CUSTOMER CHURN PREDICTION IN INSURANCE INDUSTRIES: A MULTIPRODUCT APROACH

Mauricio Henao Madrigal
Diego Restrepo Tobón
Henry Laniado

March 4, 2020

Abstract

Customers in the insurance industry usually have multiple products under the same company. Due to the importance of costumers and the increase of quality and satisfaction given in services, churn prediction for multiproduct customers has become important to support customer satisfaction. An estimation was made, using survival models, to find the determinants that most affect multiproduct customer churn. Using data from a Latin American company, the estimation shows that the first product cancellation influences customer churn. The variables that have the strongest influence in churning are the number of cancelled products, the customers portfolio, the claims paid and the distribution channel. On the other hand, the variables that increase the life time of the customer in the company after the first cancellation are the maximum number of products the customer has had and having a health insurance after the first cancellation. This study contributes by identifying the significant variables that influence customer churn in the Latin American insurance industry, and estimating the time a customer will remain in the company after the first cancellation, which will help stakeholders on the achievement of a customer centered strategy.

Key words: Customer Churn; Survival Model; Hazard Rate

JEL Classification: G22, G29, C10 .

1 Introduction

One of the major purposes of companies is to be sustainable over time. To accomplish this purpose, it is important to develop a competitive position in the short and long term and it is then when customer's loyalty starts to play a fundamental role. When customers are not satisfied, they may churn, carrying not only an increase of costs due to the absence of sales revenue but also the cost of finding new customers (such as promotion, discounts, effort to know customer needs and time to build sustainable relationships) (Athanassopoulos, 2000). Furthermore, Verbeke et al. (2011) mentions that customer retention is profitable because attracting a new customer may cost five to six times more than customer retention, and additionally because long-term customers generate higher profits, they are less sensitive to competence, and may provide new referrals by positive word-of-mouth. Consequently, Customer Relationship Management (CRM) has become relevant for companies to study customer churn (Van den Poel and Lariviere, 2004).

Different customer churn prediction models have been studied in markets that are more sensitive to customer care, such as telecommunications (Kim and Yoon, 2004; Lemmens and Croux, 2006; Óskarsdóttir et al., 2017; Verbeke et al., 2011; Ahn et al., 2006; Hung et al., 2006), subscription services (Coussement and Van den Poel, 2008;Jamal and Bucklin, 2006;Burez2007), financial services (Van den Poel and Lariviere, 2004; Athanassopoulos, 2000) and insurance (Bolancé et al., 2016; Lin, 2010; Verhoef et al., 2002; Boucher and Couture-Piché, 2015). Even though these studies provide information related to customers, they focus in a single contract as the unit of analysis while for some markets, such as the insurance market, a contractual relation usually consists of multiple contracts for each risk, such as car, house and health (Guillén et al., 2012). Due to the importance of costumers and the increase role of quality and satisfaction given in services, the strategy of insurance companies, rather than the analysis of a single contract, is centered in a relationship-based marketing view of the customer (Athanassopoulos, 2000). This change has encouraged insurance companies, which manage portfolios of multiple products with long-term contracts, to focus in a more customer-oriented approach (Brockett et al., 2008). The purpose of our study is to investigate the determinants and to predict the risk of churning of multiproduct customers in the Latin America Insurance industry.

Different determinants of insurance multiproduct customer churn have been found in the existing literature. Some literature argues that the main determinants of customer churn are competition, claims and change of home (Brockett et al., 2008). Others, suggest that competition is relevant to predict customer churn only the first year of withdrawal, finding that the most relevant variable is the cancellation of a home contract (Guillén et al., 2012). Currently, de la Llave et al. (2019) found that the location of the customer is an important variable to determine churn. Considering that these studies are centered in a European context, they may be extended to investigate the insurance market in Latin America, which to our knowledge has not been explored. The Latin

America insurance industry manages a diverse portfolio of products such as health, car, home, fire, burial, unemployment, between other products and has had through history a strategy centered in the products, where stakeholders have been interested in questions like how much premium is being sold and canceled per product and not per customer. Even though existing literature has studied customer churn prediction, relevant variables to predict churn may vary according to the different economic conditions of countries, culture, company's specific products and coverages, between others. For this reason, this study contributes to literature by identifying significant variables to predict multiproduct customer churn in the Latin-American insurance market, which has particular economic conditions and products. Additionally, based in these variables, a model is found to predict the time a customer will remain in the company after a first cancellation.

