“A logistic regression approach to estimating customer profit loss due to lapses in insurance”

AUTHORS	Montserrat Guillén Estany Ana Maria Pérez-Marín Manuela Alcañiz
ARTICLE INFO	Montserrat Guillén Estany, Ana Maria Pérez-Marín and Manuela Alcañiz (2011). A logistic regression approach to estimating customer profit loss due to lapses in insurance. *Insurance Markets and Companies*, *2*(2)
RELEASED ON	Monday, 26 December 2011
JOURNAL	"Insurance Markets and Companies"
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
0	**0**	**0**

Montserrat Guillén Estany (Spain), Ana María Pérez-Marín (Spain), Manuela Alcañiz Zanón (Spain)

A logistic regression approach to estimating customer profit loss due to lapses in insurance

Abstract

This article focuses on business risk management in the insurance industry. It proposes a methodology for estimating the profit loss caused by each customer in the portfolio due to policy cancellation. Using data from a European insurance company, customer behavior over time is analyzed in order to estimate the probability of policy cancellation and the resulting potential profit loss due to cancellation. Customers may have up to two different lines of business contracts: motor insurance and other diverse insurance, such as, home contents, life or accident insurance. The paper outlines the implications for understanding customer cancellation behavior as the core of business risk management.

Keywords: policy cancellation, customer loyalty, profit loss, customer behavior.

Introduction

Under the framework promulgated by the solvency regulation such as Solvency II in Europe, insurance companies face an increasing number of challenges which require a change in the way business risk is currently measured and managed. Risk management under this perspective requires new approaches to all components of the insurance activity that may cause losses to the company. We will study policy cancellations and we will propose a method to measure potential profit loss that follows from customers that do not renew their insurance contracts. A real case study involving two lines of business will be described and we will outline qualitative implications.

Insurance companies now operate in a more competitive environment than they used to in the past and customers easily switch from one insurer to another. Cancellations and lapses[1] have become one of the factors influencing the level of risk an insurance company and its position in the market. The proliferating extension of the Internet played an instrumental role in reducing information-gathering costs for customers who wish to change insurers. Now, the central problem for insurance companies is not only to create and launch new products for the market, but additionally to achieve commercial success by retaining customers. As a consequence of this growing interest in increasing customer loyalty, the insurance business is no longer only product-oriented, but also customer-oriented.

The fluctuations in volumes and margins due to ongoing competition are a source of risk for the company which is called *business risk* (Nakada et al., 1999 and Dhaene et al., 2006). This type of risk, which is increasingly integral to a company's operational risk[2], reflects the reality of the market's impact on the stability of the company.

Business risk management involves overcoming a number of difficulties, including the measurement of business risk itself. Numerous factors influence business risk, yet previous literature has not provided a way to quantify and assess these factors. This could, in turn, clarify specific actions to help protect companies against business risk.

Despite the difficulties, business risk management provides benefits to companies. Companies increase overall results when they invest in retaining customers generating high profits, e.g., those who pay premiums for long periods without claims, rather than those with a bad claim history, contributing to reduce overall profits. Additionally, controlling business risk also contributes to the stability and solvency of a company by preventing losses caused by a potential decrease of its market share.

Typically, insurance activity is managed separately for each line of business. The most aberrant implication of this classical managerial mistake is that different policies in the same company owned by the same customer are often managed separately. Thus, events affecting one particular policy (claims, premium increases, etc.) are evaluated without consideration of concurrent events plausibly affecting other policies within the company. With such a one-dimensional perspective, the possible relationship between these events and the customer's actions is ignored. As a result, neither the behavior of the customer nor the relationship of the company and the customer (which we call *insurance relationship*) are fully understood.

In this article, we focus on business risk management in the insurance industry and we account for the multidimensionality of the insurance relationship by taking all events affecting one policy into consideration when understanding what happens to

We thank the Spanish Ministry of Science grants ECO2010-21787-C03-01 and ECO2008-01223.

[1] Cancellations refer to policies that are not renewed. Equivalently, we also use the word *lapse* to indicate that an insurance contract is finished.

[2] We consider the same classification of risks used by Nakada et al. (1999) and Dhaene et al. (2006).

other policies owned by the same customer. We posit that in order to successfully manage business risk, an insurer needs to adopt the same integrated perspective of the insurance relationship as the one held by the customer. Therefore, the unit of analysis should be the individual policy holder (including all the policies he or she holds with the same company). However, as proposed by Guillén et al. (2006) and Guillén et al. (2008), a second level of analysis could also consider the household as a decision-making unit, as all adult members of the family generally make decisions together about insurance policies that cover their common risks. Nevertheless, real data sets on policy cancellation do not normally include detailed information on households, so it is rather difficult to carry out in practice this second level of analysis.

In this paper we analyze customer loyalty and the profit generated by a customer in different types of policies he may have in the same company. We develop a general procedure that disaggregates the profit in its different dimensions and finally we provide a valuation of the impact of cancellation on business risk in terms of the average expected yearly loss due to policy cancellation. An empirical investigation is also carried with real policy cancellation data.

Our study is organized as follows. Section 1 briefly summarizes a literature review on business risk management in the insurance industry. Section 2, describes three dimensions of the profit loss due to business risk: historical, prospective, and potential. Section 3 presents the notation and formulation for calculating profits. Section 4 describes the data set used in the empirical application. Section 5 presents all the results obtained in our empirical investigation and finally, we discuss the managerial implications of our study in the last section.

1. Literature review

Historically, there has not been much research on customer loyalty and business risk management in the actuarial field, although this has changed in recent years. During the sixties, research interests focused on investigating factors associated with the increasing demand for insurance, such as rising household incomes (Hammond et al., 1967) and women entering the labor force (Duker, 1969). Later studies, based on the portfolio theory, demonstrated that the demand for insurance products is determined simultaneously with the demand for other goods (Mayers and Smith, 1983). On the other hand, Doherty (1984) showed that the level of insurance increased as the number of insurable risks and their weight in the asset portfolio also increases. Additionally, Babbel (1985) claimed that underwriting life insurance products was inversely related to changes in the real price index.

Crosby and Stephens (1987) were the first to investigate the problem of customer loyalty in the insurance industry. They analyzed the effect of relationship marketing on life insurance premiums and customer satisfaction and retention. Their results suggested that non-lapsing customers reported a higher level of satisfaction than lapsing ones (although the insureds were only followed for 13 months).

Interest in studying customer loyalty and satisfaction grew during the eighties. Jackson (1989) observed that very few insurance companies calculated customer lifetime value (CLV) at the time. At first the author estimated CLV in the insurance sector, he emphasized its importance for determining strategies to increase the loyalty of high-valued customers. He proposed both a historical model to analyze the flow of profits generated by customers and a predictive model to determine customer value in the long term.

