd model. *Marketing Science*, *24*, 275–284.

Glady, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research*, *197*, 402–411.

Hopmann, J., & Thede, A. (2005). Applicability of customer churn forecasts in a non-contractual setting. In D. Baier, & K. D. Wernecke (Eds.), *Innovations in Classification, Data Science, and Information Systems* (pp. 330–337). Berlin: Springer-Verlag.

Jacobson, R. (1987). The validity of ROI as a measure of business performance. *The American Economic Review*, *77*, 470–478.

Keaveney, S. M., & Parthasarathy, M. (2001). Customer switching behavior in online services: an exploratory study of the role of selected attitudinal, behavioral, and demographic factors. *Journal of the Academy of Marketing Science*, *29*, 374–390.

Marler, R., & Arora, J. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, *26*, 269–395.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Massachusetts: Kluwer.

Miguéis, V. L., Camanho, A., & Falcão e Cunha, J. (2013). Customer attrition in retailing: an application of multivariate adaptive regression splines. *Expert Systems with Applications*, *40*, 6225–6232.

Miguéis, V. L., Van den Poel, D., Camanho, A. S., & Falcão e Cunha, J. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, *39*, 11250–11256.

Netzer, O., Lattin, J. M., & Srinivasan, V. (2008). A hidden markov model of customer relationship dynamics. *Marketing Science*, *27*, 185–204.

Phillips, P., & Phillips, J. (2008). *ROI fundamentals: why and when to measure ROI*. San Francisco: John Wiley and Sons.

Reichheld, F. F., & Sasser, W. (1990). Zero defections - quality comes to service. *Harvard business review*, *68*, 105–111.

Reinartz, W. J., & Kumar, V. (2000). On the profitability of long-life customers in a non contractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, *64*, 17–35.

Reinartz, W. J., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, *67*, 77–99.

Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting your customers: who are they and what will they do next? *Management Science*, *33*, 1–24.

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

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. *European Journal of Operational Research*, *218*, 211–229.

Yu, P. (1973). A class of solutions for group decision problems. *Management Science*, *19*, 936–946.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *The Journal of Marketing*, *60*, 31–46.

Zeleny, M. (1973). Compromise pogramming. In J. Cochrane, & M. Zeleny (Eds.), *Multiple Criteria Decision Making* (pp. 262–301). University of South Carolina Press.