2 Literature Review

Different models have been used to predict customer churn in a series of industries. Van den Poel and Lariviere (2004) describes the literature review on customer churn prediction until 2002. Later studies, where the dependent variable is not time dependent, use techniques such as logistic regression (Kim and Yoon, 2004; Ahn et al., 2006; Athanassopoulos, 2000; Capraro et al., 2003, Bolancé et al., 2016, Burez and Van den Poel, 2007), Bagging and boosting classification techniques (Lemmens and Croux, 2006) and social networks analytics (Óskarsdóttir et al., 2017). Óskarsdóttir et al. (2017) describes a wide literature review on social networks analytics models to predict customer churn. Other studies, where the dependent variable is time dependent, use techniques such as queuing theory (Boucher and Couture-Piché, 2015), and hazard models (Van den Poel and Lariviere, 2004; Jamal and Bucklin, 2006). Furthermore, a comparison between logistic regression, conditional trees, neural networks, support vector machine and random forest models to predict churn have also been studied (Coussement and Van den Poel, 2008). Some of these studies find that all models have a similar result (Bolancé et al., 2016; Hung et al., 2006) and others establish that logistic regression models performs as well, if not better, than the other models (Burez and Van den Poel, 2007).
In the context of companies that handle multiple products, such as the insurance industry, non-linear neural fuzzy network (Lin, 2010), spatial probit models (de la Llave et al., 2019) and hazard models (Brockett et al., 2008; Guillén et al., 2012) have been used to predict customer churn. They find that some of the variables that influence customer churn, which vary depending on each case of study, include competition, claims, a change of home and the home contract.

3 Data and Methodology

3.1 Data

In most of the cases, a signal to know if multiproduct customers have an intention to churn is found when the first product is cancelled (Brockett et al., 2008). therefore, this study focuses on customers that have had their first cancellation. The information used is from a Latin American insurance company. The data consist on 590 thousand multiproduct customers that cancelled the first product between January 2010 and October 2019. The portfolio of the company involves a combination of any of these products: life, vehicles, personal accidents, burial, fire, unemployment, health and home. Figure 1, describes the frame of time of the data. The residual time represents the lifetime of the customer after the first cancellation. It is worth noting that the lifetime for the customers that have not churn at the end of this study, are measured until the end of this study (october 2019) and are marked as censored customers, since it is not known exactly when they will churn.

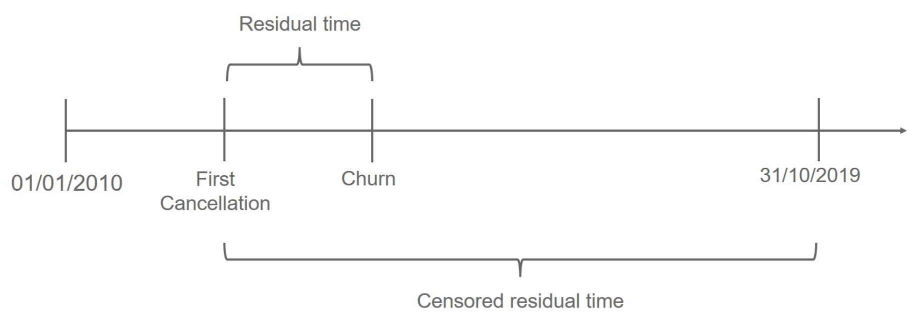

Figure 1: customer time description

Table 1 describes the variables used in the analysis. The first variable, *reentries*, counts how many times a customer has returned to the company. A customer may leave the company and return, according to different factors such as better prices or better plans of competitors, unemployment or the sale of their assets. *Topsol* specifies the maximum number of active products that the customer has had, while *numberofsolfirstcancel* specifies the number of products cancelled in the first cancellation. The area of the country where the products were sold (*state*) and the salling channel (*saleschannel*) are as well included in this study.

Another variable that may affect the customer's decision to churn is their current insurance portfolio. This study is focused in a portfolio of 11 products: vehicle, four types of life products (*lilfe1, life2, life3, life4*), personal accidents (*pa*), burial, two types of health products (*health1 and health2*), home and fire. The variables named as *autos0, life10, life20, life30, life40, pa0, burial0, health10, health20, home0, fire0* are dummy variables that state whether the customers had or did not have the products before the first cancellation. On the other hand, *autos1, life11, life21, life31, life41, pa1, burial1, health11, health21, home1, fire1* state whether the customers had or did no have the products after the first cancellation. Finally, the age and gender are included as demographic variables.

