Modeling the Number of Insureds' Cars Using Queuing Theory

JEAN-PHILIPPE BOUCHER and GUILLAUME COUTURE-PICHÉ

April 22, 2015

‡ *Quantact / Département de mathématiques, UQAM. Montréal, Québec, Canada.*

Email: `boucher.jean-philippe@uqam.ca`

Abstract

In this paper, we propose to model the number of insured cars per household. We use queuing theory to construct a new model that needs 4 different parameters: one that describes the rate of addition of new cars on the insurance contract, a second one that models the rate of removal of insured vehicles, a third parameter that models the cancellation rate of the insurance policy, and finally a parameter that describes the rate of renewal. Statistical inference techniques allows us to estimate each parameter of the model, even in the case where there is censorship of data. We also propose to generalize this new queuing process by adding some explanatory variables into each parameter of the model. This allows us to determine which policyholder's profiles are more likely to add or remove vehicles from their insurance policy, to cancel their contract or to renew annually. The estimated parameters help us to analyze the insurance portfolio in detail because the queuing theory model allows us to compute various kinds of useful statistics for insurers, such as the expected number of cars insured or the customer lifetime value that calculates the discounted future profits of an insured. Using car insurance data, a numerical illustration based on a portfolio from a Canadian insurance company is included to support this discussion.

Key Words: Queuing theory, Poisson process, Lifetime customer value, Count distribution, Statistical inference

1 Introduction

In recent years, a value that has gained interest in actuarial science is the *lifetime customer value*, see Guillén et al. (2012), Guelman et al. (2014) or Faris et al.(2010) for a general overview. This is a term from the marketing field that allows targeting of long-term clients. In insurance, it is a value that an insurer can assign to each insured, calculated by discounting all the future profits of the insured. Its application to the field of insurance is particularly recent (see Verhoef and Donkers, 2001), possibly because of the complexity of the actuarial field. To calculate the lifetime customer value of each client, as well as other useful statistics, we propose to model the number of insured cars per household. We use queuing theory (see Gross et al. 2008 for an overview) to construct a new model that needs 4 different parameters: one that describes the rate of addition of new cars on the insurance contract, a second one that models the rate of removal of insured vehicles, a third parameter that models the cancellation rate of the insurance policy, and finally a parameter that describes the rate of renewal. Using regression methods, we also identify insured profiles that are more interesting for insurers. A numerical illustration taken from a portfolio of a car insurance data from a Canadian insurance company is included to support this discussion.

In the second part of the paper, we propose the first approach to model the number of insured vehicles. This includes a process to model the arrival of new vehicles and another process to model the removal of insured vehicles from an existing insurance contract. A third process to model the renewal process will also be added to the model. In Section 3, we generalize the model of Section 2 to include a process that models cancellations during the contract. In Section 4, we propose a method to estimate the parameters of the models, for complete and for censored data. Covariates representing the characteristics of each household will then be added into each parameter of the process. In Section 5, we apply the model and calculate several useful statistics, such as the expected number of insured vehicles or the lifetime customer value. Possibles generalizations of the model will be discussed in Section 6, while Section 7 concludes the paper.

1.1 Definition of terms

The term **household** is used to designate a single customer, or an insured. This household can include several members (or drivers) and several cars grouped under one annual **insurance con-**

tract, which can be **renewed** each year. The contract represents the document that binds the insurer with the insured household. In this paper, we focus on the number of cars that the contract covers and that are owned by the same household. By extension, **added cars** and **removed cars** from the insurance contracts are also analyzed. Finally, at any time during the insurance coverage, a household can decide to cancel its contract, meaning that all the insured cars are also canceled accordingly. We call this event a **breach of contract** or a **cancellation**.

1.2 Data Used and Notations

The use of queuing theory is based on the different waiting times before a change in the number of insured cars. We base our research on empirical analyses that come from a Canadian car insurance database. This database contains general insurance information on each of the 322,174 households for the period of 2003 to 2007. Note that because of the small length of the analyzed time period, the data are censored. For each household, we have information on each of its insured cars. We also have information about new or broken contracts, contract renewal, added or removed cars. Section 4.3 analyzes the database in detail, and describes the insurance data more precisely, particularly the characteristics of each of these policies.

A graphical analysis of waiting times involved in the modeling is presented. First, in Figure 1.1, the distribution of the life of an insurance policy (in years) is shown. By the life of an insurance policy, we mean the time between the effective date of a new contract and the date of a non-renewal, between the effective date of a new contract and the date of a cancellation. In the first graph of this Figure, to avoid possible bias, we only used policies that were issued and canceled during the 2003-2007 period. This sample allows us to better understand the data. From the figure, we can see that there is a shock at each contract renewal date. Aside from that shock, we can also observe a decreasing exponential trend in the data. The color code shows that most departures happen when there is only one insured car on the insurance policy. In Figure 1.1, the time before the addition of a car on an existing insurance contract is also shown. For this graph, all the data were used. The last graph of Figure 1.1 uses again all data, and shows the time before a removal of a car on a policy, which remains in force despite the removal of a car. Again, we can see an exponential trend, shocks at each contract renewal date, even if these shocks are less important than the ones observed in the first graphs. The major purpose of our project is thus to create a mathematical model that

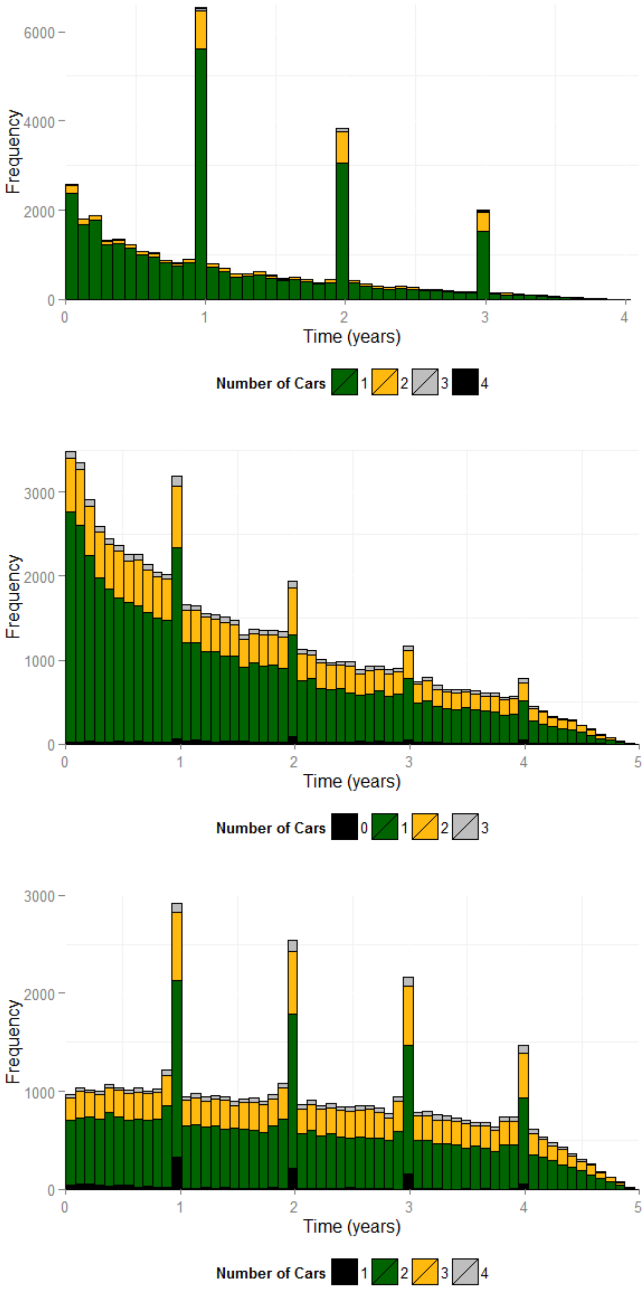

Figure 1.1: Life of an insurance policy, time prior to the addition of a car and time prior to removal of a car

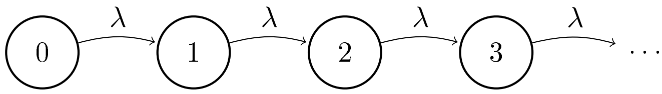

Figure 2.1: Network graph of a Poisson Process

will be able to approximate those observations.