The calculation of the CLV was continued in the insurance sector throughout the late nineties and, even more intensively, in the most recent years. Berger and Nasr (1998) believed that a systematic method of calculating CLV was missing and therefore they proposed a general mathematical model for calculating CLV in many different cases. The calculation was based on the discounted difference between the revenues and costs, including all promotional costs. To illustrate their calculation, these authors used the case of an insurance product that required a contract or needed the contract to be renewed every year. However, their example was not based on actual empirical data.

During the nineties, as the insurance market became more competitive, researchers began investigating the reasons why customers were switching companies. Schlesinger and Schulenburg (1993), for instance, analyzed a sample of German automobile policy holders and found two main reasons for changing insurers: (1) a favourable premium; and (2) a recommendation by a relative or friend. The authors also conducted parallel analyses for customers who changed insurers at some point in time and those who did not. Their findings suggested that those who changed insurers, were much more satisfied with the claim handling process by the new insurance company.

Several research articles from the nineties also explored the relationship between the insurer's quality of service and customer satisfaction. Wells and Stafford (1995) measured customer perception of service quality and compared it to the number of complaints registered by insurance companies. These authors observed that fewer complaints were significantly correlated with higher levels of perceived service quality. The results also suggested that customers who knew that they had the right to file a complaint evaluated the service quality higher than those who did not realize they had that right. Similarly, Stafford et al. (1998) found that reliability (i.e., the ability to fulfil the promised service in a timely manner) is the most important determinant of service quality and customer satisfaction in the auto casualty claim process. In more recent research on loyalty strategies, Cooley (2002) discussed a two-stage segmentation process to identify four different groups of health insurance policy holders. Using such covariates as age, sex, type of coverage, and seniority, different loyalty strategies were applied to each group based on their particular needs and as a result, customer retention was observed to increase by approximately 7%.

Following Mayers and Smith (1983), Doherty (1984), and Babbel (1985), research interest in the demand for insurance products continued into the nineties as well. Showers and Shotick (1994) investigated the effect of household characteristics and concluded that total household income and the number of members contributing to this income are positively correlated with the demand for insurance products. They also observed that as family size and age of the members increases, the marginal increment in insurance purchases decreases. Likewise, Ben-Arab et al. (1996) constructed a model for the consumption of insurance products in which current individual consumption preferences depended on past consumption patterns. This study showed that the optimal level of purchased insurance products is higher when consumption habit formation is included in the model. So, it is possible to explain why some individuals over-purchase insurance.

Ryals and Knox (2005) summarized the results of a number of studies on customer loyalty and concluded that a small increase in customer retention from 85% to 90% results in net present value profits rising from 35% to 95% amongst the business they examined. The authors also argued that CLV should be adjusted for the risk inherent in establishing a relationship with a customer, such as the volatility of future income flows provided by a new customer. Thus, the authors proposed a measure called *risk-adjusted* CLV, which combines the prediction of CLV with the future risk of the relationship. This measure can also be interpreted as a measure of the economic value (EV) of the customer. Risk-adjusted CLV included: (1) the risk of having a claim (specific to each customer, measured by the ratio between claims and premiums); and (2) the risk of establishing a relationship with the customer (measured by the probability of not retaining the customer). The authors concluded that the measurement of EV can be implemented as a beneficial management tool for addressing relationship marketing strategies.

Verhoef and Donkers (2001) and Donkers et al. (2007) presented two of the most comprehensive studies on the calculation of CLV in the insurance industry. First, Verhoef and Donkers (2001) predicted potential customer value in comparison to realized customer value. Later, Donkers et al. (2007) compared the predictive performance of various models used to calculate CLV in the insurance sector. They considered two types of approximations. The simplest ones include all policies a customer has with the same company, but only consider the total profit the company receives for them. These models are only based on the relationship. On the other hand, more complex models also take all policies into account, but disaggregate the contribution by each one and, therefore, these models are based on the product. The authors concluded that simple models provide good predictions of CLV that are only marginally improved by the more complex models.

Recent papers, by Brockett et al. (2008) and Guillén et al. (2008), provided a first approximation to business risk management in the insurance company. The authors analyzed the relationship between the customer and the insurer, the so-called insurance relationship, and based on that, their conclusions pointed to the need to consider different types of products simultaneously. In that context, the authors estimated the probability of policy cancellation and the customer lifetime duration by using real policy data. Additionally, the authors also provided some guidance for targeting business risk management in the insurance industry. More recently, Lai et al. (2011) find empirical evidence of the moderating effects of inertia and switching costs on the satisfaction-retention link in the auto liability insurance context. They show that the barriers made by switching costs and the behavioral lock-in effect produced by inertia create a pull-back effect, which prevents customers from switching to another insurance provider even in the face of dissatisfaction with the quality of service by the existing provider.

Only a few empirical studies of customer behavior focused on the particularities of the insurance sector exist and more research on business risk management is needed to control the impact of competitive markets on the stability of insurance firms.

2. Profit loss due to business risk

It is necessary to distinguish the profit that a customer is going to generate and the lifetime value of that customer from a marketing perspective. Customer lifetime value is the present value of the future profit stream expected over a given time horizon of transacting with the customer (Kotler, 1974). In this paper we only consider the profit that the customer is expected to generate for the company next year, in order to measure the expected impact of business risk on a yearly basis. Profit here refers to the one that the company is getting from each customer exclusively due to the insurance activity, essentially the balance between premiums and cost of claims.

The analysis of the profit a customer generates is based on three different dimensions:

1. *Historical profit*, which is the profit accumulated during a period of time. We calculate the historical profit of a customer by incorporating available information on underwritten policies regarding claim histories and premiums during a certain period of time. We add the aggregated premiums and subtract the aggregated costs of the claims during that period for each policy the customer has with the company. Note that *historical profit* is based on the same principles as the concept of *realized value* used by Verhoef and Donkers (2001), i.e., a measure of what is the revenue that the customer has generated until now.
2. *Prospective profit*, when the customer does not cancel his or her policies. When we look at what is going to happen in the future, prospective profit takes into account the revenue a company expects to receive from the customer in the short term if the customer keeps the policies in force. Contrary to historical profit which evaluates a customer's past behavior, prospective profit estimates future behavior regarding the policies in force. We consider how many policies the customer currently holds and measure the expected profit they would generate from the information on those policies and the probability of future renewal.
3. *Potential profit*, prospective added profit when the customer underwrites with the same company new policies of a different type than those he currently has. The profit that a customer is going to generate should also incorporate the probability of underwriting other types of products and the profits these products are expected to generate. Therefore, here we focus on the cross-buying behavior of the customer.

