Cross-buying behaviour and customer loyalty in the insurance sector

Guillén, M., Perch-Nielsen, J.; Pérez-Marín, A.M. (2009). "Cross-buying behaviour and customer loyalty in the insurance sector". EsicMarket, 132, pp. 77-105.

Abstract

Customer loyalty is one of the main business challenges, also for the insurance sector. Nevertheless, there are just a few papers dealing with this problem in the insurance field and specifically considering the uniqueness of this business sector. In this paper we define the conceptual framework for studying this problem in insurance and we propose a methodology to address it. With our methodological approach, it is possible to estimate the probability that a household with more than one insurance contract (policy) in the same insurance company (cross-buying) would cancel all policies simultaneously. For those who cancel part of their policies, but not all of them, we estimate the time they are going to stay in the company after that first policy cancellation, that is to say, the time the company has to try to retain a customer who has just given them a clear signal of leaving the company. Additionally, in this paper we present and discuss the results obtained when applying our methodology to a policy cancellation dataset provided by a Danish insurance company, and we outline some conclusions regarding the factors associated to a higher or lower customer loyalty.

Key words: Loyalty, customer lifetime duration, cross-buying, policy cancellation, insurance.

JEL Code: M31.

Montserrat Guillén / Ana María Pérez-Marín
Departamento de Econometría RFA-IREA. Universidad de Barcelona
Jens Perch-Nielsen
Festina Lente and University of Copenhagen

1. Introduction

Loosing customers is a very important problem in the insurance field. As in any other sectors, it produces a market share loss for the firm. Even though this effect can be compensated by new customers, in the insurance firm the composition and quality of insurance risks are severely distorted when contracts are massively cancelled and this has a negative impact on the solvency of the company. For that reason, the risk of loosing marked share and loosing clients is called business risk management in the insurance industry (Nakada, Shah, Koyluogo and Collignon, 1999) and is becoming increasingly central.

Insurance companies have deeply changed during last years. The increasing competition in the sector, partly caused by the introduction of the internet and by customers' information costs reduction, has forced insurance companies to orientate their management much more towards the client, while traditionally their activity has been based mainly on developing the technical aspects of the insurance products offered by them.

For this reason, to increase customer loyalty has become necessary for the insurance company. Moreover, it is also important for the insurer to understand and develop the relationship that the company keeps with the client via the internet. In this sense, the research done by Martín and Quero (2004) and Flavián and Guilaníu (2007) can be the starting point for carrying out specific studies applied to the insurance sector. On the other hand, it would be also necessary to carry out more comprehensive studies which would include all information sources considered by the insured when he decides to underwrite an insurance contract, in the same line as Molina and Blázquez (2005) and Pérez (2007). Other studies, such as Pérez (2006), have proved the importance of other variables, namely the consumer involvement, in order to explain and predict his behaviour. This variable has traditionally received more attention in marketing of tangible products. Such studies are also necessary in the insurance sector.

In this new scenario, relationship marketing is becoming more and more useful as a way how to deal with the increasing competition in the sector and the necessity to recruit new customers and, additionally, to retain them and to increase their loyalty. It is possible to find studies regar-

ding this topic, but they are applied to different type of companies (see, for example, Galguera and Méndez, 2004). On the other hand, the relationship selling has become more and more important due to the benefits of its implementation (see Roman, 2005) basically done by agents and brokers in the insurance industry. Another important change in the sector has been introduced by insurance companies which offer the possibility to underwrite policies through direct means (internet or television). Its benefits and limitations have been studied by Ruiz and Sanz (2007). All these changes in the insurance industry have forced traditional companies to strengthen their presence on the internet.

As a summary, in the marketing literature there are not many articles analysing customer loyalty in insurance companies, and most of them deal with the estimation of the probability of cancellation of one particular insurance policy (Crosby and Stephens, 1987, Schlesinger and Schulenburg, 1993). This is a limitation, because the customer is considered as the one who underwrites each single contract, but actually he/she can have more than one policy in the same company. Therefore, if we want to consider the customer of the insurance company (not just the policy holder or the individual who signed that particular type of contract) we have to consider every single policy he/she may have underwritten with the company and analyse the relationship between the insurer and the customer in all its dimensions. The existing marketing literature has not focussed much on insurance related issues, possibly because of the nature of the product. Insurance contracts are intangible goods and the buyer/seller are linked by a contract that stipulates the terms and conditions for risk coverage and economic compensation over a certain period of time.

Additionally, the estimation of the probability of a policy cancellation provides us with a short term view of the relationship between the insurance company and the customer. Actually, apart from knowing whether or not the customer is going to renew the policy in the next due date, we would like to have an approximation of the customer lifetime duration, i.e. his/her duration as a client.

This paper makes a contribution in both aspects: we consider different types of policies the customer may have underwritten with the same insu-

rance company and we estimate the time he/she is going to stay in the company. The analysis is illustrated with a policy cancellation dataset provided by a Danish insurer (Perez-Marin, 2006). Namely, we estimate the probability that a household with more than one policy with the same insurance company would decide to cancel all of them simultaneously. Secondly, for those who cancel part of their policies, but not all of them, we analyse the customer lifetime duration after the first policy cancellation by using a new non parametric estimator.

2. Background

2.1. Customer loyalty and lifetime duration

It was in the 50's when firms started to be interested in the reasons why customers are choosing a particular product or brand. The behavioural concept of loyalty was introduced by Brown (1952). According to his definition, customer loyalty is a tendency to buy one brand and it is directly related to the frequency of purchase. Nevertheless, many authors were not satisfied with a pure behavioural concept of loyalty and they included a positive attitude towards the brand in the definition of loyalty (Day, 1969; Jacoby and Chestnut, 1978). Nowadays, the idea that customer loyalty has both a behavioural and attitudinal component is widely accepted. Moreover, in recent years new factors such as sensitivity or emotions towards the brand (Fourier and Yao, 1997) and stochastic elements (Uncles and Laurent, 1997) have been considered.

It is also well accepted that people grow into loyal customers by following a step-by-step progression (Griffin, 2004). Murray (1988) was the first to introduce a scale, and he proposed five levels of loyalty: prospects, shoppers, customers, clients and advocates.

