# Modeling the Number of Insured Households in an Insurance Portfolio Using Queuing Theory

Jean-Philippe Boucher and Guillaume Couture-Piché

December 8, 2015

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

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

**Abstract**

In this paper, we use queuing theory to model the number of insured households in an insurance portfolio. The model is based on an idea from Boucher and Couture-Piché (2015), who use a queuing theory model to estimate the number of insured cars on an insurance contract. Similarly, the proposed model includes households already insured, but the modeling approach is modified to include new households that could be added to the portfolio. For each household, we also use the queuing theory model to estimate the number of insured cars. We analyze an insurance portfolio from a Canadian insurance company to support this discussion. Statistical inference techniques serve to estimate each parameter of the model, even in cases where some explanatory variables are included in each of these parameters. We show that the proposed model offers a reasonable approximation of what is observed, but we also highlight the situations where the model should be improved. By assuming that the insurance company makes a $1 profit for each one-year car exposure, the proposed approach allows us to determine a global value of the insurance portfolio of an insurer based on the customer equity concept.

**Key Words:** Queuing theory, Customer equity, Poisson process, Count distribution, Statistical inference

## 1 Introduction

In this paper, we generalize the queuing theory model developed by Boucher and Couture-Piché (2015) (subsequently called the BCP model), which estimates the number of insured cars on an insurance contract. An important novelty of this paper is the inclusion of a new household's arrival process. Because this new model includes both already insured households and new households that could be added to the portfolio, the proposed approach allows us to determine not only the customer lifetime value (see Guillén et al. (2012), Guelman et al. (2014) or Guelman et al. (2015)), but also the global value of the insurance portfolio of an insurer according to the customer equity concept (Rust et al., 2004).

We propose to model the number of insured households by using (1) the arrival process of new households, (2) a contract cancellation process and (3) a contract renewal process. For the number of insured cars per household, the model uses (4) a process that models the addition of a car to the contract and (5) another process that models the removal of an insured's car from the contract. Because the mathematical models employed are similar, the models of the number of cars per household and the number of households in the insurance portfolio will easily be nested to form a complete model that emulates the total number of insured cars for an insurance company. Our model is based on observations from an insurance portfolio from a Canadian insurance company. Statistical inference techniques are proposed to estimate each parameter of the model, even in cases where some explanatory variables are included in these parameters. We show that the proposed model offers a reasonable approximation of what is observed, but we also highlight the situations where the model should be improved.

We are interested in probability generating functions (PGF), which are an effective way to obtain all the required information in a simple equation. Thus, in Section 2, the PGF of the number of insured households is built. In Section 3, the PGF of the total number of insured cars is developed. Subsequently, in Section 4, the parameters required for models are estimated under the specific assumptions of our new model. Then, using these estimated parameters, interesting and useful statistics are calculated in Section 5, which include the number of insured cars at future time $t$, as well as an estimate of the present value of future profits. Section 6 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 contract**, which can be **renewed** each year. The contract represents the document that binds the insurer with the insured household. Finally, the expression **portfolio** is used to designate all the contracts of a single insurer.

In this paper, we focus on the number of insured households in an insurer's portfolio, and on the number of cars that the contract covers and that are owned by the same household. By extension, **cars that are added to or removed** 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. We call this event a **breach of contract** or a **cancellation**.

### 1.2 Data Used and Empirical Analysis

Our model is built from observations in our database. We base our research on empirical analyses that come from a Canadian car insurance database, which is the same as the one used in the illustration in the BCP model. Unlike the BCP model, the new approach proposed in this paper considers the arrival of new households in the portfolio.

This database contains general insurance information on each of the 322,174 households from 2003 to 2007. For each household, we have information on each of their insured cars. We also have information about new or broken contracts, contract renewal, added or removed cars. The number of new contracts by date of entry into the insurance portfolio, and the number of insured cars per contract at the entry date can be seen in Figure 1.1.

We note that depending on the calendar date, the arrival rate varies greatly and displays an apparent seasonality. The arrival of new contracts occurs more in summer than in winter. It should be noted that the number of cars insured when taking out a new contract is never equal to 0, which will require some adjustments to the equations used in the BCP model. Indeed, the BCP model allows the possibility that an insurance contract could be active without any insured cars.