Let $N(t)$ be a random variable representing the number of elements in a queuing system at time t. In our case, the number of elements is the number of insured cars owned by a specific household. The probability function of the number of insured cars will be expressed as $\Pr\{N(t) = i\} = p_i(t)$. The probability generating function (PGF) will be expressed as $P_{N(t)}(z,t) = \sum_{i=0}^{\infty} p_i(t) \times z^i$, and its partial derivative with respect to t by:

$$\frac{\partial P_{N(t)}(z,t)}{\partial t}(z,t) = \sum_{i=0}^{\infty} \frac{dp_i(t)}{dt} \times z^i. \quad (1.1)$$

Finally, the conditional probabilities will be represented and noted as $p_{(j,i)}(s,t) = \Pr\{N(t) = i | N(s) = j\}$.

2 Modeling the number of vehicles

In this section, we introduce how queuing theory, based on Newell(1982), can be used to model the number of insured cars. We first introduce the Poisson process to model the arrival of a new vehicle, and we add another process to model the removal of cars from the contract. Fewer details will be given in this part of the paper because interpreting the result requires only basic knowledge of queuing theory. Nonetheless, this introduction to queuing theory allows us to explain some tools that will be used in complex models, such as the one developed in Section 3.

2.1 Addition and removal of vehicles

In a pure birth process, also called the Poisson process and illustrated in Figure 2.1, there is only one component of arrival, defined by a parameter λ. The Chapman-Kolmogorov equation is defined in our context by:

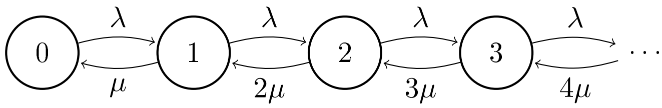

Figure 2.2: Network graph of a $M/M/\infty$ model

$$p_i(t) = \sum_{j=0}^{\infty} p_j(s) p_{(j,i)}(s,t),$$

for $s < t$. We interpret this equation by the fact that a probability can be defined by the sum of all the different paths for a short time period. These equations thus require us to find the conditional probabilities of the system, at some level i at time $t + \Delta t$, knowing that the process was at the level j at time t. This is expressed as:

$$p_{(j,i)}(t, t+\Delta t) = \begin{cases} \lambda \Delta t + o(\Delta t) & \text{when } j = i-1, \\ 1 - \lambda \Delta t + o(\Delta t) & \text{when } j = i, \\ o(\Delta t) & \text{when } j < i-1, \\ 0 & \text{otherwise}, \end{cases} \tag{2.1}$$

where $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$. For this first model, it can be shown that the probability function can be expressed as:

$$p_i(t) = \frac{e^{-\lambda t} (\lambda t)^i}{i!}, \tag{2.2}$$

which represents the classic Poisson distribution.

To obtain a model that is more realistic, we add a service component (which can also be called a death component) to the pure birth process. This is illustrated in Figure 2.2. The death component is supposed to be independent from the birth process. The resulting model is denoted as $M/M/\infty$. Like the Poisson process, the time between each arrival of a new element follows an exponential distribution. This feature is represented by the first M of the acronym, which means "Markov". For this model, each element leaves the system after a certain period of time that also follows an exponential distribution. This new process of the model defines the second M. Finally, the last symbol ∞ means that the departure process of each insured car can begin before the end of the departure process of another vehicle.

In other words, we are working with a dynamic modeling of a population that grows at a constant rate λ, and dies at a rate $i\mu$, with i representing the number of elements in the population at a specific time. In our context, the population is the number of cars in the household and therefore the new cars are added to the insurance contract at a rate λ, while vehicles leave the same insurance contract at an individual rate of μ. For this model, the conditional probabilities are:

$$p_{(j,i)}(t, t+\Delta t) = \begin{cases} \lambda \Delta t + o(\Delta t) & \text{when } j = i-1, \\ 1 - (\lambda + i\mu)\Delta t + o(\Delta t) & \text{when } j = i, \\ (i+1)\mu\Delta t + o(\Delta t) & \text{when } j = i+1, \\ o(\Delta t) & \text{otherwise.} \end{cases}$$

For this well-known model in queuing theory, it can be shown that the probability function is expressed as:

$$p_i(t) \quad = \quad \sum_{k=0}^{\min(i,a)} \frac{e^{-(1-e^{-\mu t})\lambda/\mu} \left[(1-e^{-\mu t})\lambda/\mu\right]^{i-k}}{(i-k)!} \times \binom{a}{k} e^{-\mu t k}(1-e^{-\mu t})^{a-k}. \tag{2.3}$$

The probability generating function is expressed as:

$$P_{N(t)}(z,t) = \left[(z-1)e^{-\mu t} + 1\right]^a e^{\frac{\lambda}{\mu}(1-e^{-\mu t})(z-1)} \tag{2.4}$$

where $N(0) = a$, which means that the initial number of insured cars (at time 0) is a.

2.2 Renewal of the vehicles

It is possible to add flexibility to the model $M/M/\infty$ by generalizing the component service process. Such a generalization means that the second M of the acronym should be replaced by a G (meaning general distribution). The proposed generalization allows us to incorporate a shock at the renewal time for each car. Indeed, as empirically shown in Figure 1.1, there is a higher probability that a

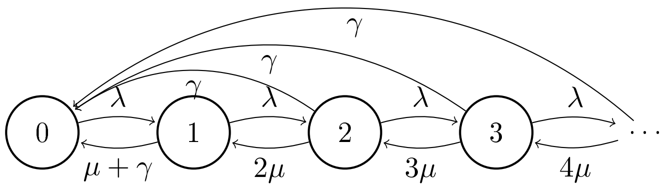

Figure 3.1: Network graph of a queuing model allowing cancellation of the insurance contract

car will be removed from the insurance contract at its annual renewal.

Based on Benes(1957a), the model is constructed by separating the process into several components modeling the number of cars. Indeed, we can suppose that the process $N(t)$ can be expressed as $N(t) = Z(t) - Y(t)$. In this case, the process $Z(t)$ will count the number of arrivals of new cars in the system until time t, and the process $Y(t)$ will count the number of departures of cars from the insurance contract until time t.

When $N(0) = 0$, it can be shown that the probability function can be expressed as:

$$\Pr(N(t) = i) = \frac{(\lambda t q_t)^i e^{-q_t \lambda t}}{i!}, \tag{2.5}$$

from which we recognize a Poisson distribution with parameters $\lambda t q_t$. The parameter q_t can be interpreted as a survival probability, and can be defined as:

$$q_t = \int_0^t \frac{S(x)}{t} dx,$$

where $S(\cdot)$ is the survival function of the service time, which corresponds in our case to the time since the car is insured. For a car insurance application, for example, we could include a shock that would happen at each renewal anniversary using the following function:

$$S(x) = e^{-\mu x} p^{\lfloor x \rfloor}, \tag{2.6}$$

where x is the age of the policy, and p represents the probability of renewal of the insurance contract.