Note that *potential profit* mentioned above is not based on the same principles as *customer potential value* in Verhoef and Donkers (2001). When we look at what is going to happen in the future, the concept of *customer potential value* is measuring the profit or value delivered by a customer if this customer behaves ideally, i.e., the customer purchases all products or services he currently buys in the market at the full prices at the focal company (Verhoef and Donkers, 2001). Therefore, it is difficult for the insurer to measure potential value as he normally only knows which products the customer currently has in the focal company. Nevertheless, Verhoef and Donkers (2001) concluded that both socio-demographic and actual purchase information at the focal company are useful predictors of purchase decisions determining potential value.

Therefore, here we slightly depart from the principles of potential value, but we still look at cross-selling opportunities: we are interested in whether or not the customer will keep the policies he currently has (*prospective profit*) and if he is going to underwrite new types of policies (*potential profit*) in order to explore his potentiality of generating profit in other lines of business in the same company. Therefore, by adding prospective profit and potential profit we have some kind of approximation of customer potential value when it is measured in a yearly basis, as current information at the focal company are useful predictors of purchase decisions determining potential customer value.

3. Notation

Let K be the number of insurance product types, T be the total number of years considered in the historical analysis, and N_t be the number of customers in year t, $t = 1,\ldots,T$. Let n_{itk} be the number of policies of product type k, $k = 1,\ldots, K$ the i^{th} individual, $i = 1,\ldots, N_t$ has in year t. Furthermore, P_{itkl} is the premium paid by the i^{th} individual in year t, for the policy l, $l = 1,\ldots, n_{itk}$ of the product type k. Similarly, S_{itkl} is the sum of the costs of claims compensated to the i^{th} individual during year t, for the policy l of the product type k.

Therefore, we can calculate the total amount of premiums for each product type k in a particular year t, that we denote by P_{tk}:

$$P_{\cdot tk\cdot} = \sum_{i=1}^{N_t} \sum_{l=1}^{n_{itk}} P_{itkl}$$

and the total costs of claims for each product type k in year t, $S_{\cdot tk\cdot}$, by

$$S_{\cdot tk\cdot} = \sum_{i=1}^{N_t} \sum_{l=1}^{n_{itk}} S_{itkl} \,.$$

In both cases, if $n_{itk} = 0$ then the term of the sum is equal to zero.

The profit of each product type k in year t can be defined as a function of the total premiums and claims. We will call this function

$$f(P_{\cdot tk\cdot}, S_{\cdot tk\cdot}) \,.$$

This profit is frequently measured by the difference between the total premiums and costs of claims, which we express as f_D,

$$f_D(P_{\cdot tk\cdot}, S_{\cdot tk\cdot}) = P_{\cdot tk\cdot} - S_{\cdot tk\cdot} \,.$$

For a longitudinal analysis of profits, we aggregate the results corresponding to different years as follows:

$$f_D(P_k, S_k) = P_k - S_k,$$

where $P_k = \sum_{t=1}^{T} P_{tk}$ and $S_k = \sum_{t=1}^{T} S_{tk}$ are the sums of premiums and costs of claims accumulated during T years for product type k.

The advantage of this latter expression is that it is additive, which means that

$$f_D(P_k, S_k) = \sum_{t=1}^{T} f_D(P_{tk}, S_{tk}) = \sum_{t=1}^{T}\sum_{i=1}^{N_t} f_D(P_{itk}, S_{itk}),$$

where $f_D(P_{itk\cdot}, S_{itk\cdot}) = P_{itk\cdot} - S_{itk\cdot} = \sum_{l=1}^{n_{itk}} P_{itkl} - \sum_{l=1}^{n_{itk}} S_{itkl}$ corresponds to the profit generated by the i^{th} individual in year t, for the product type k. Obviously, if the customer has no policy of product type k in year t, $n_{itk} = 0$, we would sum 0.

3.1. Historical profit. The historical profit of the i^{th} customer C_{iD}^{H} is measured by

$$C_{iD}^{H} = f_D(P_{i\ldots}, S_{i\ldots}) = \sum_{k=1}^{K}\sum_{t=1}^{T} f_D(P_{itk\cdot}, S_{itk\cdot}).$$

This measure, for a given customer, incorporates all information on different types of policies and years and provides the profit that the customer has generated during that period of time. It gives the same information as the *realized value* in Verhoef and Donkers (2001).

3.2. Prospective profit. The prospective profit of the customer is estimated to reflect the possibility that a customer will renew the policies he or she currently has without underwriting new contracts in the next year. Let y_{iTkl} be a binary variable equal to 1 when the i^{th} individual has the l^{th} policy, $l = 1,\ldots, n_{iTk}$ of product type k, in year T, and we estimate the probability that this policy will be renewed next year, provided that it was already in force the previous year, p_{iTkl} given by

$$p_{iTkl} = \Pr(y_{i(T+1)kl} = 1 \mid y_{iTkl} = 1).$$

Note that the estimation of this probability could incorporate other features characterizing the customer that could be used in a logistic regression model.

If the customer renews the policy, then the company expects a profit. The company's profit can be calculated as the average of the profits obtained from customers holding this particular type of product during the previous year, which we will call B_{DTk}. That is,

$$B_{DTk} = \frac{1}{M_{Tk}} \sum_{i=1}^{M_{Tk}} f_D(P_{iTk\cdot}, S_{iTk\cdot}),$$

where M_{Tk} is the number of customers holding the product type k in the year T.

The prospective profit C_{iD}^{PRO} of the i^{th} customer is obtained by summing the product of the probability of keeping policies and the expected profit from them, for all types of products the customer holds. Therefore,

$$C_{iD}^{PRO} = \sum_{k=1}^{K}\sum_{l=1}^{n_{iTk}} p_{iTkl} B_{DTk}.$$

3.3. Potential profit. We estimate the potential profit by estimating the probability that a customer will underwrite next year new policies of a different type than those he currently has. We estimate the probability that $n_{i(T+1)k}$ would be greater than zero, conditioned to $n_{iTk} = 0$, what we denote by p_{iTk}. That is,

$$p_{iTk} = Pr(n_{i(T+1)k} > 0 \mid n_{iTk} = 0),$$

where $p_{iTk} = 0$ if $n_{iTk} > 0$. Note that, again, the estimation of this probability could be done by using a logistic regression model.