Reinartz and Kumar (2003) give a brief review of the major findings of studies concerned with customer lifetime duration modelling. Firstly, the authors stress the limitations of several empirical studies (Allenby, Leone and Jen, 1999; Bolton, 1998; Dwyer, 1997; Schmittlein and Peterson, 1994) due to the general lack of customer purchase history data. Nevertheless, during last years there is an increasing availability of longitudinal

customer databases and researchers have started to take a longitudinal perspective into their work (Reinartz and Kumar, 2003). Regarding the methodology, in some of these studies the proportional regression model Cox (1972) is used, for example Li (1995) and Bolton (1998). Helsen and Schmittlein (1993) supported the superiority of these methods when handling duration-type data. Other methodologies are also applied, such as, for example, the Tobit regression model and Bayes models (Thomas, 2001; Allenby, Leone and Jen, 1999).

2.2. Applications to the insurance sector

Very few applications to the insurance market can be mentioned. Crosby and Stephens (1987) model satisfaction with the service provider in the context of life insurance. Their results suggest that nonlapsing customers (those who do not cancel contracts) report higher satisfaction than lapsed customers, but insureds were followed during a few months only.

Demand-side influences have been addressed by different authors but always considering just one type of insurance product (Schlesinger and Schulenburg 1993; Ben-Arab, Brys and Schlesinger; 1996; Kuo, Tsai and Chen, 2003; Wells and Stafford, 1995; Stafford, Stafford and Wells, 1998), investigated consumer perceptions of service quality. The demand of insurance products in the presence of specific risks factors has been investigated by Doherty and Schlesinger (1983), Schlesinger and Doherty (1985) and Gollier and Scharmure (1994) among others.

Guillén, Parner, Densgsoe and Perez-Marin (2003) considered more than one type of policy simultaneously and they estimated the probability of a policy cancellation in a three-month period for those customers in an insurance company having at least one of these three types of non-life insurance contracts: content of the house, house (the building itself)[1] and motor insurance. This research work identified some factors that are associated to a higher risk of a policy cancellation (such as the existence of recent claims and a premium increase) and confirmed the important of considering different types of policies simultaneously.

In the present paper we estimate customer lifetime duration by considering simultaneously different types of policies the customer may have

(1) The contents of the house and the house itself are insured by using two different policies because the dataset used in this study has been provided by a Danish insurer, and in Denmark these risks are insured by using two different contracts.

with the same insurer (contents, house and motor insurance), following the same ideas as Guillén, Parner, Densgsoe and Perez-Marin (2003) but extending them to understand the cancellation process more extensively. In the next section, we describe the specific conceptual framework which is the base for the analysis of customer lifetime duration in insurance.

3. Conceptual framework

3.1. The concept of cancellation

In the insurance setting, when a policy is ended two basic situations are possible: (1) the risk is going to be covered by another insurance company (e.g., an automobile insurance policy is taken out by another insurance company), or (2) the risk does not exist any more for the policy holder (e.g., a car being sold).

It is very important to distinguish between these two situations in order to decide what a policy cancellation is in our study. The first type of policy termination (brand switching via the customer purchasing from another company) is the termination of interest to understanding demand-side market dynamics and customer relationships. Therefore, the rule that we have applied to determine whether a termination is regarded as a cancellation or lapse is whether the risk still exists at the time the contract is ended.

3.2. The moment when the policy cancellation occurs

A customer normally notifies that he/she does not want to renew the policy several months before the due date. Once the notification has occurred the contract continues in force until the renewal date, when the policy termination is effective and the risk is not covered any more. Therefore, there is a period of time, of possibly several months, since the customer expresses this intention of finishing the contract until it is actually effective. It is clear that in our analysis we should consider the notification date as the moment when the policy is cancelled, because the policy termination date just identifies the moment when the risk is not covered any more.

3.3. The household as the individual in our analysis

Normally families decide to buy different types of insurance contracts in order to be covered against their common risks (those which can affect all members of the family). Very often these policies are underwritten with the same insurer even though they can be signed by different members of the family. Nevertheless, all adult members of the household usually make decisions together about cancelling or underwriting new policies. Therefore, the individual in our study is not the particular policyholder signing the contract but the family or household he/she belongs to.

4. Study design

4.1. The dataset

The dataset used in this research consists of 151,290 households having multiple insurance policies, who sent notification of cancellation of their first policy to a particular major Danish insurer between January 1, 1997 and June 1, 2001. The information was collected according to the time frame shown in Figure 1.

Once the first policy cancellation occurs, the residual household customer lifetime is measured by the number of days until all remaining policies are notified for cancellation or until the end of the study, June 1, 2001, whichever comes first (some policyholders will cancel one policy but keep others). In that second case, the lack of information is called right censoring and the statistical analysis has to account for this phenomenon.

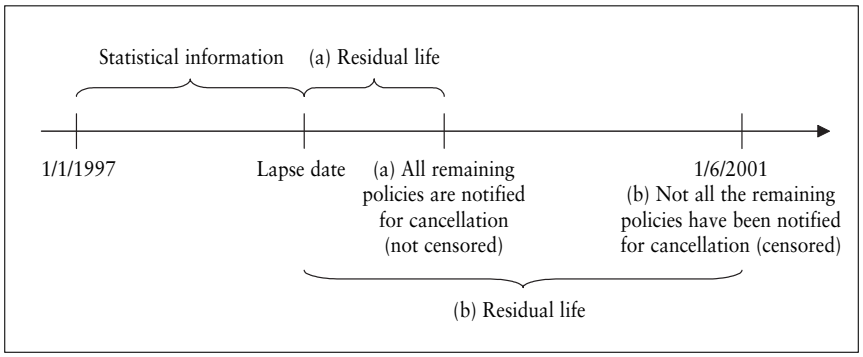
In situation (a) in Figure 1, all the remaining policies are cancelled before June 1, 2001, so the household customer residual life is the time from the first lapse date until total cancellation of all other policies occurs. In situation Figure 1 (b), at the end of the study, we only know that the residual life is greater than the time from the first lapse until June 1, 2001. In this case, the residual life is the listed as the time elapsed from first policy cancellation until June 1, 2001, but note that the observation is right censored.

Some of the household covariates refer to the occurrence of an event (a claim, a premium increase, or a change of address) from January 1, 1997

until the date of the first lapse while other covariates are measured at the time of first policy cancellation (for example, the tenure or the age of the policy holder).

Table 1 lists the variables in the database and the label given to teach.

Figure 1. Temporal frame



Variable "tenure" is the number of years the household has been a customer of the company calculated as the number of years from the first policy issued to the policy holder, within the types of policies considered here, until the date of the first lapse. Variable "notice" indicates the time interval from notification of the first policy cancellation until the actual occurrence of the corresponding cancellation.