3 Model with a period of inactivity

It would be interesting to generalize the model to include a possible cancellation of the contract, meaning that all vehicles insured in the same household leave the insurance company simultaneously. This model has to be constructed by adding a Poisson process that allows moves from levels $n \geq 1$ to 0 at a rate γ, recreating the effect of a cancellation. Figure 3.1 illustrates this model. The added process is supposed to be independent from the other ones. By studying this new model, we see, however, that it generates a conceptual problem. Indeed, the model allows the possibility that a new vehicle is added to an insurance contract even if the insurance contract has not had insured vehicles (state 0) for a while. This is counterintuitive: it seems more logical to believe that a policy without insured vehicles is simply cancelled. In this sense, the state 0 of this model should simply be an absorbent state.

However, further empirical analysis of the data points to a slightly more complex situation. Indeed, it sometimes happens that an insurance policy without an insured car is not cancelled. It means that sometimes a new vehicle is added to an existing insurance contract without insured cars. In our data, this happened 1,340 times, and the average time spent in the state 0 is almost half a year (0.4789).

This represents a form of inactivity of the contract, which allows reactivation of the contract when a new vehicle is added to the contract. One could think of situations such as storage of the last vehicle of the policy, the presence of another type of insurance (home or other) associated with an automobile policy and preventing the complete cancellation of the policy. There are also other more specific explanations for this situation. In this sense, empirically, we cannot consider the state without an insured vehicle as an absorbent level. However, the presence of an absorbent level that represents the final and decisive cancellation of a contract must be considered. This state will not be the one corresponding to a contract without an insured car, but a slightly different one.

We therefore propose a model such as the one shown in Figure 3.2. In this model, a new state is created and identified as 0^*. This new level is an absorbing state from which it is no longer possible to add vehicles[1]. In other words, the level 0^* is accessible only in cases where the contract is cancelled. It is thus noted that the passage to 0^* occurs at a rate γ for each level. Although we will

[1] If the insured cancels its contract and later wants to be covered again, insurers will often create a new insurance contract.

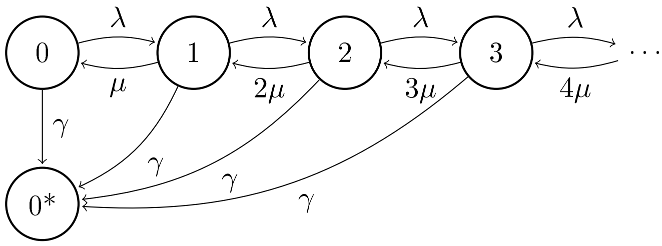

Figure 3.2: Network graph of a queuing model with absorbent state 0*

explore this situation in more detail in Section 4, we would like to highlight a conceptual difficulty with this new model. In the case of a policy with only one vehicle, empirically, the cancellation of the last vehicle is similar to the cancellation of the contract. In this sense, it is not clear whether the policy has been canceled (and enters the absorbing state 0^*) or if the insurance contract is simply in a period of inactivity (state 0). We will propose a solution to this problem later.

The construction of this new model is complex and requires new notations. Thus, in addition to $N(t)$, which represents the number of insured cars at time t, the random binary variable $M(t)$ is now introduced. This r.v. defines the life or death of the process at time t. In our context, a process is alive at time t, or $M(t) = 1$, when the insurance contract is in force, or in other words, when the contract has not been broken. A live process without insured vehicles is simply considered inactive, and at the state 0. Conversely, a dead process, or $M(t) = 0$ means that the household has cancelled its insurance contract and the absorbing state 0^* has been reached.

3.1 Modeling

For this model, Chapman-Kolmogorov equations should have been used to find the transition probabilities, from which difference-differential equations can be found. We instead use a simpler form of mathematical development. We first make the assumption that the time process related to the random variable $M(t)$ is exponentially distributed. In other words, we suppose that the time between the beginning of the contract and the end of the contract is $U \sim \text{Exponential}(\gamma)$. Consequently, we have $M(t) = 0$ when $t > U$ and:

$$\Pr(M(t) = 1) = e^{-\gamma t}. \tag{3.1}$$

We can then define the number of insured cars for both situations:

1. The case $M(t) = 1$: As long as the household is insured, the processes of adding and removing vehicles are active. As such, we can model the number of insured cars $N(t)$ with the $M/M/\infty$ queuing model seen previously.

2. The case $M(t) = 0$: When the cancellation occurs, the number of insured cars is 0. We then have $\Pr(N(t) = 0|M(t) = 0) = 1$.

Consequently, the joint PGF for $N(t)$ and $M(t)$ can easily be developed:

$$\begin{aligned}
P_{N(t),M(t)}(z, y, t) &= \sum_{m=0}^{1}\sum_{i=0}^{\infty} \Pr(M(t) = m, N(t) = i)y^m z^i \\
&= \sum_{m=0}^{1}\sum_{i=0}^{\infty} \Pr(N(t) = i|M(t) = m)\Pr(M(t) = m)y^m z^i \\
&= \Pr(M(t) = 0) + \sum_{i=0}^{\infty} \Pr(N(t) = i|M(t) = 1)\Pr(M(t) = 1)yz^i \\
&= 1 - e^{-\gamma t} + e^{-\gamma t} y P_{N(t)}(z, t) \\
&= 1 - e^{-\gamma t} + e^{-\gamma t} y \left((z-1)e^{-\mu t} + 1\right)^a e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}. \qquad (3.2)
\end{aligned}$$

The PGF of the $M/M/\infty$ model can be recognized from the right part of the equation, where equation (3.1) is added.

3.1.1 Adding Shock at the Renewal

At each policy anniversary, the insured has to renew his or her insurance contract. It is possible to modify the probability generating function by adding a Bernoulli trial that models this situation. To do this, equation (3.1) is modified to add a renewal probability p:

$$\Pr(M(t) = 1) = e^{-\gamma t} p^{\lfloor t \rfloor}. \qquad (3.3)$$

The variable t still represents the time in years, while the notation $\lfloor \cdot \rfloor$ is the floor function. At each policy anniversary for $p < 1$, this function causes an increase in the probability of cancellation.

It is easy to rewrite equation (3.2) with this new definition to have the following result:

$$P_{N(t),M(t)}(z,y,t) = 1 - e^{-\gamma t}p^{\lfloor t \rfloor} + e^{-\gamma t}p^{\lfloor t \rfloor}y\left((z-1)e^{-\mu t}+1\right)^a e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}. \qquad (3.4)$$

which now represents the final process that we proposed to model the number of insured cars.

4 Statistical Inference

It is possible to estimate parameters λ, γ, μ and p of the model by maximum likelihood. The estimation technique requires finding the joint probability density, and to adjust it based on observations of the database. Some assumptions must be made to be able to use the information available from the database. We will first summarize the steps in estimating the parameters of a simplified model where it is possible to distinguish each event. Then, by introducing data censorship and explanatory variables, the model will be adapted and generalized.

4.1 Complete Data

4.1.1 Notations

To estimate all parameters of the new model, a list of variables that are used in the likelihood function must be presented. All possible events observed during the life of the insurance policy will be noted as:

1. Type 1 event, which represents an addition of a car on the insurance contract. The random variable E represents the total number of events of this type (excluding the cars already insured at the beginning of the first observed contract);

2. Type 2 event, which represents a cancellation of the insurance policy at the policy anniversary date. The random variable Q is an indicator variable of this type of event. In other words, $Q = 1$ if there is no contract renewal;

3. Type 3 event, which represents a cancellation of the insurance policy at a different time than the anniversary of the policy. The random variable A is an indicator variable of this type of event;

4. Type 4 event, which represents a removal of a car from the insurance policy. The random variable S represents the total number of events of this type.