Therefore, we calculate the potential profit C_{iD}^{POT} by summing the product of the probability of a customer underwriting policies of a different type than his or her current policies and the average profit obtained for this type of product during the previous year B_{DTk}, namely

$$C_{iD}^{POT} = \sum_{k=1}^{K} p_{iTk} B_{DTk}.$$

Note that for prospective and potential profit calculation, we slightly depart from the way profits were determined by Donkers et al. (2007) as part of their calculation of customer value in the so-called relationship-level models. The authors assumed that the total profit each customer will generate next year will be the same as in the previous year, in case that the company retains that customer. As we are interested in analyzing the overall impact of business risk on the total portfolio, instead, we assumed that the profit that will be generated by each policy the customer holds is the average of the profits obtained from customers holding that particular type of policy the previous year. This procedure is similar to the way profit margins were determined in the so-called service level models by Donkers et al. (2007).

3.4. The loss due to business risk. The expected loss due to business risk that the company will have next year L_{BR} can be calculated as a function of the individual prospective and potential losses, C_{iD}^{PRO} and C_{iD}^{POT}. We will consider two possible situations that can generate a loss next year: a partial and a total cancellation. A partial cancellation occurs when the customer cancels some of his policies (but not all of them) and a total cancellation, when all policies are cancelled. We will assume that if a partial cancellation occurs, the loss generated by this customer is only the average prospective profit per policy. This is equivalent to assume that only one policy would be cancelled, and that policy would have generated a profit next year equal to the average prospective profit during the previous year. By doing so, we are assuming the most common and best possible situation when a partial cancellation occurs, as at least one policy would have been cancelled. On the other hand, if a total cancellation occurs, then we assume that the company looses all the prospective and potential profit. Based on these assumptions, the loss due to business risk that the company expects to have next year is:

$$L_{BR} = \sum_{\forall i} \left\{ \begin{array}{l} P(0 < n_{i(T+1)} < n_{iT} \mid n_{iT} > 0) \cdot \dfrac{C_{iD}^{PRO}}{n_{iT}} + \\ + P(n_{i(T+1)} = 0 \mid n_{iT} > 0) \cdot \left[C_{iD}^{PRO} + C_{iD}^{POT} \right] \end{array} \right\},$$

where $n_{it} = \sum_{k=1}^{K} n_{itk}$ is the total number of policies the i^{th} customer has in year t, $t = T$, $T + 1$.

4. Data

Our sample included 79599 customers who had at least one policy with a European insurance company as of December 31, 2005, of these types: automobile insurance (also called motor insurance), diverse non-automobile (which includes house and contents, funeral and accident insurance), health, and agricultural insurance[1]. We analyzed customer behavior over time from December 31, 2005 to March 31, 2008, in order to estimate the probability of policy cancellation and determine the factors affecting customer loyalty.

In Table 1 we present sample characteristics at the beginning of the period of study. A majority of the customers were men (61.70% of the sample), while 28.61% were women (the rest were either firms or the gender was not identified). The average age was 47.03 years (the standard deviation was 14.10). Most customers were married (47.85%), although 15.26% were single and civil status was not declared by 35.54%. A majority of the sample (55.34%) lived in rural villages/towns, but 37.43% lived in urban areas. Approximately 24% of customers in the sample identified their type of consumption as "affirmation" (which means they had enough earnings to afford high-quality, permanent goods and still had sufficient resources for leisure activities). The average seniority (meaning the average length of insurance relationship) in the company was 8.96 years (the standard deviation was 8.41).

In Table 2, we compare the composition of policies in force for each customer on December 31, 2005, with those in force a year later. We observe that 83.22% of customers had only one type of policy in force at the end of 2005 (mostly automobile or diverse policies), 15.69% had two types of policies, and less than 2% had three or more types of policies.

More than 94% of customers with one type of policy at the end of 2005 still had only that one type a year later. Most of the remaining 6% underwrote a new product. Among those with at least two types of policies, the percentage of customers who had the same types one year later ranges from 70.4% (those with automobile, diverse, and health policies) to 89.2% (those with automobile and diverse policies). On the other hand, the percentage of those who cancelled their policies one year later ranges from 9.9% (those with automobile, diverse, and agricultural policies) to 27.3% (those with automobile, health and agricultural policies). Finally, the largest group of customers who underwrote a new policy was the one with health and agricultural policies and later added an automobile policy (7.7%).

Table 1. Sample characteristics as of December 31, 2005

Frequency			Percent
Gender	Men	49.116	61,70%
	Women	22.772	28,61%
	Firm	3.179	3,99%
	Classification not established	4.532	5,69%
	Total	79.599	100%
Marital status	Married	38.091	47,85%
	Divorced	53	0,07%
	Separated	559	0,70%
	Single	12.148	15,26%
	Widow	456	0,57%
	Classification not established	28.292	35,54%
	Total	79.599	100%
Address	Rural (< 100.000 inhab.)	44.047	55,34%
	Urban (> 100.000 inhab.)	29.795	37,43%
	Unknown	5.757	7,23%
	Total	79.599	100%
Type of consumption	Luxurious (1)	3.333	4,19%
	Socially improving (2)	8.777	11,03%
	Affirmation (3)	19.432	24,41%
	Emulation (4)	18.097	22,74%
	Subsistence (5)	7.240	9,10%
	Classification not established	22.720	28,54%
	Total	79.599	100%

Notes: (1) surplus resources allow for access to luxuries without affecting the future possibility of consumption; (2) surplus resources allow for access to high status consumption; (3) earnings are sufficient for buying high-quality, permanent goods with enough for leisure as well; (4) earnings are for goods which are immediately consumed; and (5) no regular earnings.

[1] Unfortunately, information on premiums and claims is only available for two lines of business: automobile and diverse. Therefore, only these ones are finally included in the calculations of profits.

Table 2. Policies in force on December 31, 2005 vs. December 31, 2006. Absolute frequency and row percentage