VARIABLES
Number of times the customer has churned and cameback (REENTRIES)
Top number of products the customer has had (TOPSOL)
Number of first cancelled products (NUMBEROFSOLFIRSTCANCEL)
Gender of the customer (SEX)
1 if the customer had vehicle product before the first cancellation, 0 otherwise (AUTOS0)
1 if the customer had vehicle product after the first cancellation, 0 otherwise (AUTOS1)
1 if the customer had life1 product before the first cancellation, 0 otherwise (LIFE10)
1 if the customer had life1 product after the first cancellation, 0 otherwise (LIFE11)
1 if the customer had life2 product before the first cancellation, 0 otherwise (LIFE20)
1 if the customer had life2 product after the first cancellation, 0 otherwise (LIFE21)
1 if the customer had personal accidents product before the first cancellation, 0 otherwise (PA0)
1 if the customer had personal accidents product after the first cancellation, 0 otherwise (PA1)
1 if the customer had burial product before the first cancellation, 0 otherwise (BURIAL0)
1 if the customer had fire product before the first cancellation, 0 otherwise (FIRE0)
1 if the customer had fire product after the first cancellation, 0 otherwise (FIRE1)
1 if the customer had health1 product before the first cancellation, 0 otherwise (HEALTH10)
1 if the customer had health1 product after the first cancellation, 0 otherwise (HEALTH11)
1 if the customer had health2 product before the first cancellation, 0 otherwise (HEALTH20)
1 if the customer had health2 product after the first cancellation, 0 otherwise (HEALTH21)
1 if the customer had home product before the first cancellation, 0 otherwise (HOME0)
1 if the customer had home product after the first cancellation, 0 otherwise (HOME1)
1 if the customer had life3 product before the first cancellation, 0 otherwise (LIFE30)
1 if the customer had life3 product after the first cancellation, 0 otherwise (LIFE31)
1 if the customer had life4 product before the first cancellation, 0 otherwise (LIFE40)
1 if the customer had life4 product after the first cancellation, 0 otherwise (LIFE41)
1 if the the customer has emited a product by the comercial channel 1, 0 otherwise (SALECHANNEL1)
1 if the the customer has emited a product by the comercial channel 2, 0 otherwise (SALECHANNEL2)
1 if the the customer has emited a product by the comercial channel 3, 0 otherwise (SALECHANNEL3)
1 if the the customer has emited a product by the comercial channel 4, 0 otherwise (SALECHANNEL4)
1 if the the customer has emited a product by the comercial channel 5, 0 otherwise (SALECHANNEL5)
1 if the the customer has emited a product by the comercial channel 6, 0 otherwise (SALECHANNEL6)
1 if the the customer has emited a product by the comercial channel 7, 0 otherwise (SALECHANNEL7)
1 if the the customer has emited a product by the comercial channel 8, 0 otherwise (SALECHANNEL8)
The total amount of claims paid to the customer (CLAIMS)
1 if the customer has emited products in the state 1 of the country in study (STATE1)
1 if the customer has emited products in the state 2 of the country in study (STATE2)
1 if the customer has emited products in the state 3 of the country in study (STATE3)
1 if the customer has emited products in the state 4 of the country in study (STATE4)
1 if the customer has emited products in the state 5 of the country in study (STATE5)
1 if the customer has emited products in the state 6 of the country in study (STATE6)
The age of the customer (AGE)

Table 1: Variables of the data set

Observing the customer's portfolio before and after the first cancellation, 68 thousand customers cancelled all their products in the first cancellation and the 522 thousand left have the number of products shown in figure 2. 99% of the customers had 2 or 3 products before they cancelled the first product and the same 99% stayed with 1 or 2 products after. This study is centered in the 522 thousand customers that have time to be retain.

The first cancelled products for the customers that have time to be retain are described in table 2. Even though life1 and life2 are the products that are

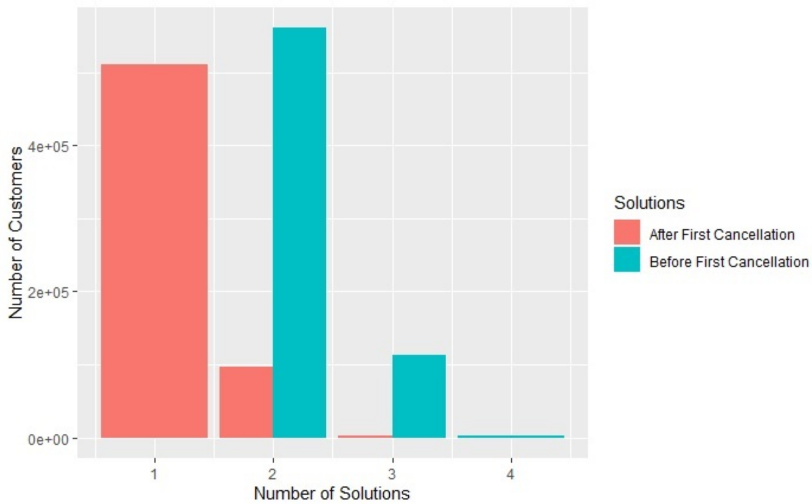

Figure 2: Number of products per customer

cancelled the most, unemployment, life3 and fire have the highest percentage of cancellations compared to the total active products before cancellation.

product	CANCELLED	ACTIVE	%
UNEMPLOYMENT	1,141	1,348	85%
LIFE 3	15,695	24,093	65%
FIRE	10,417	16,477	63%
LIFE 2	231,274	413,980	56%
PERSONAL ACCIDENTS	37,193	68,353	54%
AUTOMOBILES	59,498	122,674	49%
LIFE 4	890	2,185	41%
HOME	2,989	7,459	40%
LIFE 1	145,809	369,142	39%
BURIAL	24,682	80,659	31%
HEALTH 1	5,334	17,757	30%
HEALTH 2	1,910	7,462	26%