Thus, the total number of events, noted as K, is equal to the sum of all the previous elements, such as $K = E + A + Q + S$. We also propose notations for all the variables specifying time information about the insurance contract:

1. t_j, the time of occurrence (in years) of the j^{th} event affecting the number of insured cars;
2. $\widetilde{t_j}$ the time period (in years) between the $(j-1)^{\text{th}}$ and the j^{th} event, where $\widetilde{t_j} = t_j - t_{j-1}$, with $\widetilde{t_1} = t_1$.
3. T, the number of years the insurance contract was alive or inactive, where $T = \sum_{j=1}^{K} \widetilde{t_j}$ and hence $T = t_K$;
4. Finally, we define respectively by Ψ_1, Ψ_2, Ψ_3 and Ψ_4 the random variables that define the time before the occurrence of an event of type 1,2,3 or 4.

We then have:

1. N_j, a variable that counts the number of cars immediately before the j^{th} event;
2. V, the sum of the exposure time (in years) of all cars in a household so that $V = \int_0^T N(t)dt = \sum^K \widetilde{t_j} N_j$.

4.1.2 Likelihood function

The likelihood functions are developed based on the work of Benes(1957b) for the $M/M/\infty$ model. To estimate the parameters, we will build the conditional likelihood function based on the initial state. In this sense, we are not interested in the number of cars already insured in the portfolio (we work with this initial assumption), but only on processes that affect the number of cars in the future. Under the assumptions in the design of the system, these variables were defined as $\Psi_1 \sim \text{Exponential}(\lambda), \Psi_2 \sim \text{Exponential}(\mu), \Psi_3 \sim \text{Exponential}(\gamma)$ et $\Psi_4 \sim \text{Geometric}(p)$.

To define the joint distribution of all events for a single household, we have to analyze all possible cases. All processes are supposed to be independent, which facilitates decomposition of the model.

For example, knowing that the first event is the addition of a new vehicle on the insurance contract (type 1 event), observed at time $\widetilde{t}_1$, we obtain:

$$\Pr(\Psi_1 = \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1) = \lambda e^{-\lambda \widetilde{t}_1} e^{-N_1 \mu \widetilde{t}_1} e^{-\gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)}, \tag{4.1}$$

where $h(t)$ is a function that counts the number of times a specific household was in a renewal position before time t. This equation was developed using the properties of the first event of a joint distribution of exponentials. Then, by a similar development, it is possible to find the probability of events 2,3 or 4[2]:

$$\begin{aligned}
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 = \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1) &= e^{-\lambda \widetilde{t}_1} N_1 \mu e^{-N_1 \mu \widetilde{t}_1} e^{-\gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)}; && (4.2)\\
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 = \widetilde{t}_1, \Psi_4 > \widetilde{t}_1) &= e^{-\lambda \widetilde{t}_1} e^{-N_1 \mu \widetilde{t}_1} \gamma e^{-\gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)}; && (4.3)\\
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 = \widetilde{t}_1) &= e^{-\lambda \widetilde{t}_1} e^{-N_1 \mu \widetilde{t}_1} e^{-\gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)-1}(1-p). && (4.4)
\end{aligned}$$

Thereafter, knowing that exponential processes do not have memory, it is possible to calculate the product of all K events to obtain the joint distribution of all events of a single household:

$$\begin{aligned}
f(E, A, S, Q, T, V | \lambda, \gamma, \mu, p) &= \lambda^E e^{-\lambda \sum^K \widetilde{t}_j} \left(\prod^S N_s \right) \mu^S e^{-\mu \sum^K N_j \widetilde{t}_j} \gamma^A e^{-\gamma \sum^K \widetilde{t}_j} (1-p)^Q p^{\sum^K h(\widetilde{t}_j) - Q} \\
&\propto \lambda^E \gamma^A \mu^S (1-p)^Q e^{-(\lambda+\gamma)T} e^{-\mu V} p^{\lfloor T \rfloor - Q}, && (4.5)
\end{aligned}$$

where the constant terms have been removed from the second equation because they are not used in computing of the maximum likelihood method. For all Φ households, the loglikelihood function can be expressed by:

$$\ell = \sum_{i=1}^{\Phi} E_i \ln \lambda + A_i \ln \gamma + S_i \ln \mu + Q_i \ln(1-p) + (\lfloor T_i \rfloor - Q_i) \ln p - (\lambda + \gamma) T_i - \mu V_i, \tag{4.6}$$

[2] This technique is slightly biased by the fact that it does not consider the possibility of cancellation of the insurance contract at the renewal date. However, the impact is minimal because it corresponds to approximately 1/365.

where the subscript i identify each household. Finally, by maximizing the loglikelihood function, we can obtain the following estimators:

$$\hat{\lambda} = \frac{\sum^{\Phi} E_i}{\sum^{\Phi} T_i}, \quad \hat{\gamma} = \frac{\sum^{\Phi} A_i}{\sum^{\Phi} T_i}, \quad \hat{\mu} = \frac{\sum^{\Phi} S_i}{\sum^{\Phi} V_i}, \quad \hat{p} = \frac{\sum^{\Phi} \lfloor T_i \rfloor - Q_i}{\sum^{\Phi} \lfloor T_i \rfloor}.$$

4.2 Incomplete Data

Given that the time horizon of the sample is only 5 years, it is not always possible to observe the whole lifetime of an insurance policy. Consequently, the data that we are working with are necessarily censored and it is not possible to perform the estimation techniques based on complete data. To estimate the parameters correctly, important assumptions must be made. In this sense, we note a new variable, Ω, the observation date of the database. For example, in the case of our empirical analysis, this date is December 31, 2007.

Thus, we do not know what happened to a policy that had not been canceled before that date. Maybe the policy was canceled in 2008, or is still active in 2014. For those contracts, we must add to the likelihood function the time period between the last observed event K and Ω, i.e. a probability function that indicates that no events occur during that time interval. This corresponds to the joint probability of survival of 4 events. It is therefore fair to say, by independent random variables, that:

$$\begin{aligned} &\Pr(\Psi_1 > \Omega - t_K, \Psi_2 > \Omega - t_K, \Psi_3 > \Omega - t_K, \Psi_4 > \Omega - t_K) \\ &= \Pr(\Psi_1 > \Omega - t_K) \Pr(\Psi_2 > \Omega - t_K) \Pr(\Psi_3 > \Omega - t_K) \Pr(\Psi_4 > \Omega - t_K) \\ &= e^{-\lambda(\Omega - t_K)} e^{-N_K \mu(\Omega - t_K)} e^{-\gamma(\Omega - t_K)} p^{h(\Omega - t_K)}. \end{aligned} \tag{4.7}$$

4.2.1 Inactive vs Canceled

When there is censorship, we must distinguish between inactive insurance contracts (state 0) and canceled insurance contracts (state 0*). Indeed, in cases where only one vehicle was insured at time t_K, i.e. $N_K = 1$, an event of type 3 and an event of type 4 looks similar.

Let us first note this situation by a new indicator variable $AS(\Omega - t_K) = AS = 1$, meaning that this unknown state has a duration of $\Omega - t_K$. Because we are working with exponential random variables, we can sum equations (4.2) and (4.3). We obtain the probability that the household enters the unknown state at time t_K. This probability is expressed by $e^{-(\lambda+\mu+\gamma)\widetilde{t}_K} p^{h(\widetilde{t}_K)}(\mu + \gamma)$, which one may incorporate in the likelihood function.