Since the types of policies held by the household could conceivably affect the retention attributes of the client with respect to the insurer, the following dummy variables were created: "contents0", "house0" and "motor0". They indicate whether the household has contents, house, or automobile policies respectively before the first lapse. Variables "contents1", "house1" and "motor1" indicate whether the household has contents, house, or automobile policies respectively after the first lapse. Variables "newcontents", "newhouse" and "newmotor" indicate whether or not the household has underwritten the first contents, house, or automobile policies, respectively, within the 12 months prior to the date of the first lapse.

Table 1. Explanatory variables

Age of the customer when the first policy is cancelled ("Age")
Gender of the customer ("Gender")
Number of days elapsed between the notification of the cancellation and the day when the policy is no longer in force ("Notice")
Tenure of the customer ("Tenure")
Customer with special advantages. Apart from the contents policy he also has other two types of policies in the same company ("Advantages")
Change of address before the first cancellation ("Address", six subcategories)
First Cancellation notice furnished by external company A ("External Company A")
First Cancellation notice furnished by external company B ("External Company B")
First Cancellation notice furnished by external company C ("External Company C")
First Cancellation notice furnished by external company D ("External Company D")
First Cancellation notice furnished by another known external company ("Another Known External Company")
Claims ("Claims", six subcategories)
Contents policy before the first cancellation ("Contents0")
Contents policy after the first cancellation ("Contents1")
House policy before the first cancellaton ("House0")
House policy after the first cancellaton ("House1")
Motor policy before the first policy cancellation ("Motor0")
Motor policy after the first policy cancellation ("Motor1")
Indicator of household having underwritten the first contents policy within the 12 months previous to the date of the first lapse ("Newcontents")
Indicator of household having underwritten the first house policy within the 12 months previous to the date of the first lapse ("Newhouse")
Indicator of household having underwritten the first motor policy within the 12 months previous to the date of the first lapse ("Newmotor")
Premium increase ("Premium increase", broken into three subcategories)

Variable "advantages" indicates whether the customer has a core customer status. A core customer is a customer that has a contents policy and at least two other types of policies (they could be automobile, house, or others like life insurance) with the insurer. In the insurance company that has been analyzed here, core customers have lower premiums, bonuses, and special advantages. From a marketing perspective core customers having multiple policies tend to be more profitable and, hence, deserve special consideration.

Information on whether a change of address has occurred was included, as it can affect the probability of house and contents cancellations. When a family buys a house, the financial institution in charge of the mortgage normally tries to persuade the customer to underwrite the contents and house policies with their insurance company or the one with whom they have some kind of commercial agreement. Six categories were developed for this variable: no change of address, change of address less than 2 months before the date of the first lapse, between 2 and 6 months before the date of the first lapse, between 6 and 12 months before the date of the first lapse, between 12 and 24 months before the date of the first lapse, and more than 2 years before the date of the first lapse. The same categories have been considered for the variable claims, which indicate whether or not there has been any claim during the considered time-period.

Since premium increases might impact customer retention, information was included on whether the time period included a substantial increase in premium of 20 to 50%. Such premium increases are commonly termed pruning, since the insurer wants to persuade the customer to lapse, possibly due to a very bad claims history. Three categories were developed: no pruning, pruning within the past 12 months, and pruning more than one year before the date of the first lapse.

Finally, considering the competitive nature of the marketplace and the marketing dynamics of alternative brands in a brand switching model, we have also included information on whether there was any external company involved in the cancellation notice. The customer has a choice of notifying the current insurer him/herself of cancellation or of having the new insurer notify the current insurer. It is clear that when the new insurer

does the notification, that a brand switch has already occurred and, at least for that policy, the customer is entrenched with the new insurer for at least the next year. It is likely, also, that the new insurer will wait until the last moment to signal their competitor of the upcoming brand switch, lest the competitor take measures to try to retain their customer. Further, the new insurer will likely be discussing other insurance policy needs with their newly acquired customer, so subsequent policy cancellations are likely. We considered the four most important competitors, coded as A, B, C and D and developed six categories for this variable: no external company (notification by the customer himself), company A, company B, company C, company D, and finally another known external company. We considered a competitor to be involved if the notification was communicated by an insurance company on behalf of the customer.

Table 2 presents a description of the policy portfolio state before versus the state after the first lapse, thus comparing the types of policies the household had before and after the occurrence of the first lapse. This information is represented with a string of three characters of 0's and 1's where 1 (0) indicates that the household had (had not) one particular type of policy. The sequence order is contents - house - automobile. For example, if the state before the first lapse is 011 and the state after the first lapse is 010, then the household had house and automobile policies before the first lapse, but no automobile policy after the lapse.

As shown in Table 2 the most frequent state before the first lapse is 111 (the customer has the three types of policies), while the most frequent state after the first policy cancellation is 000. We also observe that 77,337 (51.20%) households have more than one policy before the first policy and 73,953 households (48.80%), just one. Among those with more than one policy, 10,642 (13.76%) make a total cancellation, while 66,695 (86.24%) make a partial cancellation.

Table 2. Policies before vs. after the first cancellation (absolute frequency)

Policies before* \ Policies after*	000	100	010	001	110	101	011	111	Total
000	0	0	0	0	0	0	0	0	0
100	34998	0	0	0	0	0	0	0	34998
010	10757	0	0	0	0	0	0	0	10757
001	28198	0	0	0	0	0	0	0	28198
110	2690	3060	4090	1	0	0	0	0	9841
101	3397	13764	1	10613	0	0	0	0	27775
011	166	1	1409	1042	0	0	0	0	2618
111	4389	471	2488	4535	12488	5957	6775	0	37103
Total	84595	17296	7988	16191	12488	5957	6775	0	151290

*The underwritten policies are represented by a string of 0's and 1's where 0 denotes presence of the policy and 1 indicates the absence of the policy type, with the ordering of the state triad being contents insurance - house - automobile.

4.2. Methodology

Our modelling process includes two stages. Firstly, we consider those households with more than one policy in the insurance company. Some of them would cancel all their policies simultaneously (total cancellation) and some of them would make a partial cancellation. For those households making a total cancellation the insurer has no time to react after this first cancellation. For those households who make a partial cancellation the insurer can estimate the remaining lifetime (the time between the first cancellation and the moment when all the remaining policies would be cancelled). Therefore, the modelling process includes a first step where the probability of a total cancellation is estimated for those households with more than one policy in the insurance company. A logistic regression model will be used to estimate this probability.