2006 2005	----	---G	--S-	--SG	-D--	-D-G	-DS-	-DSG	A---	A--G	A-S-	A-SG	AD--	AD-G	ADS-	ADSG	Total row
---G	18 0.8%	2027 94.9%	0 0.0%	0 0.0%	3 0.1%	17 0.8%	0 0.0%	0 0.0%	7 0.3%	60 2.8%	0 0.0%	0 0.0%	0 0.0%	4 0.2%	0 0.0%	0 0.0%	2136 2.68%
--S-	35 1.7%	0 0.0%	1999 94.1%	3 0.1%	8 0.4%	0 0.0%	39 1.8%	0 0.0%	5 0.2%	0 0.0%	28 1.3%	0 0.0%	2 0.1%	0 0.0%	5 0.2%	0 0.0%	2124 2.67%
--SG	0 0.0%	3 11.5%	1 3.9%	20 76.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	26 0.03%
-D--	135 0.6%	3 0.0%	2 0.0%	0 0.0%	23719 97.5%	27 0.1%	22 0.1%	0 0.0%	39 0.2%	0 0.0%	0 0.0%	0 0.0%	380 1.6%	0 0.0%	1 0.0%	0 0.0%	24328 30.56%
-D-G	0 0.0%	14 4.6%	0 0.0%	0 0.0%	17 5.6%	265 87.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	7 2.3%	0 0.0%	0 0.0%	303 0.3%
-DS-	4 0.8%	0 0.0%	23 4.6%	0 0.0%	83 16.6%	0 0.0%	370 73.9%	0 0.0%	1 0.2%	0 0.0%	1 0.2%	0 0.0%	6 1.2%	0 0.0%	13 2.6%	0 0.0%	501 0.63%
-DSG	0 0.0%	1 5.3%	0 0.0%	0 0.0%	0 0.0%	3 15.8%	1 5.3%	14 73.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	19 0.02%
A---	387 1.0%	5 0.0%	3 0.0%	0 0.0%	58 0.2%	1 0.0%	1 0.0%	0 0.0%	36180 96.1%	104 0.3%	42 0.1%	0 0.0%	870 2.3%	4 0.0%	2 0.0%	0 0.0%	37657 47.31%
A--G	2 0.1%	55 3.8%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	76 5.2%	1301 88.7%	0 0.0%	4 0.3%	1 0.1%	27 1.8%	0 0.0%	0 0.0%	1466 1.84%
A-S-	3 0.7%	0 0.0%	24 5.4%	0 0.0%	1 0.2%	0 0.0%	1 0.2%	0 0.0%	66 14.9%	0 0.0%	327 73.8%	0 0.0%	4 0.9%	1 0.2%	16 3.6%	0 0.0%	443 0.56%
A-SG	0 0.0%	0 0.0%	0 0.0%	2 9.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	4 18.2%	0 0.0%	16 72.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	22 0.03%
AD--	25 0.3%	0 0.0%	0 0.0%	0 0.0%	444 4.6%	2 0.0%	1 0.0%	0 0.0%	535 5.5%	1 0.0%	4 0.0%	0 0.0%	8693 89.2%	32 0.3%	13 0.1%	0 0.0%	9750 12.25%
AD-G	1 0.2%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	8 1.7%	0 0.0%	0 0.0%	3 0.7%	22 4.7%	0 0.0%	0 0.0%	16 3.5%	412 88.8%	0 0.0%	1 0.2%	464 0.58%
ADS-	0 0.0%	0 0.0%	1 0.3%	0 0.0%	4 1.2%	0 0.0%	10 3.0%	0 0.0%	4 1.2%	0 0.0%	11 3.3%	0 0.0%	66 19.5%	2 0.6%	238 70.4%	2 0.6%	338 0.42%
ADSG	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 4.6%	0 0.0%	1 4.6%	0 0.0%	1 4.6%	3 13.6%	16 72.7%	22 0.03%
Total row																	79599

Notes: The policies the customer has at any moment are represented by a string of four characters in the following order ‘A’ (automobile policy), ‘D’ (diverse policy), ‘S’ (health policy) and ‘G’ (agricultural policy). The symbol ‘-’ indicates that the customer does not have the corresponding policy.

5. Empirical application

We estimate the probability of policy cancellation within three time periods, starting from December 31, 2005: short term (180 days), medium term (1 year), and long term (2 years and 3 months, the whole time period being analyzed). We use logistic regression to predict the probability of observing a cancellation in each time period, based on a set of explanatory covariates. The variable descriptions and regression results are presented in the Appendix (Tables A1 to A4).

We use these models to estimate the probability of cancellation for different types of customers. In Table 3, we fix the customer's age, seniority (longevity in the company), and premium at the sample mean[1]. Then, we calculate the probability of cancellation depending on gender, marital status, and type of policy (assuming the customer only has one) for each time period. For example, the probability of cancellation for a married man with a health policy is 14.01% in the short term, 18.83% in the medium term, and 26.90% in the long term. Furthermore, the probability of cancellation for a single woman with an automobile policy ranges from 1.60% (short term) to 6.93% (long term).

In Table 4, we show the average historical profit generated by customers and also the estimation of prospective and potential profits for next year depending on gender, marital status and types of policies in force. As previously mentioned, data on premiums and claims are only available for automobile and diverse policies, therefore, in these predictions we only consider customers having one of these two types of policies. The historical profit in Table 4 is simply the average profit (average of the difference between premiums and costs of claims) generated by these customers during 2006 and 2007. Married men with a diverse policy have the highest historical profit at 863,88 euros. In general, the historical profit of customers with policies in the diverse line of business is higher than of those with automobile policies, except for single women. Women have lower historical values than men, and single customers have higher historical values than those who are married.

To calculate the prospective profit, we use a logistic regression model (see Appendix, Tables A5 to A8) to predict the probability of renewal for each customer and then multiply that probability by the average profit of the corresponding type of policy for the previous year[2]. We find that, on average, customers in the diverse line of business have slightly higher prospective profits than those in the automobile line of business. On the other hand, there are only very small differences among customers in terms of gender or marital status. Actually, the estimation of prospective profit for customers having a diverse policy is the same for different gender or marital status because these two variables have no significant parameters in the corresponding logistic regression model.

Table 3. Probability of cancellation by customer characteristics, time period, and policy

			Policy			
		Marital status	Auto	Health	Diverse	Agro
Men	Short term	Married	2.72%	14.01%	4.96%	3.34%
		Single	1.77%	9.51%	3.26%	2.18%
	Medium term	Married	6.64%	18.83%	9.82%	7.55%
		Single	4.24%	12.61%	6.34%	4.84%
	Long term	Married	14.16%	26.90%	19.77%	15.95%
		Single	9.09%	18.23%	13.00%	10.32%
Women	Medium term	Married	2.46%	12.80%	4.49%	3.02%
		Single	1.60%	8.64%	2.94%	1.97%
	Short term	Married	5.32%	15.49%	7.92%	6.07%
		Single	3.38%	10.24%	5.08%	3.86%
	Long term	Married	10.93%	21.50%	15.50%	12.38%
		Single	6.93%	14.24%	10.01%	7.89%

[1] Age is 47.03 years; longevity in the company is 8.96 years; and premium is 381.72 euros.

[2] The average profit for the automobile line of business is 240.67 euros if the customer has only one policy and 513.01 euros if he has two or more policies. Average profit for the diverse line of business is 285.23 euros for only one policy and 510.78 euros for two or more policies.