By the properties of the exponential distribution, the probability that the household is in state 0^* (event of type 3) or in state 0 (event of type 4) are respectively:

$$\Pr(M(t_K) = 0|AS = 1) = \frac{\gamma}{\gamma + \mu} \tag{4.8}$$

$$\Pr(M(t_K) = 1, N(t_K) = 0|AS = 1) = \frac{\mu}{\gamma + \mu}. \tag{4.9}$$

It is possible to include the time $\Omega - t_K$ in the likelihood function to know which event between type 3 and 4 is the most probable.

When no possibility of renewal is possible between time t_k and Ω, 3 situations are possible at time t_K:

1. An event of type 3 occurred at t_K, which implies that the insurance policy is canceled;
2. An event of type 4 occurred at t_K and no other event has occurred up to Ω. This implies that the insurance policy is inactive.
3. An event of type 4 occurred in t_K and an event of type 3 occurs during the period (t_K, Ω), thus canceling the insurance policy.

Instead of calculating each of these probabilities, it is easier to calculate the complementary probability. This complementary probability corresponds to a single event: the occurrence of an event of type 4 at time t_K followed by an arrival during the period (t_K, Ω). This probability can be calculated as follows:

$$\Pr(M(t_K) = 1, N(t_K) = 0|AS = 1) \Pr(\Psi_1 < \min(\Psi_3, \Omega - t_K)) = \frac{\mu}{\gamma + \mu} \left(1 - e^{-(\gamma+\lambda)(\Omega - t_K)}\right) \frac{\lambda}{\lambda + \gamma}. \tag{4.10}$$

This corresponds to the product of three probabilities: the probability that an event of type 4 occurs at t_K, the probability that events of type 1 or type 3 occur before Ω and the probability that this last event is a vehicle entrance (type 1 event). It is important to note that this equation should be used only in cases where there is no possibility of renewal in the unknown period. Consequently, we need to generalize this equation to include each contract renewal in the calculation of the probability of remaining in the unknown state. To do this, it is necessary to separate the probability of arrival of a vehicle according to each year.

But first, several elements must be explained in detail. First, it should be noted that time t_K, corresponding to the moment when the status of the insured becomes unknown, does not necessarily match the time of the policy renewal. This means that the time period before the first renewal, and the last period before Ω (if different) are less than one year apart.

Consequently, if there is at least one renewal during the unknown period $\Omega - t_K$, a way to write the unknown time part of the year before the first renewal is $1 - \{t_K\}$, where $\{\cdot\}$ is the fractional part function. Similarly, the time part of the year before Ω can be written as $\{\Omega\}$, and the total number of renewals will be $\lfloor \Omega \rfloor - \lfloor t_K \rfloor = R$. We then have:

$$
\begin{aligned}
& \underbrace{\Pr(\Psi_1 < \min(\Psi_3, 1 - \{t_K\}))}_{\text{Before first renewal after } t_K} \\
+ \quad & \underbrace{\Pr(\min(\Psi_3, \Psi_1, \Psi_4) > 1 - \{t_K\}) \times \Pr(\Psi_1 < \min(\Psi_3, 2 - \{t_K\}) | \min(\Psi_3, \Psi_4, \Psi_1) > 1 - \{t_K\})}_{\text{Between the first and second renewal after } t_K} \\
+ \quad & \ldots
\end{aligned} \tag{4.11}
$$

This equation can be understood as the sum of many possibilities occurring between each renewal. The first element is simply the probability that an event E occurs before the first renewal after t_K. The second element of the sum is also the probability that an event E occurs, but between the first and the second renewal after t_K. In this case, the probability must also consider the fact that the insurance policy should have been renewed at the first renewal after t_K. All other situations, between potential renewal j and $j+1$, for $j=3,...$ must also be calculated, so that:

$$\begin{aligned}
& \Pr(\Psi_1 < \min(\Psi_3, \Psi_4, \Omega - t_K)|R \geq 1) \\
= \quad & \frac{\lambda}{\gamma+\lambda}\left(1 - e^{-(\gamma+\lambda)(1-\{t_K\})}\right) + \frac{\lambda}{\gamma+\lambda}e^{-(\gamma+\lambda)(1-\{t_K\})}p\left(1 - e^{-(\gamma+\lambda)}\right) + \dots \\
& + \frac{\lambda}{\gamma+\lambda}e^{-(\gamma+\lambda)(R-\{t_K\})}p^R\left(1 - e^{-(\gamma+\lambda)\{\Omega\}}\right) \\
= \quad & \frac{\lambda}{\lambda+\gamma}\left[\left(1 - e^{-(\gamma+\lambda)(1-\{t_K\})}\right) + e^{-(\gamma+\lambda)(1-\{t_K\})}p\left(1 - e^{-(\gamma+\lambda)}\right)\frac{1-\left(pe^{-(\gamma+\lambda)}\right)^{R-1}}{1-pe^{-(\gamma+\lambda)}}\right. \\
& \left. + e^{-(\gamma+\lambda)(R-\{t_K\})}p^R\left(1 - e^{-(\gamma+\lambda)\{\Omega\}}\right)\right].
\end{aligned}$$

where a simplification is done using the properties of geometric series. Cases when $R=0$ and $R>0$ can be combined to obtain:

$$\begin{aligned}
& \Pr(\Psi_1 < \min(\Psi_3, \Psi_4, \Omega - t_K)) \times \Pr(M(t_K) = 1, N(t_K) = 0|AS = 1) \\
= \quad & \frac{\mu}{\mu+\gamma}\frac{\lambda}{\gamma+\lambda}\left[\mathbb{1}(R=0)\left(1 - e^{-(\gamma+\lambda)(\Omega-t_K)}\right) + \mathbb{1}(R \geq 1)\left[\left(1 - e^{-(\gamma+\lambda)(1-\{t_K\})}\right)\right.\right. \\
& + e^{-(\gamma+\lambda)(1-\{t_K\})}p\left(1 - e^{-(\gamma+\lambda)}\right)\frac{1-\left(pe^{-(\gamma+\lambda)}\right)^{R-1}}{1-pe^{-(\gamma+\lambda)}} \\
& \left.\left. + e^{-(\gamma+\lambda)(R-\{t_K\})}p^R\left(1 - e^{-(\gamma+\lambda)\{\Omega\}}\right)\right]\right] \equiv G(t_K, \Omega). \qquad (4.12)
\end{aligned}$$

The case where $p=1$ gives the result shown by equation (4.10). Finally, for the sake of simplification, as already noticed, we calculated the complementary probability.

Following the correction of the model to include situations where it is difficult to distinguish between inactive and canceled insurance policy, the likelihood function (4.5) must be adjusted by adding the term $1 - G(t_K, \Omega)$:

$$L(\lambda, \gamma, \mu, p) \propto e^{-(\lambda+\gamma)T}e^{-\mu V}\lambda^E\mu^S\gamma^A(1-p)^Q p^{\lfloor T \rfloor - Q}\left[(\mu+\gamma)(1-G)\right]^{A^*}, \qquad (4.13)$$

where $A^* = 1$ if the insured is in the unknown state. Using the logarithm of this function, and summing for all households of a given portfolio, it becomes possible to estimate the four parameters of models, λ, γ, μ et p.

Car Arrival λ	Cancellation γ	Car Removal μ	Renewal p
0.0730 (0.0003)	0.0400 (0.0003)	0.0828 (0.0003)	0.9179 (0.0003)

Table 4.1: Parameters Estimations (std. err.)