In the second stage we focus on those households who made a partial cancellation. The risk that all the remaining policies would be cancelled (therefore, the insurer loose the customer) and the customer lifetime is estimated as a function of some covariates by using the proportional hazards regression model and a new non parametric estimator.

a) Predicting the probability of a total cancellation

In the first stage, we use logistic regression to determine the probability (based upon explanatory covariates) that a household originally having more than one policy will cancel all the policies simultaneously. For household *i*, in = 1,..., we assume that

$$P(R_i = 1 \mid x_i) = \frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)} \quad (1)$$

where $R_i = 0$ for a partial cancellation and $R_i = 1$ for a total cancellation, x_i is a vector of the observed explanatory variables, β is a vector of unknown parameters. Consistent and asymptotically efficient estimates of the parameters in the logistic regression model (1) are obtainable using the conditional maximum likelihood method (Snell and Cox, 1989; Agresti, 1990) implemented in many common statistical packages. In this manner we are able to look at the effect of household characteristics (covariates) on the likelihood of total simultaneous cancellation.

b) Customer lifetime duration

In order to estimate the customer lifetime duration for those customers who make a partial cancellation we apply a proportional hazards regression model. According to that model, the hazard for a random survival time *T*

$$\alpha(t) = \lim_{dt \to 0} \frac{P(t \le T < t + dt \mid T \ge t)}{dt} \quad (2)$$

is the product of function which depends on time, the baseline hazard $\alpha_0(t)$, and an exponential function of the covariates, $\alpha(t \mid z_i) = \alpha_0(t)\exp(\beta' z_i)$where z_i is the column vector p dimensional of explanatory variables corresponding to the *i*th individual and β is the column vector of unknown parameters. The baseline hazard represents the instantaneous risk when all covariates are equal to zero.

If we consider two individual with covariates z_0 y z_1, the ratio of their hazards is constant (exp $[\beta'(z_0 - z_1)]$) over time. This is the reason why this model is called the proportional hazards regression model. The hazard $\alpha(t)$

can be used to calculate the survival function $S(t) = 1 - F(t)$, where $F(t)$ is the distribution function, according to

$$S(t) = \exp\left(-\int s_0 \alpha(s) ds\right). \quad (3)$$

The lifetime duration for a particular customer can be obtained by integrating the corresponding survival function. The vector of parameters β is obtained maximizing the partial likelihood function without previously specifying the baseline hazard (Efron, 1977).

Most of the covariates in our application are binary and can be understood as indicators of the presence of a risk factor (for example, a change of address or a claim). The sign of the parameter estimate can be interpreted as the effect of the corresponding covariate on the expected time to final withdrawal from the company. When the parameter estimate is positive, we conclude that the hazard for the household with the associated covariate (risk factor) is larger than in the absence of the indicator of this covariate. On the basis of proportionality, the corresponding resulting survival function is also steeper. Thus, a positive parameter estimate is associated to a shorter time to total withdrawal for those households that have the risk factor signalled by the covariate, compared to those without the risk factor. The interpretation is exactly the opposite in case that the parameter is negative.

In order to obtain the survival function for each customer, it is necessary to estimate the corresponding baseline hazard function. In order to do that, we have used a modification of the Nelson-Aalen estimator (Aalen, 1978; Nelson, 1969; Nelson, 1972), called *naive local constant estimator*. This estimator was introduced by Guillen, Nielsen y Perez-Marin (2007) and has a better efficiency that other traditional non-parametric estimators (namely, the Nelson-Aalen estimator) used in survival analysis. On the basis of the original formulation of the *naive local constant*, here we will use the following reformulation of this estimator in order to adapt it to the estimation of the baseline hazard function in the Cox model

$$\hat{\Lambda}_{NLC}(t) = \sum_{t_j=0}^{max(t-b,0)} \frac{d_j}{\sum_{l \in R_j} exp(\beta' z_l)} + \sum_{t_j=max(t-b,0)}^{t+b} \frac{d_j}{\sum_{l \in R_j} \gamma_{t,b} \, exp(\beta' z_l)} \quad (4)$$

where $t_1 < t_2 < ... < t_D$ denote the D different ordered survival times, d_j the total number of cancellations in t_j, R_j the set of all households who are at risk just prior to time t_j, b is the bandwidth parameter and $\gamma_{t,b}$ is a constant given by $\gamma_{t,b} = \{t + b - \max(t - b,0)\}/\{t - \max(t - b,0)\}$. Guillen, Nielsen and Perez-Marin (2007) proved that the value of b providing the maximum efficiency gain with respect to the Nelson-Aalen estimator is given by

$$b_{opt} = \left\{ \frac{\alpha(t)}{2Y(t)\alpha'(t)^2} \right\}, \qquad (5)$$

where $Y(t)$ represents the risk exposure at time t. In the boundary region, the optimal solution is $b_{\text{opt}} = t$.

5. Results

We consider the subset of customers with more than one policy before the first cancellation, 77,337 households. We firstly apply a logistic regression model to this subset in order to estimate the probability of a total cancellation. Among those 77,337 households, 66,695 of them make a partial cancellation. For that second subset of customers we estimate the customer lifetime duration applying a Cox model. The results for both models are presented in this section.

5.1. Estimation of the probability of a total cancellation

The covariates described in Table 1 are used, except for *contents1*, *house1* and *motor1* which are, of course, all zero after a total cancellation has occurred. The data set used in the estimation of the model consists of 74,969 households (a few observations were eliminated due to missing values on some covariates). Among those 74,969 observations, 10,317 simultaneously effected a total cancellation of all policies with the insurer.

The overall statistical test of no covariate effect provided a likelihood ratio statistic of LR = 9,178 with 28 degrees of freedom ($p<.001$), therefore these results support the overall significance of the model. Individual parameter estimates are shown in Table 3. All of the parameters in the model are significant, except for having added a new house policy in the

last 12 months (*newhouse*) and having had a premium increase more than one year prior to the household first giving a cancellation notice (*pruning more than one year past)*.