Table 2: First cancelled products

Furthermore, from these 522 thousand customers that did not cancelled all their products, 514 thousand customers belong to a portfolio of products where more than 50% of the customers have churned, making it the most important group to be retained. Table 3, 4, 5 describe these portfolios in accordance to the number of customers and the combination of products before and after the first cancellation. It is observed that when customers with life1 and life2 cancel one of the products, 57% remain with life1, 38% with life2 and only 5% cancel both. A possible interpretation of this result is that for customers can be redundant to have these tow life products. The same can be concluded when analyzing the portfolios with 3 products involving life1 and life2, finding that 53% of the customers with funeral, life1 and life2 remain with burial and life1; and 64% of the customers with vehicle, life1 and life2 remain with vehicle and either life1 or life2. Finally, it is observed that 79% of the customers with a portfolio of personal accidents and funeral remain with personal accidents, concluding that the funeral product is not generating value for customers with this portfolio.

	LIFE 3	UNEMPLOYMENT	LIFE 1	LIFE 2	PERSONAL ACCIDENTS	BURIAL	LIFE 4	VEHICLES	FIRE	HOME	HEALTH 2	HEALTH 1	TOTAL
LIFE2, LIFE 1	-	-	57%	38%	-	-	-	-	-	-	-	-	208,239
VEHICLES, LIFE2	-	-	-	43%	-	-	-	53%	-	-	-	-	45,444
VEHICLES, LIFE1	-	-	54%	-	-	-	-	38%	-	-	-	-	29,382
PERSONAL ACCIDENTS, LIFE 2	-	-	-	42%	32%	-	-	-	-	-	-	-	27,659
BURIAL, LIFE2	-	-	-	29%	-	66%	-	-	-	-	-	-	26,893
PERSONAL ACCIDENTS, LIFE 1	-	-	51%	-	35%	-	-	-	-	-	-	-	17,121
LIFE 3, LIFE 1	19%	-	40%	-	-	-	-	-	-	-	-	-	16,120
BURIAL, LIFE1	-	-	12%	-	-	18%	-	-	-	-	-	-	9,512
PERSONAL ACCIDENTS, VEHICLES	-	-	-	-	46%	-	-	47%	-	-	-	-	8,371
LIFE 3, LIFE 2	38%	-	-	57%	-	-	-	-	-	-	-	-	5,735
HEALTH1, LIFE 1	-	-	42%	-	-	-	-	-	-	-	-	46%	5,577
HEALTH1, LIFE 2	-	-	-	27%	-	-	-	-	-	-	-	70%	3,981
HEALTH2, LIFE 2	-	-	-	20%	-	-	-	-	-	-	60%	-	3,845
FIRE, LIFE2	-	-	-	61%	-	-	-	-	35%	-	-	-	3,551
FIRE, LIFE1	-	-	49%	-	-	-	-	-	38%	-	-	-	3,469
HOME, LIFE2	-	-	-	27%	-	-	-	-	-	70%	-	-	3,240
LIFE4, LIFE2	-	-	-	33%	-	-	48%	-	-	-	-	-	2,185
HOME, LIFE1	-	-	49%	-	-	-	-	-	-	40%	-	-	2,177
VEHICLES, HOME	-	-	-	-	-	-	-	42%	-	54%	-	-	2,042
PERSONAL ACCIDENTS, HEALTH 2	-	-	-	-	24%	-	-	-	-	-	74%	-	1,871
VEHICLES, FIRE	-	-	-	-	-	-	-	43%	54%	-	-	-	1,576
PERSONAL ACCIDENTS, BURIAL	-	-	-	-	79%	16%	-	-	-	-	-	-	1,390
UNEMPLOYMENT, LIFE2	-	12%	-	68%	-	-	-	-	-	-	-	-	1,348
VEHICLES, HEALTH1	-	-	-	-	-	-	-	21%	-	-	-	-	1,007
VEHICLES, HEALTH2	-	-	-	-	-	-	-	26%	-	-	-	-	475

Table 3: Percentage of customers per portfolio of 2 products

	LIFE 1	LIFE 2	BURIAL	LIFE 3	PERSONAL ACCIDENTS	FIRE	VEHICLES	HEALTH 1	TOTAL
BURIAL, LIFE2, LIFE1	1%	20%	1%	-	-	-	-	-	41,322
VEHICLES, LIFE2, LIFE 1	3%	3%	-	-	-	-	2%	-	16,795
PERSONAL ACCIDENTS, LIFE 2, LIFE 1	9%	3%	-	-	1%	-	-	-	6,136
FIRE, LIFE2, LIFE 1	24%	2%	-	-	-	0%	-	-	6,109
PERSONAL ACCIDENTS, VEHICLES, LIFE 2	-	3%	-	-	3%	-	7%	-	4,272
LIFE3, LIFE2, LIFE1	3%	20%	-	1%	-	-	-	-	2,238
VEHICLES, FIRE, LIFE2	-	2%	-	-	-	-	11%	-	1,748
PERSONAL ACCIDENTS, VEHICLES, LIFE 1	3%	-	-	-	3%	-	2%	-	1,533
VEHICLES,HEALTH 1, LIFE 2	-	-	-	-	-	-	1%	-	238
VEHICLES,HEALTH 1, LIFE 1	-	-	-	-	-	-	-	1%	28