Table 4. Historical, prospective and potential average profits (premiums minus total claims) for 2006-2007 in euros by type of policy

	Marital status	Profit	Policy	
			Auto	Diverse
Men	Married	Historical	475.86	863.88
		Prospective	239.56	265.86
		Potential	7.34	78.39
	Single	Historical	547.29	732.23
		Prospective	239.16	265.86
		Potential	5.62	42.01
Women	Married	Historical	440.85	573.23
		Prospective	240.16	265.86
		Potential	5.43	29.89
	Single	Historical	522.21	488.79
		Prospective	239.90	265.86
		Potential	4.15	14.54

To estimate potential profit, we specify a logistic regression model (see Appendix, Tables A9 and A10) to predict the probability of underwriting policies of a different type than those the customer currently has and multiply that probability by the average profit of the corresponding type of policy during the previous year[1]. We show in Table 4 that customers with one policy covering diverse risks have higher average potential values than those with one policy of automobile. This is in part because it is more likely that a customer having a diverse policy would underwrite an automobile policy than the other way around. We also observe that men generally have a higher potential profit than women, and married customers have a higher profit than single customers.

The expected profit loss for next year due to business risk can be estimated as described in section 3. Here we split the contribution of partial cancellations and total cancellation on the expected total loss. The expected loss due to partial cancellations for the total sample of 43133[2] customers is 0.87 million euros. Here, the probability of cancellation has been estimated using logistic regression model in Table A3 in the Appendix, which corresponds to cancellations in the medium term (one year time horizon).

On the other hand, in order to measure the expected impact of total cancellations on the loss due to business risk we need to know the probability of a total cancellation, which we assumed the one observed in the sample, which is 0.76%. We multiply this probability by the total customer profit (prospective profit plus expected profit) and find an additional loss of 0.04 million euros.

Therefore, the total expected loss this sample portfolio will incur due to cancellations is 0.87 + 0.04 = 0.91 million euros. Namely, 0.91 million euros / 43133 = 21.1€ per customer. However, this estimation is only an expected value and does not represent a risk measure, which should be calculated with confidence levels of 95% for obtaining the quantile value of loss at this level.

The most straightforward way to optimize the way funds aimed at increasing retention are used, is to invest them in customers generating the highest profits, while trying to avoid those with the lowest contribution to profits. In Table 5, we identify four groups of customers according to their loyalty and profit. Profits here are calculated by adding prospective and expected profits, that is, revenue the company anticipates in the next year. Historical profit is not included, as it is a measure of the customer's past value, but its calculation is essencial as we saw that future profit estimation is based on the realized or historical profit observed in the past.

We divide customers by high and low profit (profit that is greater or smaller than the median of approximately 240€) and by high and low loyalty (probability of cancellation that is smaller or greater than the median of 6.4%). We classify the total number of customers in our sample accordingly in order to suggest different retention strategies. Finally, we find that the company in our example should concentrate its retention efforts on the customer generating the highest profit (the right column of Table 5).

Table 5. Loyalty vs. profit

Loyalty	Value	
	Low profit Profit < 240	High profit Profit ≥ 240
Low loyalty Prob. ≥ 6.4%	Age: 39.55 Longevity: 5.79 83.99% men 89.53% married 7.75% single	Age: 47.50 Longevity: 10.09 82.05% men 83.53% married 13.86% single

[1] The average profit is 299.07€ for automobile and 266.26€ for diverse policies (in the calculation of the average, all customers having these particular type of policies have been considered, independently on how many of them).

[2] The final sample size was 43.133 after delating some observations containing missing values.

Table 5 (cont.). Loyalty vs. profit

Loyalty	Value	
	Low profit Profit < 240	High profit Profit ≥ 240
High loyalty Prob. < 6.4%	Age: 46.05 Longevity: 8.44 73.63% men 50.23% married 49.01% single	Age: 54.00 Longevity: 13.47 57.00% men 85.12% married 14.06% single

Conclusions

An analysis of loyalty and the profit generated by the customer is the foundation for managing business risk. In this article, we propose a method for determining expected losses due to policy cancellation by estimating the profit generated by each customer and predicting their probability of cancellation.

The analysis of the profit the customer generates is done in three dimensions: historical, prospective and potential profit. The first of them considers the profit generated in the past and the other two are the ones that can be generated in the next period as a result of keeping the policies in force (prospective profit) or underwriting policies of a different type (potential profit). Based on them, the expected profit loss due to business risk is formulated.

An empirical application is carried out by analyzing a sample of customers. Different products are considered in our analysis simultaneously. We observe that factors such as gender or civil status affects the probability of cancellation, and also that health policies are more likely to be cancelled than the other types of policies being considered here. We additionally calculate the historical, prospective and potential profits for customers in our sample holding diverse or automobile policies.

Based on these calculations, the total expected loss the company will incur due to business risk is calculated for our sample. This is not a risk measure but only the expected value of the loss due to business risk. Nevertheless, it is a valuable information for managers, as it is the limit of funds that should be invested in customer loyalty. Segmentation strategies can be applied in order to decide how to invest these funds in customer retention. The research carried out here can be extended in order to define suitable measures that could better represent the exposure of the company to business risk, in the context of classical risk measures.