Table 3. Logit model. Estimation

Parameter	Estimation	Std. Error	OR	p-value
Constant	-2.2010	0.1117	-	<.0001
Change of address, less 2 m. ago	-0.5956	0.0596	0.551	<.0001
Change of address, 2 - 6 m. ago	-0.1219	0.0519	0.885	0.0189
Change of address, 6 - 12 m. ago	-0.0951	0.0484	0.909	0.0492
Change of address, 12 - 24 m. ago	0.2294	0.0412	1.258	<.0001
Change of address, more 24 m. ago	0.5312	0.0443	1.701	<.0001
Tenure	-0.0111	0.0014	0.989	<.0001
Claims, less 2 months ago	0.2300	0.0404	1.259	<.0001
Claims, 2 and 6 months ago	0.3244	0.0355	1.383	<.0001
Claims, 6 and 12 months ago	0.4400	0.0346	1.553	<.0001
Claims, 12 and 24 months ago	0.4689	0.0372	1.598	<.0001
Claims, more 2 years ago	0.5463	0.0536	1.727	<.0001
Contents0	0.2773	0.0875	1.319	0.0015
Advantages	0.1096	0.0254	1.116	<.0001
Age	0.0040	0.0009	1.004	<.0001
External company A	2.5478	0.0418	12.78	<.0001
External company B	2.1654	0.0458	8.718	<.0001
External company C	1.8927	0.0479	6.637	<.0001
External company D	2.2700	0.0467	8.834	<.0001
Another known external company	1.6860	0.0353	9.68	<.0001
Gender (male)	0.0933	0.0279	1.104	0.0004
House0	-0.6572	0.0302	0.518	<.0001
Motor0	-1.2532	0.0328	0.286	<.0001
Newcontents	-0.1133	0.0428	0.893	0.0081
Newhouse	0.0728	0.0600	1.076	0.2252
Newmotor	-0.2076	0.0504	0.813	<.0001
Notice	-0.0018	0.0001	0.998	<.0001
Premium increase past 12 months	-0.1875	0.0725	0.829	0.0097
Premium increase + 1 year ago	0.0856	0.1114	1.089	0.4421

By looking at the odds ratios (OR), we see that external companies, claims, change of address more than one year ago are the most relevant factors influencing the probability that a total cancellation occurs. Addi-

tionally, we observe that contents policy is associated to a higher probability of total cancellation.

The tenure of the customer contributes to reduce the probability of a total cancellation, while the age of the customer has the opposite effect. We also observe that men have a higher risk of total cancellation than women.

The most important determinant of the probability of a total cancellation is, however, the intervention of an external company in the first cancellation. We also observe differences among the competitors (the one coded as A is the most aggressive). We would like to stress this last result, because at the time the statistical analysis began, the insurance company that provided us with the data set had not realised that the competitor coded A was involved in so many policy cancellations. Our analysis provided evidence of that flow of clients from one insurer to the competitor.

A surprising result is that being a customer with special advantages increases the probability of a total cancellation. Probably the reason is that, in order to keep these advantages in the new insurance company, the customer is very often required to switch all his policies to the new insurer.

We also observe that "notification" has a negative parameter estimation, therefore, the more in advance the cancellation is notified, the lower the probability of a total cancellation. Similarly, the parameter estimates associated to "newcontents" and "newmotor" are negative (reduce the risk of total cancellation) while "newhouse" is positive, so it increases that probability.

Regarding the ability of the model to discriminate between total and partial cancellations, the results are satisfactory, as it is able to detect 71% of all total cancellations.

5.2. Estimation of the customer lifetime duration

For those households with more than one policy that do not cancel all their policies at the same time, we analyze expected amount of time between the first cancellation occurs and the final termination of all policies with the company by using the Cox model. The likelihood ratio test for the overall significance of the model is high. LR = 32,623.98, which is chi-squared

distributed with 31 degrees of freedom (p<.001), indicating that the covariates have a significant effect on the hazard and thus on the amount of time the insurer has to react to a partial withdrawal. Table 4 displays the parameter estimates of this model.

Table 4. Cox model. Estimation

Parameter	Estimation	St. Error	Risk ratio	p-value
Change of address less 2 m. ago	-0.245	0.019	0.783	<0.001
Change of address 2 - 6 m. ago	-0.083	0.020	0.920	<0.001
Change of address 6 - 12 m. ago	-0.023	0.020	0.978	0.252
Change of address 12 - 24 m. ago	0.044	0.019	1.045	0.021
Change of address more 24 m. ago	0.157	0.025	1.170	<0.001
Tenure	-0.003	0.001	0.997	<0.001
Claims, less 2 m. ago	0.096	0.016	1.100	<0.001
Claims, 2 - 6 m. ago	0.161	0.015	1.175	<0.001
Claims, 6 - 12 m. ago	0.185	0.015	1.203	<0.001
Claims, 12 - 24 m. ago	0.228	0.017	1.256	<0.001
Claims, more 24 m. ago	0.257	0.027	1.294	<0.001
Contents0	0.681	0.029	1.975	<0.001
Contents1	-0.869	0.016	0.419	<0.001
Advantages	-0.042	0.011	0.959	<0.001
Age	-0.003	<0.001	0.998	<0.001
External company A	1.727	0.018	5.625	<0.001
External company B	1.528	0.020	4.611	<0.001
External company C	1.652	0.019	5.217	<0.001
External company D	1.778	0.020	5.919	<0.001
Another known external company	1.643	0.012	5.170	<0.001
Gender (male)	0.103	0.012	1.109	<0.001
House0	0.222	0.017	1.249	<0.001
House1	-0.559	0.017	0.571	<0.001
Motor0	0.415	0.021	1.515	<0.001
Motor1	-0.529	0.018	0.589	<0.001
Newcontents	-0.059	0.016	0.942	<0.001
Newhouse	-0.109	0.024	0.897	<0.001
Newmotor	-0.076	0.017	0.927	<0.001
Notice	>-0.001*	<0.001	1.000	<0.001
Premium increase past 12 months	0.188	0.029	1.207	<0.001
Premium increase + 1 year ago	0.095	0.053	1.100	0.075

(*) Negative parameter estimate

We observe that all parameters associated to the covariates in the model are significant except for the change of address (between 6 and 12 months before the first policy cancellation) and premium increase (more than one year ago). External companies, claims, premium increase (during last year) and change of address are the factors causing the largest reduction in the customer lifetime duration after the first policy cancellation.

The parameter associated to "tenure" is negative and significant, therefore the longer the customer was in the company the longer he is going to stay after the first policy cancellation. We also observe that women have longer customer lifetime duration than men, and that age also contributes to increase that duration.

We also observe that men with special advantages in the company have a longer customer lifetime duration alter the first policy cancellation, while we saw that this factor is associated to a higher probability of total cancellation.