Table 4: Percentage of customers per portfolio of 3 products

	BURIAL, LIFE1	LIFE2, LIFE1	LIFE3, LIFE2	LIFE3, LIFE1	BURIAL, LIFE2	VEHICLES, LIFE2	FIRE, LIFE2	PERSONAL ACCIDENTS, LIFE 1	PERSONAL ACCIDENTS, LIFE 2	PERSONAL ACCIDENTS, VEHICLES	VEHICLES, LIFE1	FIRE, LIFE1	VEHICLES, FIRE	TOTAL
BURIAL, LIFE2, LIFE 1	53%	6%	-	-	18%	-	-	-	-	-	-	-	-	41,322
VEHICLES, LIFE2, LIFE 1	-	29%	-	-	-	27%	-	-	-	-	37%	-	-	16,795
PERSONAL ACCIDENTS, LIFE 2, LIFE 1	-	37%	-	-	-	-	-	24%	24%	-	-	-	-	6,136
FIRE, LIFE2, LIFE 1	-	47%	-	-	-	-	19%	-	-	-	-	8%	-	6,109
PERSONAL ACCIDENTS, VEHICLES, LIFE 2	-	-	-	-	-	32%	-	-	34%	21%	-	-	-	4,272
LIFE 3, LIFE 2, LIFE 1	-	31%	11%	31%	-	-	-	-	-	-	-	-	-	2,238
VEHICLES, FIRE, LIFE 2	-	-	-	-	-	47%	24%	-	-	-	-	-	15%	1,748
PERSONAL ACCIDENTS, VEHICLES, LIFE 1	-	-	-	-	-	-	-	34%	-	20%	38%	-	-	1,533
VEHICLES, HEALTH 1, LIFE 2	-	-	-	-	-	13%	-	-	-	-	-	-	-	238
VEHICLES, HEALTH 1, LIFE 1	-	-	-	-	-	-	-	-	-	-	-	-	-	28

Table 5: Percentage of customers per portfolio of 3 products

3.2 Methodology

Survival analysis is used to analyze and predict the time that customers might take between the cancellation of the first product and their churn. Defining the random variable T as the duration of the customer in the company from the first cancellation to the cancellation of the last product; the survival function $S(t)$, the hazard function $\lambda(t)$ and the cumulative hazard function $H(t)$ can be derived by the density function $f(t)$ or by its cumulative distribution function (CDF) $F(t) = P(T \leq t)$ (Cleves et al., 2008):

The survival function, which measures the probability of churning beyond a time t is defined as:

$$S(t) = 1 - F(t) = P(T > t) \tag{1}$$

The hazard function, which measures how long customers stay with the company since the first cancellation until they churn, is defined as:

$$\lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt} = \frac{f(t)}{S(t)} \tag{2}$$

This hazard function is defined as the probability of a customer to churn at time t given that the customer did not cancel all the products up to time t. It is worth noting that this duration is right-censored because there are customers that have not churn at the time that this analysis was performed.

The cumulative hazard function $H(t)$ defines the total risk accumulated to time t and is useful due to its easy estimation compared to the hazard function (Therneau and Grambsch, 2013):

$$H(t) = \int_0^t \lambda(s) ds \tag{3}$$

$S(t), F(t)$ and $f(t)$ can be written in terms of the cumulative hazard:

$$S(t) = e^{-H(t)} \tag{4}$$

$$F(t) = 1 - e^{-H(t)} \tag{5}$$

$$f(t) = h(t)e^{-H(t)} \tag{6}$$

Parametric, semi-parametric and non-parametric models are used in survival analysis. Semi-parametric and non-parametric models have the advantage of not assuming a specific form of the survival function (Kalbfleisch and Prentice, 2011). In this study we use the non parametric methodology of Nelson and Aalen and Kaplan Meier.

3.2.1 Homogeneous Population:

When the failure time does not depend on exploratory variables, survival distributions can be used to model the customer's churn. Even though our data depend on explanatory covariates, these survival distributions are described because they form part of the models involving covariates.

Nelson Aalen Estimator (Nelson, 1969):

To estimate the cumulative hazard function, Nelson Aalen estimator is the most common non-parametric estimator for homogeneous variables. It has good properties for small samples and is defined as:

$$\hat{H}(t) = \sum_{j|t_j \leq t} \frac{d_j}{n_j} \tag{7}$$

Where n_j are the customers at risk at time t_j and d_j is the number of churned customers at time t_j.

Kaplan Meier Estimator (Kaplan and Meier, 1958):

The Kaplan Meier estimator is also a non-parametric method and estimates the survivor function by:

$$\hat{S}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j} \tag{8}$$

Where n_j are the customers at risk at time t_j and d_j is the number of churned customers at time t_j.