## 2 Number of Insured Households

In this section, we describe how queuing theory, based on Newell (1982), can be used to model the number of insured households. By explaining how Boucher and Couture-Piché (2015) used the queuing model to predict the number of insured cars, we introduce the Poisson process from which we add the death component to model departures from the system. Fewer details will be

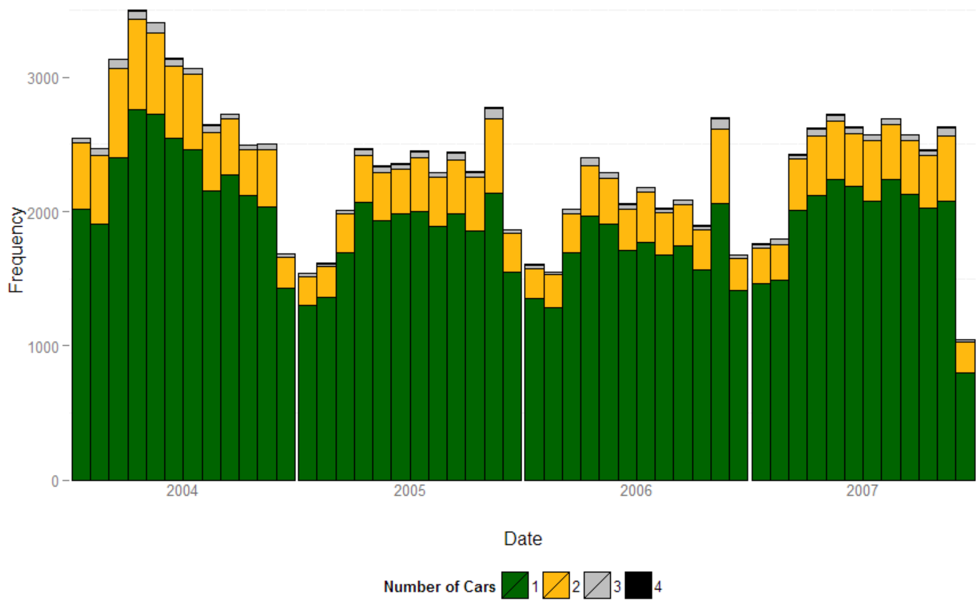


Figure 1.1: Number of new households by arrival date

given in this part of the paper because the results require only a basic knowledge of queuing theory. Moreover, some results have already been introduced with the BCP model. This introduction to queuing theory allows us to explain some tools that will be used in complex models.

## 2.1 General Characteristics of the Model

Let $N(t)$ be a random variable representing the number of elements in a queuing system at time $t$. In our case, it will first represent the number of insured households in the insurer's portfolio, while it will model the number of insured cars for a specific household in Section 3. In a pure birth process, also called the Poisson process, there is only one component of arrival, defined by a parameter $\tau$. It can be shown that the probability function of the number of elements in the system at time $t$ can be expressed as:

$$\Pr(N(t) = i) = \frac{e^{-\tau t}(\tau t)^i}{i!}, \tag{2.1}$$

where $\tau$ can be considered the rate of arrival of new elements in the system. This distribution represents the classic Poisson distribution.

To obtain a more realistic model, we add a constant service component (which can also be called a death component) to the pure birth process. The resulting model is denoted $M/M/\infty$. The first $M$ of the acronym, which means "Markov", denotes the exponential distribution of the time between the arrival of each new element. Because each element leaves the system after a certain period of time that also follows an exponential distribution, a second $M$ is used in the acronym. Finally, the last symbol identifies the number of parallel servers. Because the departure process of each element can begin before the end of the departure process of another element, there is no waiting time and the symbol $\infty$ is chosen. See Gross et al. (2008) for more details about queuing theory.

It is possible to add flexibility to the $M/M/\infty$ model 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 departure rate that can change over time. 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 households in the portfolio until time $t$, and the process $Y(t)$ will count the number of departures of households from the insurer's portfolio until time $t$.

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

$$\Pr(N(t) = i) = \frac{(\tau t q_t)^i e^{-q_t \tau t}}{i!}, \tag{2.2}$$

from which we recognize a Poisson distribution with parameters $\tau t q_t$. The parameter $\tau$ again represents the rate of arrival of new households in the portfolio, while the parameter $q_t$ has to 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 any survival function of the service time. An interpretation of Gross et al. (2008, p.258), allows us to explain the parameter $q_t$ as the probability of an arbitrary household that enters the portfolio between time $(0, t)$ still being in the insurer's portfolio at time $t$.

To continue our construction of the model, we have to consider two different kinds of households.

Next, we separate the modeling of the number of household into two components:

1. New insureds that enter the portfolio after the start date of the analysis, insureds which will be called **new households**;

2. Insureds already in the portfolio when the portfolio is analyzed, which will be called **old households**.

## 2.2 New Households

Let us first note $\mathcal{K}(t)$, the random variable representing the number of new households that are still in the insurer's portfolio at time $t$. We suppose that this random variable follows a $M/G/\infty$ process. Because, by definition, with $\mathcal{K}(0) = 0$, meaning that there are no new households in the portfolio at time 0, we can begin the model by using equation (2.2). We will, however, generalize the departure rate $\gamma$ of this equation by adding a shock at the renewal time. Indeed, because we expect that more insureds will not renew their insurance contracts at the renewal date, compared with other days, we have to modify the model to have a higher cancellation probability at the renewal date.