The presence of claims reduces residual life, and the effect becomes more dramatic as the time since the claim has occurred increases. This could be connected to some delay in the compensation of claims that may result in a delay in the assessment of the claims handling process made by the household.

Without any doubts, the factor most significantly related to a reduction in residual life of the client household with the insurer. When a new insurer is involved in the first cancellation, the customer lifetime duration in the first company is dramatically reduced, especially for the external companies coded as D and A.

The parameter estimates associated to "newcontents", "newhouse" and "newmotor" are negative, therefore recent business is contributing to an increase residual lifetime duration after the first policy cancellation.

The parameter associated with "notice" is negative. This indicates that the sooner the notification is made the longer the customer lifetime duration is after the first policy cancellation.

Finally, regarding the covariates describing the composition of the customer portfolio before the first policy cancellation, the one associated to the lowest reduction of the customer lifetime duration is "house0",

"motor0" and finally "contents0". Regarding the composition of the customer portfolio after the first policy cancellation, the variables associated to the highest increase of the customer lifetime duration are "contents1", "house1" and finally "motor1".

5.3. Examples of survival functions for different customers

We apply the methodology described above to estimate the survival function and lifetime duration of different types of customers.

In the first example, we illustrate how important is to consider different types of policies simultaneously. We consider a 55 year-old male customer, with ten years of tenure with the insurer, no change of address within the last two years, a claim between 2 and 6 months ago, and no external company involved in the notification, giving 150 days of notice before renewal, no new business with the insurer within the past twelve months, no special advantages, no premium increase and just with the contents policy after the first cancellation. In Figure 2 we compare the survival function for that particular customer depending on the policies he had before the first cancellation. As expected, the survival curve with the steepest slope, and the lower estimation of the customer lifetime duration, corresponds to the case when the customer had the three policies initially and cancels two of them simultaneously (662 days).

In the second example, we consider the same customer but with contents and motor policies before the first cancellation and just the motor policy after the first cancellation (he cancelled the contents). In that case, the survival functions and the lifetime durations are compared depending on whether or not there has been any external company involved in the first cancellation. The results are shown in Figure 3. External company coded D is the one associated with the largest reduction in the customer lifetime duration (just 45 days). On the other hand, when there is no external company involved in the first cancellation, the customer lifetime duration in substantially larger (600 days).

Figure 2. Survival functions. Comparison depending on the policies being cancelled

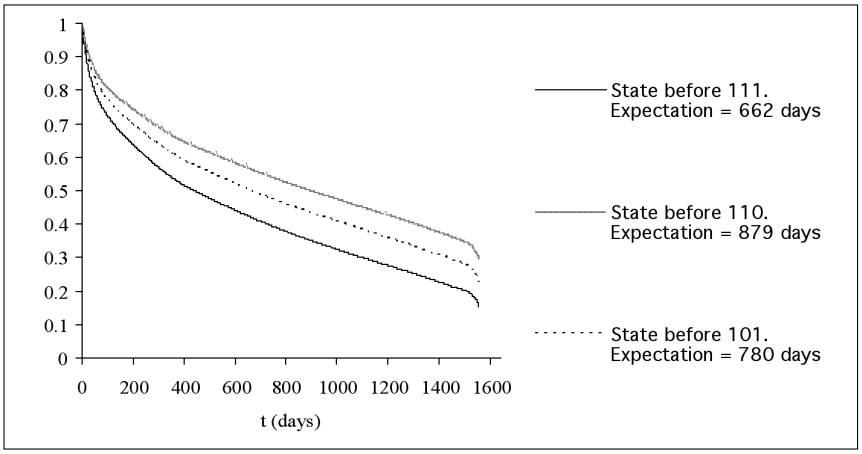
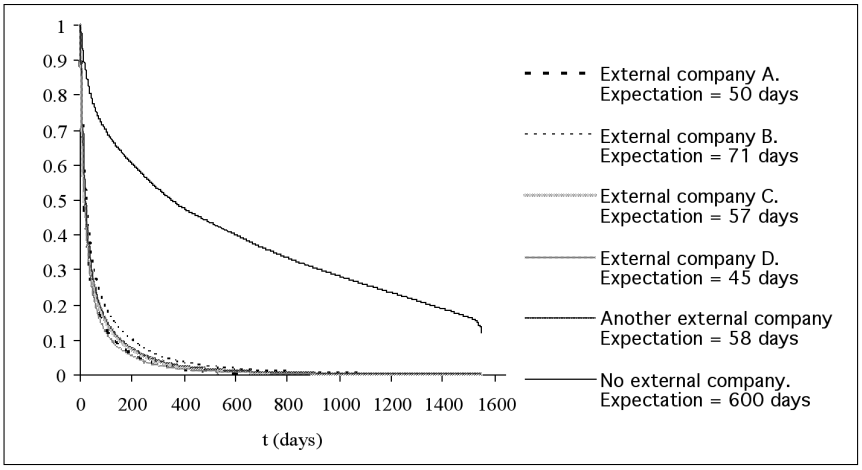


Figure 3. Survival functions. Comparison for different external companies



5.4. Estimation of survival probabilities

Finally, we applied the proposed methodology to estimate the probabilities of cancellation in a given period of time. The objective is to evaluate the model's ability to detect customers with a high probability to completely leave the company within short time periods: 3 and 6 months. The estimations of the probabilities for each individual in the dataset are compared with what we actually observe for him (see Tables 5 and 6).

Table 5. Predicted probabilities for the three-month period

	Observed customer lifetime duration (c.l.d.)		
p = P(c.l.d. < 3 months)	c.l.d. > 3 months	c.l.d. ≤ 3 month	Total
$p < 0.25$	19929	4559	24488
$0.25 \leq p < 0.5$	4162	2662	6824
$0.5 \leq p < 0.75$	6433	13099	19532
$p \geq 0.75$	1853	8228	10081
Total	32377	28548	60925

Table 6. Predicted probabilities for the six-month period

	Observed customer lifetime duration (c.l.d.)		
p = P(c.l.d. < 6 months)	c.l.d. > 6 months	c.l.d. ≤ 6 month	Total
$p < 0.25$	11664	3606	15270
$0.25 \leq p < 0.5$	8328	6197	14525
$0.5 \leq p < 0.75$	2946	7720	10666
$p \geq 0.75$	3129	17335	20464
Total	26067	34858	60925

Regarding the results corresponding to the three-month period, we observe that, among those who actually cancel all the remaining policies in that period, 28,548 customers, 74.7% have a probability equal or higher than 50% of cancelling them, according to the model. Regarding those who continue in the company more than three months, 32,377 customers, 74.4% have an estimated probability lower than 50% according to the model.