The cumultive hazard function can be found by:

$$H(t) = -lnS(t) \tag{9}$$

3.2.2 Regression Models:

The failure time of customers in a company can not be treated only with the distributions described above because the data is not homogeneous and it depends on explanatory variables. To consider these aspects, more generalized regression models are used.

Two assumptions are made in this type of models (Brockett et al., 2008):

- A reference distribution of the time a multiproduct customer takes to churn and
- The relative risk of the customers to churn changes completely this distribution due to their individual covariates.

Cox Model:

This model is a semi-parametric method that does not assume a specific distribution. Cox, 1972 establishes the hazard function as a reference hazard function $\lambda_0(t)$ multiplied by the covariate dependent factor:

$$\lambda(t|z) = \lambda_0(t)exp(Z^{'}\beta) \tag{10}$$

Where $\lambda_0(t)$ is an unspicified reference hazard function. When $\lambda_0(t) = \frac{\sigma}{\eta}\left(\frac{t}{\eta}\right)^{\sigma-1}$ the Cox model is the Weibull regression model.

The conditional density is:

$$f(t|z) = \lambda_0(t)exp(Z'\beta)exp\left[-exp(Z'\beta)\int_0^t \lambda_0(u)du\right] \tag{11}$$

The conditional survival function is:

$$S(t;z) = [S_0(t)]^{exp(Z'\beta)} \tag{12}$$

Where,

$$S_o(t) = exp\left[\int_0^t \lambda_0(u)du\right] \tag{13}$$

To estimate the parameters, there is not need to specify $\lambda_0(t)$. The estimates are obtained by maximum likelihood.

4 Results

This study focuses on 522 thousand customers that did not cancel all their products at once, which gives the company some time to react and to make and effort to retain them before they churn. 38% of these customers have not left the company at the end of this study, making them censored customers. The average age of the customers is 41 years old (standard deviation of 12) and 47% of the customers are female. The time the customers last in the company after the first product is cancelled has a median of 2 years (by Kaplan Meier estimator). As shown in figure 4, the cumulative hazard function is increasing, implying that as time goes on the customer has a greater probability to churn. Looking at the residual time in relation to the products before and after the first cancellation, as shown in table 6, is observed that customers that cancel one product, stay on average, more time in the company than those whose cancel

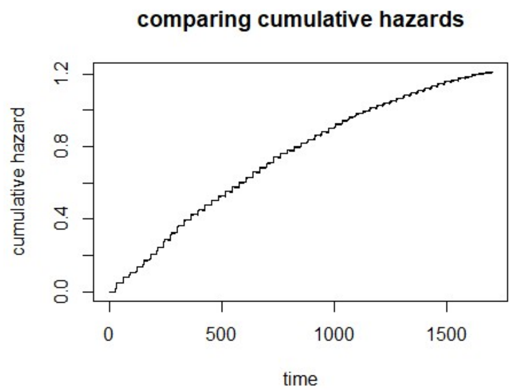

Figure 3: Hazard of the time since first cancellation to churn

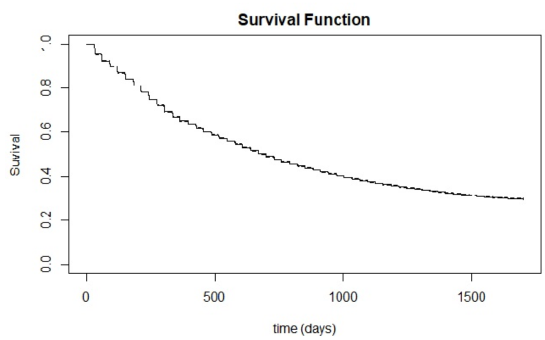

Figure 4: Survival Function of the time since first cancellation to churn

State Before	State After	Average (days)
3	2	1,394
3	1	795
2	1	1,110

Table 6: Residual time after first lapse

two products. Additionally, the customers that after having 3 products cancel one, stay longer on average.

Cox Regression

The the expected customer remaining time between the first cancelled product and the withdrawal is better analyzed using the cox regression. Table 7 describes the estimated parameters. The model has a good fit with a likelihood ratio test of 111691 being chi-squared distributed with 41 degrees of freedom ($p < 0.001$), allowing us to conclude a significant effect of the covariates on the hazard. The model confirms that the company has time to take action to retain