To include a departure shock that happens at each renewal anniversary, we use the following function:

$$S(x) = e^{-\gamma x} p^{\lfloor x \rfloor}, \tag{2.3}$$

where $p$ represents the probability of renewal of the insurance contract and $\lfloor x \rfloor$ is a floor function that allows the parameter $p$ to affect the survival function only at each renewal of the contract. Thus, we have:

$$\begin{aligned} q_t &= \int_0^t \frac{e^{-\gamma x} p^{\lfloor x \rfloor}}{t} dx \\ &= \left(1 - e^{-\gamma}\right) \frac{1 - \left(pe^{-\gamma}\right)^{\lfloor t \rfloor}}{\gamma t \left(1 - pe^{-\gamma}\right)} + \frac{p^{\lfloor t \rfloor} \left(e^{-\gamma \lfloor t \rfloor} - e^{-\gamma t}\right)}{\gamma t}. \end{aligned} \tag{2.4}$$

It can be shown that the PGF of this $\mathcal{K}(t)$ is:

$$\begin{aligned} P_{\mathcal{K}(t)}(z,t) &= \sum_{i=0}^{\infty} \frac{z^i e^{-\tau q_t t} (\tau q_t t)^i}{i!} \\ &= e^{(z-1)\tau q_t t}. \end{aligned} \tag{2.5}$$

which is the PGF of a Poisson distribution (see Gross et al. 2008).

Note that although the data shown in Figure 1.1 seem to exhibit seasonality, we assume in our model that the arrivals of new households in the portfolio will be modeled by a Poisson process with fixed parameter $\tau$. A model that is closer to reality would have had a rate of new arrivals that is a function of time and would exhibit variations. Regrettably, supposing this kind of flexible arrival rate would mean that some of the Markov properties of the model would be lost, and would have generated more complex equation systems (for example, a $G/G/\infty$ process). As mentioned in Eick et al. (1993), a fixed arrival rate is an obvious approximation strategy, commonly applied in practice. This allows us to suppose that the number of insured households at time $t$ follows a $M/G/\infty$ process, from which the resulting generating function is already known.

## 2.3 Old Households

We now define by $\mathcal{R}(t)$ the number of households that were initially present in the portfolio at the time the portfolio was analyzed, and that are still in the portfolio at time $t$. As opposed to $\mathcal{K}(t)$, the random variable $\mathcal{R}(t)$ is not equal to 0 at time 0. We will note $R_0 = \mathcal{R}(0)$ as the initial number of insured households.

By construction, the random variable $\mathcal{R}(t)$ is affected by the departure process only because arrivals of new households are modeled by $\mathcal{K}(t)$, already defined. Consequently, the process is only a function of the rate of renewal of the contract (modeled by the parameter $p$), and a function of the cancellation rate (modeled by the parameter $\gamma$).

To model the number of old households at time $t$, we will need to define a new random variable. We thus note by $\mathcal{M}_i(t)$, a random variable that indicates whether household $i$ is still insured at time $t$. In which case $\mathcal{M}_i(t) = 1$. Consequently, it is possibl e to use the equation $\mathcal{R}(t) = \sum_{i=1}^{R_0} \mathcal{M}_i(t)$ to model the number of old households still insured at time $t$. The random variable $\mathcal{M}_i(t)$ is a Bernoulli random variable and has the following PGF:

$$P_{\mathcal{M}_i(t)}(z) = e^{-\gamma t} p^{\lfloor t+c_i \rfloor}(z-1) + 1. \tag{2.6}$$

where $p$ represents the probability of renewal of the insurance contract and $\lfloor t+c \rfloor$ is the same floor function as equation (2.3), which allows the parameter $p$ to affect the survival function at each renewal of the contract. Because $t$ is a unique calendar time that affects all households of the portfolio, and because the renewal date of each household $i$ is not the same throughout the portfolio, we included a new constant $c_i$ in the model. This constant $c_i$ allows the correct use of the parameter $p$. Indeed, the function $\lfloor t + c_i \rfloor$ will change units at each anniversary of the insurance contract, i.e. at the time of the renewal, and will cause a departure shock in the survival function. For example, a household $i$ that is analyzed at time $t$ when its contract will be renewed in three months, will have a value of $c_i = 9/12$.