4.3 Numerical Applications

The model was applied to the insurance dataset described in Section 1.2, and all 4 estimated parameters appear in Table 4.1. The value of $\widehat{\lambda}$ means that, for an active contract, at each $0.07303^{-1} = 13.69$ years in average, a new car will be added to the contract. The value of $\widehat{\mu}$ means that each car has an average life of $0.08277^{-1} = 12.08$ years into an insurance contract. Note that the arrival rate of cars is not enough to compensate for the departure rate of cars because $\widehat{\lambda} < \widehat{\mu}$. Note also that the annual renewal rate is about 92 % and that the probability of cancellation is approximately equal to 4%. For the insurer analyzed in this paper, arrival of new insureds would then be needed to ensure long-term profitability.

4.3.1 Covariates

Intuitively, we know that some household profiles are more likely to add or remove cars on their insurance contracts. Similarly, we may think that some profiles cancel more than others or that certain types of policyholders have a lower or higher rate of renewal. Thus, the addition of covariates into each parameters of the queuing process seems justified.

Covariates selected to define the vector $\boldsymbol{X}_i$ of each household are provided in Table 4.3.1. The effective dates of the insurance contract were used to show the stability of households insuring in July. Even if the characteristics of a household can change over the year, to simplify, only the characteristics observed in the first contract are considered. We consider that the effect is minimal, because the time horizon of the database is quite short, and most households do not change their risk characteristics over the year. However, future research might improve the modeling.

A link function $g(\boldsymbol{X}_i\boldsymbol{\beta})$ is then associated with each parameter, where $\boldsymbol{\beta}$ is the vector of parameters to be estimated. In our model, the parameters satisfy $\lambda, \gamma, \mu \in \mathbb{R}^+$, consequently a logarithmic link function is chosen because this link function allows parameters to be always positive. Moreover, because the parameter that models the renewal probability must satisfy $p \in [0, 1]$, we use the logit

Variable	Description
X1	equals 1 if the household comes from the general market (as opposed to group insurance)
X2	equals 1 if the household has at least one rented car
X3	equals 1 if the insureds are not married
X4	equals 1 if the household is with the insurance company for less than 9 years
X5	equals 1 if the effective date of the insurance contact is between January and July
X6	equals 1 if the effective date of the insurance contact is in July
X7	equals 1 if the effective date of the insurance contact is on the first day of a month

Table 4.2: Binary variables summarizing the information available about each household

Parameters	β_λ	β_γ	β_μ	β_p
β_0	-2.494 (0.015)	-3.709 (0.039)	-2.726 (0.014)	2.769 (0.019)
β_1	-0.264 (0.009)	-0.070 (0.019)	0.122 (0.009)	-0.152 (0.010)
β_2	-0.142 (0.011)	·	-0.048 (0.011)	0.092 (0.013)
β_3	-0.454 (0.009)	0.214 (0.018)	0.410 (0.008)	-0.246 (0.009)
β_4	0.201 (0.014)	0.677 (0.038)	0.130 (0.013)	-0.399 (0.017)
β_5	·	·	·	0.137 (0.009)
β_6	-0.098 (0.012)	-0.181 (0.029)	-0.034 (0.012)	0.400 (0.016)
β_7	-0.129 (0.009)	-0.647 (0.021)	-0.129 (0.009)	0.089 (0.010)

Table 4.3: Estimated parameters for the process with covariates

link, i.e. $p_i = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_p)}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_p)}$. The estimated values of the vector parameters $\boldsymbol{\beta}$ are shown in Table 4.3.

The objective of an insurer should be to maximize the number of cars insured at time t. In this case, we are looking for a high value of parameters λ and p, combined with low values for the other parameters μ and γ. The results suggests that some covariates have a great impact, such as marital status, or X6 identifying insured whose effective date is July 1^{st}. In addition, policyholders renewing their insurance contract on the first of each month also offer increased stability. Finally, as expected, the covariate X4, which identifies insured with the insurance company for less than 9 years, shows higher loyalty to their insurer. To test whether the explanatory variables are statistically significant, a Wald test was performed according to a confidence interval of 95% for each parameter. Consequently, β_5 was not included in $\boldsymbol{\beta}_\lambda$, $\boldsymbol{\beta}_\gamma$ and $\boldsymbol{\beta}_\mu$ and β_2 was not included in $\boldsymbol{\beta}_\gamma$.

5 Analysis

In this section, applications are presented using the estimated parameters found by regression in the previous section and shown in Table 4.3. Thus, for example, even if we are working with $\widehat{\lambda}$, $\widehat{\gamma}$,

Parameter	X1	X2	X3	X4	X5	X6	X7
Household A	1	0	1	1	0	0	0
Household B	1	1	1	0	0	0	0
Household C	0	0	1	0	0	0	1
Household D	1	1	0	0	0	0	0
Household E	0	1	0	0	0	1	1

Table 5.1: Covariates of each profile

Parameter	λ	γ	μ	p
Household A	0.0492	0.0557	0.1270	0.87795
Household B	0.0349	0.0283	0.1063	0.92154
Household C	0.0460	0.0159	0.0868	0.93163
Household D	0.0550	0.0228	0.0705	0.93756
Household E	0.0570	0.0107	0.0530	0.96608

Table 5.2: Parameter values for each profile

$\widehat{\mu}$ or $\widehat{p}$, for simplicity, we will note those parameters by λ, γ, μ and p.

We selected 5 profiles to represent the impact of market segmentation. Indeed, given that there are 96 possible profiles, only some typical insured will be used to show the results of our analyses. The first selected profile is the best type of household E in terms of expected insured cars, while the worst profile corresponds to household A. For illustration, we also used 3 average types (B, C, D). Table 5.1 expresses each profile in terms of their covariates.

We can see that the only difference between household types B and D lies in marital status. As we can see by the results shown in Table 4.3, this covariate has a significant impact on each parameter of the model. In Table 5.2, the value of each parameter λ, γ, μ and p is shown for each profile.

5.1 Expected number of insured cars

Using the probability generating function of the complete process presented in Section 3.1.1, interesting properties can be found. To simplify notations and computations, we will use the variable $H(t)$ as the number of insured cars from an active or inactive insured, i.e. $H(t) = M(t) \times N(t)$. Knowing that $\Pr(X = n) = \frac{d^n P(z=0)}{dz^n}$, we can compute the expected number of insured cars $H(t)$ by using:

$$\mathbb{E}[H(t)] = \frac{\partial P_{N(t),M(t)}(z=1,y=1,t)}{\partial z} = \sum_{i=0}^{\infty} i p_i^{(1)} 1^{i-1} 1.$$

We compute the expected value for our process by using equation (3.4). We first took the derivative in respect to z:

$$\begin{aligned}
\frac{\partial P_{N(t),M(t)}(z,t)}{\partial z} &= \frac{\partial \left[1 - e^{-\gamma t} p^{\lfloor t \rfloor} \left(1 - \left((z-1)e^{-\mu t}+1\right)^a y e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}\right)\right]}{\partial z} \\
&= e^{-\gamma t} p^{\lfloor t \rfloor} \left((z-1)e^{-\mu t}+1\right)^{a-1} y e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}} \\
&\quad \times \left[a e^{-\mu t} + \left((z-1)e^{-\mu t}+1\right)\left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right]. \qquad (5.1)
\end{aligned}$$

After, setting $z = 1$ and $y = 1$, the expected value can be found:

$$E[H(t)] = \frac{\partial P_{H(t)}(z=1,t)}{\partial z} = e^{-\gamma t} p^{\lfloor t \rfloor} \left(a e^{-\mu t} + \left(1 - e^{-\mu t}\right)\frac{\lambda}{\mu}\right). \qquad (5.2)$$

With similar computations, other moments of the distribution can be found, such as the variance or higher moments. Those computations can also be applied to all the other probability generating functions shown in the paper.