Parameter	Estimate	Standard Error	z	p-Value
REENTRIES	0.038	0.0040	9.126	0.0000
TOPSOL	- 1.521	0.0110	- 135.247	0.0000
NUMBEROFSOLFIRSTCANCEL	0.525	0.0120	44.468	0.0000
SEXM	0.094	0.0040	26.391	0.0000
AUTOS01	0.597	0.0160	36.752	0.0000
AUTOS11	0.281	0.0130	21.522	0.0000
LIFE101	0.991	0.0150	64.594	0.0000
LIFE111	0.067	0.0100	6.439	0.0000
LIFE201	0.787	0.0150	50.944	0.0000
LIFE211	0.229	0.0100	23.369	0.0000
PA01	0.794	0.0170	47.176	0.0000
PA11	0.345	0.0140	24.257	0.0000
BURIAL01	1.058	0.0130	82.935	0.0000
FIRE01	0.504	0.0200	25.483	0.0000
FIRE11	0.473	0.0250	19.296	0.0000
HEALTH101	0.582	0.0290	20.291	0.0000
HEALTH111	- 0.440	0.0330	- 13.174	0.0000
HEALTH201	0.687	0.0390	17.481	0.0000
HEALTH211	- 0.334	0.0450	- 7.464	0.0000
HOME01	0.556	0.0330	16.696	0.0000
HOME11	0.191	0.0390	4.855	0.0000
LIFE301	1.247	0.0190	66.572	0.0000
LIFE311	0.496	0.0210	24.127	0.0000
LIFE401	0.853	0.0470	18.217	0.0000
LIFE411	0.478	0.0560	8.567	0.0000
SALECHANNEL11	0.081	0.0070	11.516	0.0000
SALECHANNEL21	- 0.471	0.0070	- 63.271	0.0000
SALECHANNEL41	- 0.696	0.0710	- 9.745	0.0000
SALECHANNEL51	- 0.483	0.0620	- 7.810	0.0000
SALECHANNEL61	- 0.573	0.0140	- 40.091	0.0000
SALECHANNEL71	- 0.329	0.0330	- 10.094	0.0000
SALECHANNEL81	- 0.247	0.0060	- 42.077	0.0000
SALECHANNEL91	- 0.079	0.0420	- 1.867	0.0620
CLAIMS	21.425	1.3480	15.894	0.0000
STATE11	0.360	0.0080	42.999	0.0000
STATE21	- 0.142	0.0060	- 23.651	0.0000
STATE31	- 0.155	0.0080	- 20.165	0.0000
STATE41	- 0.246	0.0090	- 28.370	0.0000
STATE51	- 0.232	0.0090	- 26.176	0.0000
STATE61	- 0.079	0.0110	- 7.357	0.0000
AGE	- 0.010	0.0000	- 68.637	0.0000

Table 7: Cox Regression

the customers before they churn. Additionally, the concordance of the model is 0.66, indicating that the model identifies correctly the order of the survival times of each pair of customers 66% of the time.

The estimates in table 7 show that the variable that the most increases the risk to churn is the number of products cancelled in the first cancellation (*numberofsolfirstcancel*). On the other hand, the variable that most reduces the risk to churn is the maximum number of products that a customer has had (*topsol*). This result supports the analysis made in table 6. The first cancellation and the portfolio of the customers before and after the first cancellation affects the lifetime of the customer in the company which is consistent with Brockett et al. (2008). Analyzing the portfolio of products, it is observed that having any of the products before the first cancellation, increase the probability to churn. Moreover, it can be concluded that customers with life3 before the first cancellation have a higher risk than the others, followed by burial and life 1. It is worth nothing that the risk to churn is higher having any of these products before than after the first cancellation. On the other side, having health1 or health2 after the first cancellation, decreases the probability to churn. Finally, the number of times a customer has come back to the company (*reentries*), the paid claims (normalized variable) and the male gender increases the probability of a customer to churn while the sale's channel and the states decrease the risk to churn, except for the sale's channel1 and state1.

Prediction according to the first products cancelled

To analyze how the first cancellation affects the residual time of the customers, we used a customer with the same characteristics except for the first product cancelled: a male customer of 43 years old, that has never been in the company before, with issued products from the sale's channel 3, in the states 1 and 2, with no claims and that has had a maximum of 2 products. Figure 5 shows the survival estimates according to the first policy being cancelled if the customer has health, life1 and life2. It is observed that when the customer cancels health, the residual time is less compared to the same customer cancelling individual life or group life. This means that having health after the first cancellation retains the customer for longer. Likewise, it is also observed that when life 2 is the first cancelled product, the customer stays for a longer time in the company. This means that life 2 does not retain the customer. In general, these results hold for all the portfolios involving health and life2, allowing us to conclude that health is a product that holds the customer from churning while life 2 shortens the time of the customer in the company.

5 Conclusions

It was found that the churn median time after the first cancellation of the customers is 2 years, which for the insurance industry and for the type of policies, is not a long time. Still, 2 years is a good time for the companies to retain the customers. This result highlights the importance of studying, monitoring and taking action on customers at risk of churning, in order to keep them loyalty. Furthermore, it is found that not only the number of products, but also the type of product cancelled affects the risk of a customer to churn. Specifically, health insurance has a strong effect on the lifetime of the customer, finding that customers that cancel health in the first cancellation have a higher risk to leave the company than the ones that cancel other products. In contrast, life2 is a product that does not retain the customer due to the evidence found that there is a strong risk to leave the company in the near future if the customer stays with this product after the first cancellation. Additionally, the model shows that customers with products issued by channel1 have a higher risk to churn, compared to other channels; it is important for the company to evaluate the portfolios of products and the characteristics presented in this channel to identify flaws and to implement changes able to prevent this risk.