References

1. Babbel, D. (1985). The price elasticity of demand for whole life insurance, *The Journal of Finance*, 40 (1), pp. 225-239.
2. Ben-Arab, M., Brys, E. & Schlesinger, H. (1996). Habit formation and the demand for insurance, *Journal of Risk and Insurance*, 63 (1), pp. 111-119.
3. Berger, P.D. & Nasr, N. (1998). Customer lifetime value: marketing models and applications, *Journal of Interactive Marketing*, 12, pp. 17-30.
4. Brockett, P.L., Golden, L., Guillén, M., Nielsen, J.P., Parner, J. & Pérez-Marín, A.M. (2008). Survival analysis of household insurance policies: How Much Time Do You Have to Stop Total Customer Defection? *Journal of Risk and Insurance*, 75 (3), pp. 713-737.
5. Cooley, S. (2002). Loyalty strategy development using applied member-cohort segmentation, *Journal of Consumer Marketing*, 19 (7), pp. 550-563.
6. Crosby, L.A. & Stephens, N. (1987). Effects of relationship marketing on satisfaction, retention, and prices in the life insurance industry, *Journal of Marketing Research*, 24 (4), pp. 404-411.
7. Dhaene, J., Vanduffel, S., Goovaerts, M.J., Kaas, R., Tang, Q. & Vyncke, D. (2006). Risk measures and comonotonicity: a review, *Stochastic Models*, 22, pp. 573-606.
8. Doherty, N.A. (1984). Portfolio efficient insurance buying strategies, *The Journal of Risk and Insurance*, 51 (2), pp. 205-224.
9. Donkers, B., Verhoef, P.C. & Jong, M.G. (2007). Modeling CLV: a test of competing models in the insurance industry, *Quantitative Marketing and Economics*, 5 (2), pp. 163-190.
10. Duker, J.M. (1969). Expenditures for life insurance among working-wife families, *The Journal of Risk and Insurance*, 36 (5), pp. 525-533.
11. Guillén, M., Nielsen, J.P. & Pérez-Marín, A.M. (2006). La duración de distintos contratos de seguros en los hogares, Un enfoque integrado, *Gerencia de Riesgos y Seguros*, 96, pp. 23-32.
12. Guillén, M., Nielsen, J.P. & Pérez-Marín, A.M. (2008). The need to monitor customer loyalty and business risk in the European insurance industry, *Geneva Papers on Risk and Insurance – Issues and Practice*, 33, pp. 207-218.
13. Hammond, J.D., Houston, D.B. & Melander, E.R. (1967). Determinants of household life insurance premium expenditures: an empirical investigation, *The Journal of Risk and Insurance*, 34 (3), pp. 397-408.
14. Jackson, D. (1989). Determining a customer's lifetime value, part three, *Direct Marketing*, 52 (4), pp. 28-30.
15. Kotler, P. (1974). Marketing during periods of shortage, *Journal of Marketing*, 38 (3), pp. 20-29.
16. Lai, Li-Hua, Liu, Chun Ting, Lin, Jinn Tyan (2011). The moderating effects of switching costs and inertia on the customer satisfaction-retention link: auto liability insurance service in Taiwan, *Insurance Markets and Companies: Analyses and Actuarial Computations*, 2 (1), pp. 69-78.
17. Mayers, D. & Smith, C.W.Jr. (1983). The interdependence of individual portfolio decisions and the demand for insurance, *The Journal of Political Economy*, 91 (2), pp. 304-311.

18. Nakada, P., Shah, H., Koyluoglu, H.U. & Collignon, O. (1999). P&C RAROC: a catalyst for improved capital management in the property and casualty insurance industry, *The Journal of Risk Finance*, 1 (1), pp. 1-18.
19. Ryals, L.J. & Knox, S. (2005). Measuring risk-adjusted customer lifetime value and its impact on relationship marketing strategies and shareholder value, *European Journal of Marketing*, 39 (5/6), pp. 456-472.
20. Schlesinger, H. & Schulenburg, J.M. (1993). Customer information and decisions to switch insurers, *Journal of Risk and Insurance*, 60 (4), pp. 591-615.
21. Showers, V.E. and Shotick, J.A. (1994). The effects of household characteristics on demand for insurance: a tobit analysis, *Journal of Risk and Insurance*, 61, pp. 492-502.
22. Stafford, M.R., Stafford, T.F. & Wells, B.P. (1998). Determinants of service quality and satisfaction in the auto casualty claims process, *Journal of Services Marketing*, 12 (6), pp. 426-40.
23. Verhoef, P.C. and B. Donkers (2001). Predicting customer potential value: an application to the insurance industry, *Decision Support Systems*, 2, pp. 189-199.
24. Wells, B.P. & Stafford, M.R. (1995). Service quality in the insurance industry. Customer perception versus regulatory perceptions, *Journal of Insurance Regulation*, 13 (4), pp. 462-477.

Appendix

Table A1. Explanatory covariates

Variable	Description
Longevity	Number of whole years passed since the customer underwrote the first policy in the company until December 31, 2005.
Age	Age of the customer as of December 31, 2005.
Mars	Marital status, specified by four binary covariates: Mars_M – married, Mars_W – widow, Mars_D – Divorced, and Mars_S – Separated. The reference group is composed of single individuals. Those for whom the marital status is unknown have been eliminated from the analysis, including firms.
Gender	Gender of the customer, indicated by a binary covariate which is equal to 1 when the customer is male.
Premium	Premium paid by the customer[1]. If a cancellation has occurred, we took the premium paid for the policy being cancelled as representative of the premium level of that customer. If there has not been any cancellation, we took the premium of the first policy that has been modified during the period considered in the analysis, usually the first one that was underwritten or renewed during the period.
Auto	Automobile policy. Binary covariate equal to 1 if the customer has an automobile policy in force as of December 31, 2005.
Health	Health policy. Binary covariate equal to 1 if the customer has a health policy in force as of December 31, 2005.
Diverse	Diverse policy. Binary covariate equal to 1 if the customer has a diverse policy in force as of December 31, 2005.
Agro	Agricultural policy. Binary covariate equal to 1 if the customer has an agricultural policy in force as of December 31, 2005.

Table A2. Logistic regression model estimation for the probability of cancellation in the short term (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	-3.9717	0.1227	1047.5313		<0.0001
Longevity	-0.0204	0.0040	26.5386	0.980	<0.0001
Age	-0.0090	0.0021	18.9263	0.991	<0.0001
Mars_M	0.4393	0.0646	46.1738	1.552	<0.0001
Mars _W	0.9319	0.2107	19.5693	2.539	<0.0001
Mars _D	0.9149	0.4201	4.7419	2.496	0.0294
Mars _S	0.1668	0.1560	1.1445	1.182	0.2847
Gender	0.1045	0.0556	3.5335	1.110	0.0601
Premium	0.0002	0.0001	9.4206	1.000	0.0021
Auto	0.3907	0.0806	23.4857	1.478	<0.0001
Health	2.1537	0.0815	698.3346	8.616	<0.0001
Diverse	1.0152	0.0488	432.0603	2.760	<0.0001
Agro	0.6027	0.0937	41.4016	1.827	<0.0001

Notes: (*) After deleting observations with incomplete information, we estimated the probability of cancellation in the short term for a sample of 48798 customers. This includes 2052 cancellations (4.2%) during the first 180 days of 2006. Results support the overall significance of the model. The likelihood ratio is 1380.6, with 12 degrees of freedom, and p-value less than 0.0001.

Table A3. Logistic regression model estimation for the probability of cancellation in the medium term (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	-3.0913	0.0924	1120.4427		<0.0001
Longevity	-0.0192	0.0028	46.6432	0.981	<0.0001
Age	-0.0151	0.0015	102.6480	0.985	<0.0001
Mars_M	0.4746	0.0457	107.6842	1.607	<0.0001

[1] Tariff premium without taxes, gross yearly amount.