For each profile, the expected number of insured cars has been computed. The resulting function is shown in Figure 5.1. For illustration, we set the value $c = 0$, meaning that the insureds are at the beginning of their contract, and that the next renewal will be in one year. In the Figure, we can clearly see the shock of each contract renewal, modeled by the parameter p. Numerical values can be found in Table 5.3, where the expected number of insured cars after 5 years, i.e. just after the renewal, is shown. We see that the initial number of insured cars linearly increases the expected value.

The results point to a large difference between each profile, where for example, household E seems to be much more advantageous for insurers than household A. Consequently, we believe that an insurer could adapt its marketing efforts to target certain types of insureds, in this case households of type E instead of household of type A. Indeed, because they stay longer in the company, insurers should be able to offer discounts to some insured profiles that generate lower administrative costs because they stay with the company longer.

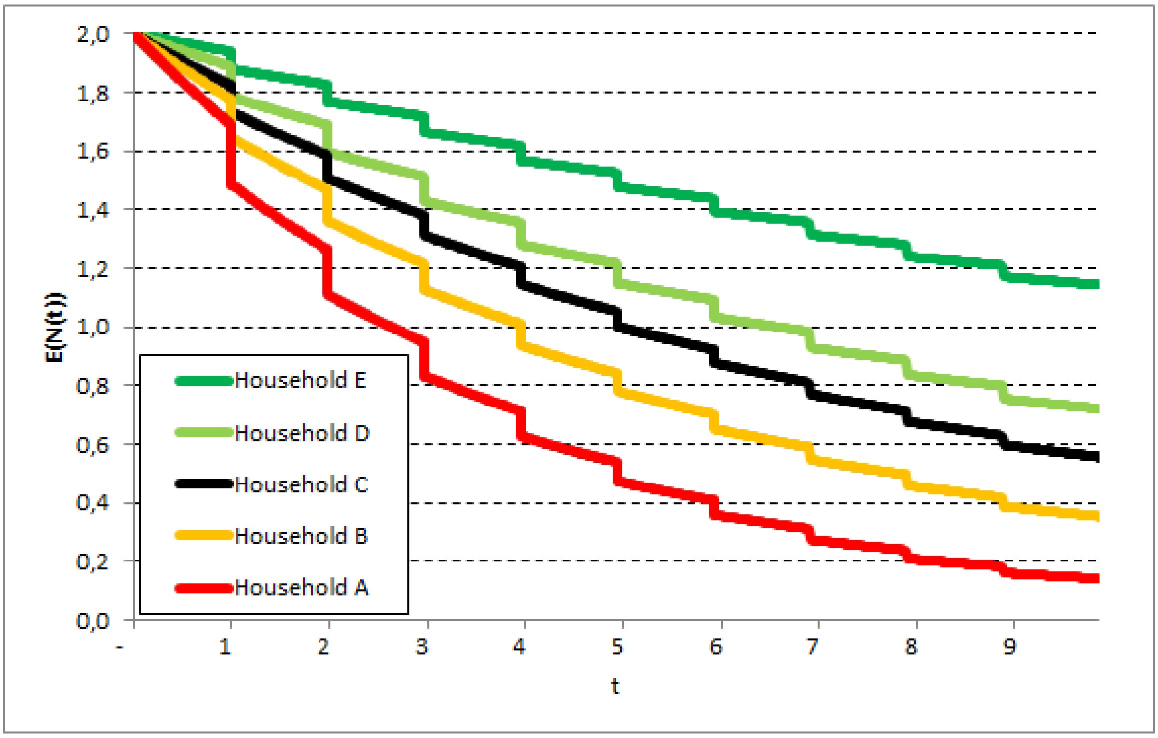

Figure 5.1: Expected number of insured cars at time t

Initial number of insured cars (a)	1	2	3	4
Household A	0.281	0.490	0.700	0.909
Household B	0.417	0.756	1.096	1.435
Household C	0.541	0.961	1.381	1.801
Household D	0.604	1.058	1.513	1.967
Household E	0.812	1.424	2.036	2.648

Table 5.3: Expected number of insured cars after 5 years

5.2 Discounting risk exposures

The discounted monetary value of a household can be found by calculating the present value of cash flows in the future. In marketing terms, this result is called customer lifetime value. This value is analyzed in several articles, for example Donkers et al.(2007), which compares different calculation methods, Guillén et al. (2012) and Guelman et al. (2014). In our case, we will assume that the insurance company makes a \$1 profit for each one-year car exposure. This assumption can easily be modified to be more realistic. We also use an instant discount rate of δ. Therefore, to calculate the customer lifetime value, it is necessary to calculate the present value of future exposure until time T using a continuous δ discount rate. The customer lifetime value will be denoted by ω_T, and is computed by integrating the variable $H(t)$ such that:

$$\begin{aligned} \omega_T &= \int_0^T H(t)e^{-\delta t}dt \\ &= \lim_{n\to\infty} \frac{T}{n}\sum_{i=1}^{n} H\left(\frac{iT}{n}\right)e^{-\frac{\delta iT}{n}}, \end{aligned} \tag{5.3}$$

by using Riemann sums. The expected value can be easily calculated such as:

$$\begin{aligned} E(\omega_T) &= \lim_{n\to\infty} \frac{T}{n}\sum_{i=1}^{n} E\left(H\left(\frac{iT}{n}\right)\right)e^{-\frac{\delta iT}{n}} \\ &= \int_0^T E(H(t))e^{-\delta t}dt. \end{aligned} \tag{5.4}$$

In Table 5.4, we show the expected value of each household if we suppose that the household has two insured cars at time $t = 0$. In the long term, the differences between household profiles become apparent. Indeed, household E is on average 2.5 times more profitable after 20 years than household A.

Classically, in the marketing literature, it is uncommon to calculate the customer lifetime value over a very large time horizon. However, we found it useful to compute this result because it allows us to obtain simple equations. Using equation (5.3), we can find the total profit for each household when T goes toward infinity. Define this value as $\omega \equiv \lim_{T\to\infty} \omega_T$. Its expectation can be calculated

Number of years (T)	1	5	10	20
Household A	1.833	5.468	6.720	7.122
Household B	1.869	6.359	8.581	9.731
Household C	1.904	7.003	10.104	12.237
Household D	1.917	7.286	10.823	13.506
Household E	1.946	8.218	13.553	19.433

Table 5.4: Customer lifetime value for households having two insured cars at time $t = 0$, with a 2% discount rate

by:

$$\begin{aligned} E(\omega) &= \int_0^\infty E(H(t))e^{-\delta t}dt \\ &= \int_0^\infty e^{-(\gamma+\delta)t}p^{\lfloor t\rfloor}\left(ae^{-\mu t} + \left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right)dt \\ &= \sum_{i=0}^\infty \int_i^{i+1} e^{-(\gamma+\delta)t}p^i\left(ae^{-\mu t} + \left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right)dt \\ &= \sum_{i=0}^\infty a\frac{\left(e^{-i(\gamma+\delta+\mu)} - e^{-(i+1)(\gamma+\delta+\mu)}\right)}{\gamma+\delta+\mu}p^i + \frac{\left(e^{-i(\gamma+\delta)} - e^{-(i+1)(\gamma+\delta)}\right)\lambda}{(\gamma+\delta)\mu}p^i \\ &\quad - \frac{\left(e^{-i(\gamma+\delta+\mu)} - e^{-(i+1)(\gamma+\delta+\mu)}\right)\lambda}{(\gamma+\delta+\mu)\mu}p^i \\ &= \frac{\left(1-e^{-(\gamma+\delta+\mu)}\right)\left(a-\frac{\lambda}{\mu}\right)}{(\gamma+\delta+\mu)\left(1-e^{-(\gamma+\delta+\mu)}p\right)} + \frac{(1-e^{-(\gamma+\delta)})\lambda}{(\gamma+\delta)\mu(1-e^{-(\gamma+\delta)}p)}, \end{aligned} \tag{5.5}$$

where a is the number of insured cars at time $t = 0$. The numerical results, shown in Table 5.5 [3] , show a great disparity between the best and worst households. Consequently, following our assumptions, we can see once again that an insurance company should target policyholders having the covariates of household E. This analysis is based on a \$1 per household profit. This can be generalized in a profit that can be a proportion of the premium. Moreover, using this model and an economic model incorporating the elasticity of the price, the premium can be set as a function of future profits.