Even though there are external factors that companies do not have an easy

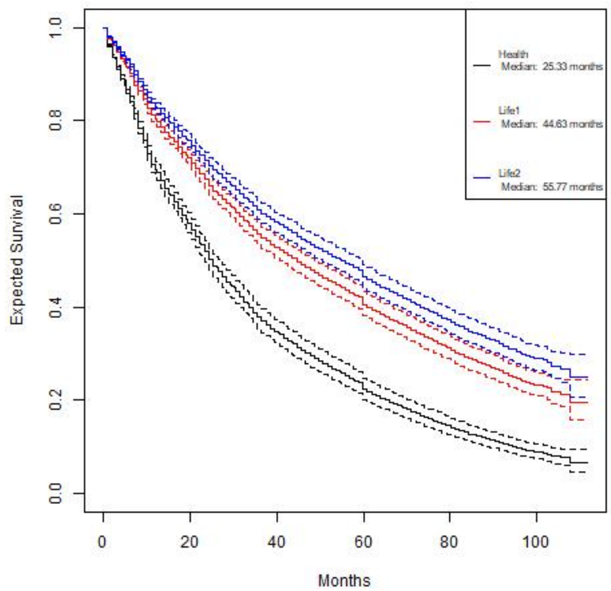

Figure 5: Survival function depending on the first cancelled product

access to, such as competitors, customer unemployment, among others, internal factors can help identify customer churn before it happens. The results of this analysis will strength customer's loyalty, contributing the Latin American insurance companies to anticipate customer churn which, to our knowledge, has not been explored for a multiproduct approach.

References

J.-H. Ahn, S.-P. Han, and Y.-S. Lee. Customer churn analysis: Churn determinants and mediation effects of partial defection in the korean mobile telecommunications service industry. *Telecommunications policy*, 30(10-11): 552–568, 2006.

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

C. Bolancé, M. Guillen, and A. E. Padilla-Barreto. Predicting probability of customer churn in insurance. In *International Conference on Modeling and Simulation in Engineering, Economics and Management*, pages 82–91. Springer, 2016.

J.-P. Boucher and G. Couture-Piché. Modeling the number of insureds' cars using queuing theory. *Insurance: Mathematics and Economics*, 64:67–76, 2015.

P. L. Brockett, L. L. Golden, M. Guillen, J. P. Nielsen, J. Parner, and A. M. Perez-Marin. Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection? *Journal of Risk and Insurance*, 75(3):713–737, 2008.

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

A. J. Capraro, S. Broniarczyk, and R. K. Srivastava. Factors influencing the likelihood of customer defection: the role of consumer knowledge. *Journal of the Academy of Marketing Science*, 31(2):164–175, 2003.

M. Cleves, W. Gould, W. W. Gould, R. Gutierrez, and Y. Marchenko. *An introduction to survival analysis using Stata*. Stata press, 2008.

K. Coussement and D. Van den Poel. 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, 2008.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

M. Á. de la Llave, F. A. López, and A. Angulo. The impact of geographical factors on churn prediction: an application to an insurance company in madrid's urban area. *Scandinavian Actuarial Journal*, 2019(3):188–203, 2019.

M. Guillén, J. P. Nielsen, T. H. Scheike, and A. M. Pérez-Marín. Time-varying effects in the analysis of customer loyalty: A case study in insurance. *Expert systems with Applications*, 39(3):3551–3558, 2012.

S.-Y. Hung, D. C. Yen, and H.-Y. Wang. Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3):515–524, 2006.

Z. Jamal and R. E. Bucklin. Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. *Journal of Interactive Marketing*, 20(3-4):16–29, 2006.

J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

H.-S. Kim and C.-H. Yoon. Determinants of subscriber churn and customer loyalty in the korean mobile telephony market. *Telecommunications policy*, 28(9-10):751–765, 2004.

A. Lemmens and C. Croux. Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2):276–286, 2006.

W.-B. Lin. Service failure and consumer switching behaviors: Evidence from the insurance industry. *Expert Systems with Applications*, 37(4):3209–3218, 2010.

W. Nelson. Hazard plotting for incomplete failure data. *Journal of Quality Technology*, 1(1):27–52, 1969.

M. Óskarsdóttir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen. Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Systems with Applications*, 85:204–220, 2017.

T. M. Therneau and P. M. Grambsch. *Modeling survival data: extending the Cox model*. Springer Science & Business Media, 2013.

D. Van den Poel and B. Lariviere. Customer attrition analysis for financial services using proportional hazard models. *European journal of operational research*, 157(1):196–217, 2004.

W. Verbeke, D. Martens, C. Mues, and B. Baesens. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3):2354–2364, 2011. ISSN 09574174. doi: 10.1016/j.eswa.2010.08.023. URL `http://dx.doi.org/10.1016/j.eswa.2010.08.023`.

P. C. Verhoef, P. H. Franses, and J. C. Hoekstra. The effect of relational constructs on customer referrals and number of servi... (1998), 2002.