Table A3 (cont.). Logistic regression model estimation for the probability of cancellation in the medium term (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Mars _D	1.0369	0.3631	8.1535	2.820	0.0043
Mars _S	0.0505	0.1435	0.1238	1.052	0.7250
Gender	0.2353	0.0406	33.6287	1.265	<0.0001
Premium	-0.0001	0.0001	1.3384	1.000	0.2473
Auto	0.6411	0.0650	97.1906	1.899	<0.0001
Health	1.8232	0.0699	680.8117	6.192	<0.0001
Diverse	1.0668	0.0349	932.3473	2.906	<0.0001
Agro	0.7796	0.0644	146.3683	2.181	<0.0001

Notes: (*) Our sample of 48798 customers includes 4327 cancellations (8.9%) during 2006. Results support the overall significance of the model. The likelihood ratio is 1885.06, with 12 degrees of freedom, and p-value less than 0.0001.

Table A4. Logistic regression model estimation for the probability of cancellationin the long term (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	-2.1840	0.0724	911.1926		<0.0001
Longevity	-0.0141	0.0021	44.8662	0.986	<0.0001
Age	-0.0187	0.0011	275.0211	0.981	<0.0001
Mars_M	0.5007	0.0343	213.0012	1.650	<0.0001
Mars _W	0.7679	0.1340	32.8407	2.155	<0.0001
Mars _D	1.1500	0.3236	12.6318	3.158	0.0004
Mars _S	-0.0096	0.1309	0.0053	0.990	0.9418
Gender	0.2953	0.0308	91.7539	1.343	<0.0001
Premium	-0.0003	0.0001	46.0513	1.000	<0.0001
Auto	0.7065	0.0531	176.7519	2.027	<0.0001
Health	1.5089	0.0628	578.0557	4.522	<0.0001
Diverse	1.1082	0.0273	1.642.6860	3.029	<0.0001
Agro	0.8471	0.0511	275.1147	2.333	<0.0001

Notes: (*) Our sample of 48798 customers includes 8407 cancellations (17.2%) from December 31, 2005 to March 31, 2008. Results support the overall significance of the model. The likelihood ratio is 2831.06, with 12 degrees of freedom, and p-value less than 0.0001.

Table A5. Estimation of the logistic regression model for the probability of renewing an automobile policy if that policy was in force in the previous year (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	4.5436	0.2687	285.9964		<0.0001
Longevity	0.0421	0.0135	9.7259	1.043	0.0018
Age	0.0175	0.0067	6.8680	1.018	0.0088
Mars_M	0.3091	0.1749	3.1248	1.362	0.0771
Mars_D	-2.4697	1.0553	5.4774	0.085	0.0193
Mars_S	-2.3700	0.4186	32.0535	0.093	<0.0001
Gender	-0.6811	0.1860	13.4107	0.506	0.0003
Diverse	-0.4683	0.1699	7.6017	0.626	0.0058

Notes: (*) The sample used consists of 34340 observations corresponding to customers with one automobile policy in force at the end of 2006. Most of them, 34147 customers, renewed their policies (99.44%) the next year. Results support the overall significance of the model. The likelihood ratio is 80.57, with 7 degrees of freedom, and p-value less than 0.0001.

Table A6. Estimation of the logistic regression model for the probability of renewing two or more automobile policies if that policies were in force in the previous year (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	1.1296	0.0742	231.6066		<0.0001
Longevity	0.0230	0.0041	32.2088	1.023	<0.0001
Mars_M	0.2325	0.0708	10.7684	1.262	0.0010

Notes: (*) The sample used consists of 10637 observations corresponding to customers with two or more automobile policies in force at the end of 2006. Most of them, 8790 customers, renewed their policies (82.64%) the next year. Results support the overall significance of the model. The likelihood ratio is 45.70, with 2 degrees of freedom, and p-value less than 0.0001.

Table A7. Estimation of the logistic regression model for the probability of renewing a diverse policy if that policy was in force in the previous year (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	1.4435	0.1609	80.5108		<0.0001
Longevity	0.0230	0.0075	9.3749	1.023	0.0022
Age	0.0206	0.0037	30.3589	1.021	<0.0001

Notes: (*) The sample used consists of 8581 observations corresponding to customers with one diverse policy in force at the end of 2006. Most of them, 7992 customers, renewed their policies (93.14%) the next year. Results support the overall significance of the model. The likelihood ratio is 61.94, with 2 degrees of freedom, and p-value less than 0.0001.

Table A8. Estimation of the logistic regression model for the probability of renewing two or more diverse policies if that policies were in force in the previous year (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	1.3689	0.3016	20.6056		<0.0001
Longevity	0.0294	0.0113	6.8108	1.030	0.0091
Age	0.0151	0.0057	7.1198	1.015	0.0076
Gender	-0.4907	0.1890	6.7418	0.612	0.0094

Notes: (*) The sample used consists of 2335 observations corresponding to customers with two or more automobile policies in force at the end of 2006. Most of them, 2061 customers, renewed their policies (88.26%) the next year. Results support the overall significance of the model. The likelihood ratio is 23.08, with 3 degrees of freedom, and p-value less than 0.0001.

Table A9. Estimation of the logistic regression model for the probability of underwriting an automobile policy if the customer did not have any policy of this type the previous year (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	-0.8139	0.1519	28.7177		<0.0001
Longevity	-0.0322	0.0078	16.9401	0.968	<0.0001
Age	-0.0398	0.0038	107.5857	0.961	<0.0001
Mars_M	0.7765	0.1034	56.3831	2.174	<0.0001
Mars_W	1.0061	0.4315	5.4362	2.735	0.0197
Mars_S	-1.6627	0.3331	24.9204	0.190	<0.0001
Gender	1.1627	0.0961	146.4152	3.199	<0.0001

Notes: (*) The sample used consists of 4533 observations corresponding to customers who do not have any automobile policy in force at the end of 2006. Only 810 of them underwrote an automobile policy the next year (17.87%). Results support the overall significance of the model. The likelihood ratio is 478.05, with 6 degrees of freedom, and p-value less than 0.0001.

Table A10. Estimation of the logistic regression model for the probability of underwriting a diverse policy if the customer did not have any policy of this type the previous year (*)

Variable	Estimated parameter	Standard error	Chi-square	OR	p-value
Constant	-3.7794	0.1256	905.1696		<.0001
Age	-0.0078	0.0028	7.9808	0.992	0.0047
Mars_M	0.2737	0.0874	9.8181	1.315	0.0017
Mars_W	1.0975	0.3121	12.3651	2.997	0.0004
Mars_D	1.9489	0.6194	9.9010	7.021	0.0017
Gender	0.3092	0.0799	14.9779	1.362	0.0001

Notes: (*) The sample used consists of 38594 observations corresponding to customers who do not have any diverse policy in force at the end of 2006. Only 942 of them underwrote an automobile policy the next year (2.44%). Results support the overall significance of the model. The likelihood ratio is 36.54, with 5 degrees of freedom, and p-value less than 0.0001.