[3] We made an approximation with covariate x_4 that we kept fixed, but that should change over time. This is an approximation done only for household A.

Number of insured cars at time $t = 0$, (a)	1	2	3	4
Household A	3.964	7.161	10.358	13.555
Household B	5.588	9.992	14.395	18.799
Household C	7.661	13.011	18.361	23.711
Household D	8.782	14.586	20.390	26.194
Household E	16.130	24.735	33.340	41.945

Table 5.5: Customer lifetime value, with a 2% discount rate

6 Discussion

As mentioned previously, numerical results of our empirical study suggest that an insurer should pay more attention to some households than to others. We wanted to introduce to the actuarial literature queuing processes that are realistic and directly applicable to model risk exposures in insurance, but it is important to understand that these conclusions were based on some assumptions of our model.

In the previous section of the paper, we mentioned that the assumption of the profit component of the premium should be generalized to be more realistic. Other changes are also possible. The explanatory variables should be more dynamic to reflect the change in household characteristics over time. For example, the covariate modeling the time insured within the company (x_4) should logically change over the year. For example, we could add claims experience to the model. This could improve the fit of the model, mainly for parameter p, which represents the renewal rate. Indeed, we intuitively believe that the insured's behavior will not be the same if the insured files a claim during a year. This change in behavior would probably be related to a premium increase. In this sense, in such a generalization of the model a system of experience rating could be introduced. The hunger for bonus phenomenon (see Lemaire 1976, or Boucher et al. 2009) should also be added to obtain a more useful model and a more realistic approach.

Other interesting generalizations of the model are also possible. As mentioned by an anonymous referee, dependence between the four processes used the model should be added in a future modeling. For example, dependence between the cancellation rate and the renewal date seems logical. Indeed, an insured who is aware of the prices on the insurance market should have a higher probability of changing insurers during an insurance term and upon renewal. Similarly, it seems logical to believe that an insured who adds a car on his insurance contract will be more likely to cancel one of the other vehicles. More generally, an insured who adds many new cars on his or her contract should also remove many vehicles. Other kinds of dependence between processes are possible. However, even if the dependence structure seems to be obvious, the addition of covariates in the model should diminish the impact of such dependence.

Even if the inclusion of dependences between the processes can be an interesting generalization of the model, we think that simpler generalizations of the model should be more useful. For example, the use of other waiting time distributions (gamma instead of the exponential distribution, for example) to model the processes should be envisioned. Other uses of the model proposed in the paper, to be used with other lines of business (such as home insurance or group insurance) should also be interesting avenues. Mixing several lines of business can also be considered.

7 Conclusion

We wanted to model the number of insured cars for each household. Starting with a simple Poisson process, it is possible to generalize many types of queuing models. The model proposed in this paper can be seen as a generalization of the $M/M/\infty$ process. We add a new death level to the system, which allows the possibility of cancellation or non-renewal of contracts. Justified by empirical data, we also propose a distinction between a canceled policy and an inactive insurance contract.

The proposed new model needs 4 parameters: one parameter that models the rate of addition of new cars on the insurance contract, a second parameter that models the rate of removal of insured vehicles, a third parameters that models the cancellation rate of the insurance policy, and finally a parameter that describes the rate of renewal. Statistical inference techniques allowed us to estimate each of these parameters, often by using the properties of the exponential distribution, and by conditioning on all possible events. Finally, because we worked with censored data, we have developed a way to estimate the parameters in the case where it is not possible to distinguish between the inactive contracts and the canceled insurance contracts.

We also proposed to generalize this new queuing system by adding some explanatory variables into each of the 4 parameters of the model. It was then possible to segment the portfolio and to determine which policyholders' profiles are more likely to add or remove vehicles from their insurance policy, cancel their contract or renew annually. The estimated parameters obtained help us to analyze the insurance portfolio in detail because we developed various kinds of useful statistics for insurers, such as the expected number of insured cars or customer lifetime value that calculates the future profits associated with an insured.

We believe that this model offers a good approximation of the empirical data and proposes an interesting first step in the modeling of exposure time in insurance. It may be appropriate to continue to improve the proposed model in the future. Several types of generalizations have been proposed.

Finally, note that this analysis considers only the evolution of households already in the portfolio. Thus, the model does not include the arrival of new policies in the portfolio. Consequently, the number of insured households in the insurance portfolio decreases over time. However, the main purpose of the paper was to analyze the change in the number of insured cars in a specific household, as well as the insured's behavior and the impact of household characteristics. If an insurer wants to analyze the evolution of its entire portfolio, it should take into account the arrival of new households. The insurer could thus calculate the monetary value of the portfolio (also called customer equity), analogous to the calculation leading to customer life value. We are currently working on that modeling.

Acknowledgements

Jean-Philippe Boucher gratefully acknowledge the financial support from the Natural Sciences and Engineering Research Council of Canada.

References

[1] V. E. Benes (1957a) Fluctuations of telephone traffic. *Bell System Technical Journal*, 36, 965–973.

[2] V. E. Benes (1957b) A sufficient set of statistics for a simple telephone exchange model. *Bell System Technical Journal*, 36, 939–964.

[3] J.-P. Boucher, M. Denuit and M. Guillén (2009) Number of Accidents or Number of Claims? An Approach with Zero-inflated Poisson Models for Panel Data. *Journal of Risk and Insurance*, 76(4), 821 - 846.

[4] B. Donkers, P. Verhoef, and M. Jong (2007) Modeling clv: A test of competing models in the insurance industry. *Quantitative Marketing and Economics*, 5(2), 163–190.

[5] P. W. Farris, N. T. Bendle, P. E. Pfeifer, and D. J. Reibstein (2010). *Marketing Metrics: The Definitive Guide to Measuring Marketing Performance (2nd Edition).* Pearson Prentice Hall, 2nd edition.

[6] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris (2008) *Fundamentals of Queueing Theory (Wiley Series in Probability and Statistics).* Wiley-Interscience, 4th edition.

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

[8] Guelman, L., Guillén, M. and Pérez-Marín, A.M. (2014) *Uplift Random Forests* Cybernetics & Systems Special issue on Intelligent Systems in Business and Economics, in press.

[9] J. Lemaire (1976) La Soif du Bonus, *ASTIN Bulletin*, 9, 181-190.

[10] G. Newell (1982) *Applications of queueing theory.* Monographs on statistics and applied probability. Chapman and Hall.

[11] P. C. Verhoef and B. Donkers (2001) Predicting customer potential value an application in the insurance industry. *Decision Support Systems*, 32(2), 189–199.