The same probabilities are estimated for the six-month period, the results are shown in Table 6. Among those who actually cancel all the remaining policies in that period of time, 34,858 customers, 71.88% have an estimated probability equal or higher than 50%. For those who actually continue more than six months, 26,067 customers, 76.7% have an estimated probability lower than 50% according to the model.

Therefore, we conclude that the model proposed here detect reasonably well hose customers with a high probability of cancelling all their remaining policies in a short period of time.

6. Conclusions

This research leads us to a number of conclusions, from both the business and the academic perspective.

From the academic perspective, it is necessary to remark the methodological contribution of this research in order to analyse the loyalty and the cross-buying behaviour of the customers in the service sector. In this study we consider customers who may have different types of policies underwritten with the same insurer. This research also makes a contribution to the investigation of the insurance sector, which is valuable due to the lack of specific studies which would be focused on the particularities of this sector.

This analysis includes two stages where we identify the factors with a negative influence on the insurance customer loyalty: the estimation of the probability that the customer would decide not to renew the contract (total cancellation) and the study of the customer lifetime duration after the first time the client decides not to renew the contract (the first policy cancellation). It is important to remark that in this second stage we use for the first time one reformulation of the *naïve local constant* estimator (introduced by Guillén, Nielsen and Pérez-Marín, 2007) for estimating the baseline hazard in the Cox model in order to approximate the customer lifetime duration. Therefore, we propose an alternative methodology to those traditionally used in marketing which introduce an efficiency improvement with respect to the classical estimators (proved by Guillén, Nielsen and Pérez-Marín, 2007).

Certainly, the procedure which has been applied in this research is suitable for analysing customer loyalty in the insurance sector. Regarding the prediction performance of the models, the method identifies a high percentage (more than 70%) of all total cancellations and all households which cancel all the remaining policies in a short period of time (3 or 6 months) after their first cancellation.

Finally, as we mentioned before, customer loyalty is crucial in order to control the business risk assumed by the insurance company. This research can be considered the starting point in order to approximate the distribution of losses which are caused by the competitive environment where the company operates and, later on, define suitable risk measures.

From the business perspective, this research has identified the factors having a more remarkable effect on customer loyalty. As a summary, we conclude that external companies being involved in the first policy cancellation is the factor with the highest impact on the risk of total cancellation. Similarly, the occurrences of a claim and a premium increase (more than one year ago) are associated to a higher probability of total cancellation. We also observe differences when considering different types of policies the customer may have, and we see that having a contents policy is associated to a higher risk of total cancellation. Additionally, and contrary to what we expected, customers with special advantages in the company are among those with a higher probability of total cancellation.

Regarding the second stage, again external companies reduce very dramatically customer lifetime duration after the first policy cancellation. Claims and premium increase have the same influence on customer lifetime duration, but not so intensively as external companies. On the other hand, those with special advantages have a longer customer lifetime duration, but they have a higher probability of a total cancellation. Finally, the different composition of the customer portfolio before and after the first policy cancellation is a relevant factor in order to explain the customer lifetime duration after that first cancellation.

Therefore, from these results we can establish some recommendations for insurance companies in order to increase customer loyalty. The first one would be to detect the competitors which more successfully attract

their customers and to know the reasons for their success. In his way, it would be possible to address loyalty strategies as soon as these customers announce that they are going to move the first policy to that competitor, or even before they would do it.

It is also necessary to pay special attention to those customers who have one contents policy and/or special advantages in the company, as these customers have a higher probability of making a total policy cancellation. As a conclusion, it seems that offering special advantages to the customers is not an effective loyalty strategy, therefore it is necessary to apply additional actions for these customers. On the other hand, the occurrence of claims and premium increases negatively affect loyalty and it is necessary to identify these circumstances in order to compensate their effects.

Apart from detecting these groups of customers who have a higher risk of cancellation, it would be very useful for the company to establish (based on the methodology proposed in this research) a ranking of the customers in the portfolio based on their risk of cancellation. Based on that ranking, it would be possible to address specific actions to these customers with a high risk and with a high value for the company. This ranking can be updated periodically, or each time a relevant event affecting customer loyalty (claims, premium increase, …) would take place.

Moreover, by calculating the survival function for each customer the insurer is able to approximate the evolution of his loyalty in order to predict his behaviour and avoid losing him. The procedure described here can be implemented in the company, in such a way that each time a cancellation occurs, the remaining lifetime duration of that customer (in case that he still keeps other policies in the company) would be calculated as a measure of the time the insurer has in order to react and try to keep that customer. This information, together with the customer value, would let the company to decide which would be the best strategy to apply in each case.

All these recommendations can be used in order to improve customer loyalty and manage business risk in the insurance company. Nevertheless, as we mentioned before, it is necessary to carry out specific studies in order to have a more detailed knowledge of customer loyalty in the insurance

sector. These studies should take into account a number of topics, including the information sources used in the decision-making process of the insured, the role of relationship selling, the possibility of underwriting policies by using direct means and the role of the internet inside the relationship marketing in the insurance company, among many others.

7. References

AALEN, O. O. (1978). "Nonparametric Inference for a Family of Counting Processes," *Annals of Statistics*, Vol. 6, pp. 701-726.

AGRESTI, A. (1990). *Categorical Data Analysis*. Wiley, New York.

ALLENBY, G. M., LEONE, R. P. and JEN, L (1999). "A dynamic model of purchase timing with application to direct marketing", *Journal of the American Statistical Association* Vol. 94, No. 446, pp. 365-374.

BEN-ARAB, M., BRYS, E. and SCHLESINGER, H. (1996). "Habit Formation and the Demand for Insurance", *Journal of Risk and Insurance*, Vol. 63, No. 1, pp. 111-119.

BOLTON, R. N. (1998). "A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction", *Marketing Science*, Vol. 17, No. 1, pp. 45-65.

BROWN, G. H. (1952). "Brand loyalty - fact or fiction?" *Advertising Age*, Vol. 9, pp. 53-55.

COX, D. R. (1972). "Regression Models and Life Tables", *Journal of the Royal Statistical Society B*, Vol. 34, pp. 187-220.

CROSBY, L. A. and STEPHENS, N. (1987). "Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry", *Journal of Marketing Research*, Vol. 24, No. 4, pp. 404-411.

DAY, G. S. (1969). "A two dimensional concept of brand loyalty," *Journal of Advertising Research*, Vol. 9 (September), pp. 29-36.

DOHERTY, N. A. and SCHLESINGER, H. (1983). "Optimal Insurance in Incomplete Markets," *Journal of Political Economy*, Vol. 91, pp. 1045-1054.

DWYER, F. R. (1997). "Customer lifetime valuation to support marketing decision making," *Journal of Direct Marketing*, Vol. 11 (Fall), pp. 6-13.

EFRON, B. (1977). "The Efficiency of the Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association*, Vol. 72, pp. 557-565.

FLAVIÁN, C. and GUINALÍU, M. (2007). "Development and validation of familiarity, reputation and loyalty scales for internet relationships", *ESIC Market,* Vol.126, pp. 157-188.

FOURIER, S. and YAO, J. L. (1997). "Reviving brand loyalty: a reconceptualization within the framework of customer-brand relationships", *International Journal of Research in Marketing*, Vol. 14, No. 5, pp. 451-472.

GALGUERA, L. and MÉNDEZ, M. P. (2004). "An empirical approache to relationship marketing:air companies' loyalty cards", *ESIC Market,* Vol. 119, pp. 195-222.

GOLLIER, C. and SCARMURE, P. (1994). "The Spillover Effect of Compulsory Insurance", *Geneva Papers on Risk and Insurance Theory*, Vol. 19, No. 1, pp. 23-34.

GRIFFIN, J. (2004). *Customer Loyalty*, Second Edition. Josey-Bass, New York.

GUILLEN, M., NIELSEN, J. P. and PEREZ-MARIN, A. M. (2007). "Improving the efficiency of the Nelson-Aalen estimator: the naive local constant estimator", *Scandinavian Journal of Statistics,* Vol. 34, 2, 419-431.

GUILLEN, M., PARNER, J., DENSGSOE, C. and PEREZ-MARIN, A. M. (2003). "Using logistic regression models to predict and understand why customer leave an insurance company" in *Intelligent and other Computational Techniques in Insurance. Theory and Applications*, Lakhmi Jain and Arnold Shapiro eds. World Scientific, pp 465-490.

HELSEN, K. and SCHMITTLEIN, D. C. (1993). "Analysing duration times in marketing: evidence for the effectiveness of hazard rate models", *Marketing Science*, Vol. 11 (Fall), pp. 395-414.

JACOBY, J. and CHESNUY, R. (1978). *Brand loyalty: measurement and management*. Willey, New York.

KUO, W., TSAI, C. and CHEN, W. -K. (2003). "An Empirical Study on the Lapse Rate: The Cointegration Approach", *Journal of Risk and Insurance*, Vol. 70, No. 3, pp. 489-501.

LI, S. (1995). "Survival analysis", *Marketing Research*, Vol. 7 (Fall), pp. 17-23.

MARTÍN, J. and QUERO, J. M. (2004). "The website: a relationship marketing tool?", *ESIC Market,* Vol. 199, pp. 117-129.

MOLINA, A. and BLÁZQUEZ, J. J. (2005). "The role of information sources in touristic consumers' choice", *ESIC Market*, Vol. 120, pp. 243-262.

MURRAY, R. (1988). "Up the loyalty ladder", *Direct Marketing*, December.

NAKADA, P., SHAH, H., KOYLUOGO, H.U. and COLLIGNON, O. (1999). "P&C RAROC: A catalyst for improved capital management in the property and casualty insurance industry", *The Journal of Risk Finance*, Fall, pp. 1-18.

NELSON, W. (1969) "Hazard plotting for incomplete failure data", *Journal of Quality Technology*, Vol. 1, pp. 27-52.

NELSON, W. (1972). "Theory and applications of hazard plotting for censored failure data", *Technometrics*, Vol. 14, pp. 945-965.

PÉREZ, C. (2006). "Consumer involvement in goods and service purchases", *ESIC Market*, Vol. 123, pp. 73-91.

PÉREZ, C. (2007). "A comparative study on the external search for goods and services", *ESIC Market*, Vol. 127, pp. 27-43.

PEREZ-MARIN, A. M. (2006). *Survival analysis methods for analysing customer lifetime duration in insurance*. PhD Dissertation. University of Barcelona. [http://guillen.eco.ub.es/~montse/perezmarinPhD.pdf].

REINARTZ, W. J. and KUMAR, V. (2003). "The impact of customer relationship characteristics on profitable lifetime duration". *Journal of Marketing*, Vol. 67, pp. 77-99.

ROMÁN, S. (2005). "Conceptualization, implementation, benefits and limitation of relationship selling", *ESIC Market*, Vol. 121, pp. 169-192.

RUÍZ, C. and SANZ, S. (2007). "Buying throught direct methods: benefits and limitations from the consumer point of view", *ESIC Market*, Vol. 126, pp. 95-123.

SCHLESINGER, H. and DOHERTY, N. A. (1985). "Incomplete Markets for Insurance: An Overview", *Journal of Risk and Insurance*, Vol. 52, pp. 402-423.

SCHLESINGER, H. and SCHULENBURG, J. M. (1993). "Customer Information and Decisions to Switch Insurers", *Journal of Risk and Insurance*, Vol. 60, No. 4, pp. 591-615.

SCHMITTLEIN, D. C. and PETERSON, R. A. (1994). "Customer base analysis: an industrial purchase process application", *Marketing Science*, Vol. 13, No. 1, pp. 41-67.

SNELL, E. J. and COX, D. R. (1989). *Analysis of Binary Data*. Chapman and Hall, London.

STAFFORD, M. R., STAFFORD, T. F. and WELLS, B. P. (1998). "Determinants of Service Quality and Satisfaction in the Auto Casualty Claims Process", *Journal of Services Marketing*, Vol. 12, No. 6, pp. 426-40.

THOMAS, J. S. (2001). "A methodology for linking customer acquisition to customer retention," *Journal of Marketing Research*, Vol. 38, No. 2, pp. 262-268.

UNCLES, M. and LAURENT, G. (1997) Editorial. *International Journal of Research in Marketing*, Vol. 14, No. 5, pp. 399-404.

WELLS, B. P. and STAFFORD, M. R. (1995). "Service Quality in the Insurance Industry. Customer Perception versus Regulatory Perceptions", *Journal of Insurance Regulation*, Vol. 13, No. 4, pp. 462-477.