### 2.4 All Households

We define by $\mathcal{W}(t)$ the total number of insured households at time $t$, i.e. the number of old households and the number of new households still insured at time $t$, or $\mathcal{W}(t) = \mathcal{K}(t) + \mathcal{R}(t)$. Because these two random variables are independent, we have the relation $P_{\mathcal{W}(t)}(z,t) = P_{\mathcal{K}(t)}(z,t) \times P_{\mathcal{R}(t)}(z,t)$, from which we get the generating function of $\mathcal{W}(t)$:

$$P_{\mathcal{W}(t)}(z,t) = \prod_{i=1}^{R_0} \left[ e^{-\gamma t} p^{\lfloor t+c_i \rfloor}(z-1) + 1 \right] \times e^{\tau t q_t (z-1)}. \tag{2.7}$$

This PGF of $\mathcal{W}(t)$ will be used in the next section of the paper to model the number of insured cars.

## 3 Total Number of Insured Cars

Each insurance contract covers several vehicles: new vehicles can be added to the insurance contract and vehicles can be removed. Consequently, from an household point of view, we are still working with a queuing theory process, this time for the number of insured cars. By combining the number of households, and the number of cars per household, we are able to estimate the total number of insured cars in the portfolio. To construct this model, we use a combination of the model for the number of households, and a classic queuing theory model for the number of insured cars per household. By defining $\mathcal{W}(t)$, the number of insured households, the total number of insured cars $\mathcal{J}(t)$ can be calculated using the equation:

$$\mathcal{J}(t) = \sum_{j=1}^{\mathcal{W}(t)} \mathcal{N}_j(t), \tag{3.1}$$

where $\mathcal{N}_j(t)$ is the number of insured cars at time $t$ for household $j$. We suppose that $\mathcal{W}(t) = 0$ means $\mathcal{J}(t) = 0$.

Before proceeding to the development of the model, it is important to explain again what the time, represented by the parameter $t$, represents. As opposed to a household point of view that has its own time variable $t$, as in the BCP model, in the proposed model the time $t$ should be seen as the calendar time, and is therefore shared by all households. We assume that $t = 0$ is the time when the insurance portfolio is analyzed. Thus, the process $\mathcal{N}_j(t)$ depends on the calendar time and not the age of the contract. The number of cars $\mathcal{N}_j(t)$ of household $j$ therefore changes over time $t$, even if this household $j$ is not even insured by the insurer that we are analyzing.

Once again, the total number of insured cars that come from already insured (old) households ($\mathcal{J}_A(t)$) and the total number of insured cars that come from new households ($\mathcal{J}_N(t)$) will be analyzed separately. This will allow us to treat the problem specifically for each case.

## 3.1 New Households

To model $\mathcal{J}_N(t)$, the number of insured cars owned by new households, we will use the following equation:

$$\mathcal{J}_N(t) = \sum_{j=1}^{\mathcal{K}(t)} \mathcal{N}_j(t),$$

where the variable $\mathcal{N}_j(t)$ follows a $M/M/\infty$ process.

As defined in Section 2.3, $\mathcal{K}(t)$ is the random variable that represents the number of new households that are still in the insurer's portfolio at time $t$. The random variable $\mathcal{K}(t)$ is a $M/G/\infty$ process that assumes no initial insured households. The PGF of $\mathcal{K}(t)$ is given by equation (2.5). For $\mathcal{N}_j(t)$, the PGF of a $M/M/\infty$ process is a well-known model in queuing theory and expressed as:

$$P_{N(t)}(z,t) = \left[(z-1)e^{-\mu t} + 1\right]^a e^{\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)(z-1)}, \tag{3.2}$$

with $a$, the number of insured cars for a specific household, at the time the insurance portfolio was analyzed. $\mathcal{N}_j(t)$ models the number of insured cars for a specific household $j$, which means that new cars are added to the insurance contract at a rate of $\lambda$, while vehicles leave the same insurance contract at an individual rate of $\mu$.

By the properties of the PGF, we can develop the equation:

$$\begin{aligned} P_{\mathcal{J}_N(t)}(z,t) &= P_{\mathcal{K}(t)}(P_{\mathcal{N}(t)}(z,t),t) \\ &= \exp\left[\tau t q_t \left[\left((z-1)e^{-\mu t}+1\right)^a e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}} - 1\right]\right]. \end{aligned} \tag{3.3}$$

We see that the PGF requires us to give a value to $a$, the number of insured cars at time $t = 0$, the time the insurance portfolio was analyzed. For old households, the initial number of insured cars, $\mathcal{N}(0) = a$, was always considered to be known, and this information was available in the database analyzed. In the case of a new household, we must make an assumption about the distribution of the number of insured cars at arrival.

One possible solution is to set a fixed value for $a$, but it is clear that the number of cars per household at time $t = 0$ is not the same for everyone. Another way to determine the number of insured cars per household is to take the stationary distribution of $\mathcal{N}(t)$, i.e. the distribution of $\mathcal{N}(t)$ when $t \to \infty$. The resulting PGF in this case is simply:

$$\lim_{t\to\infty} P_{\mathcal{N}(t)}(z,t) = e^{\frac{\lambda}{\mu}(z-1)}, \tag{3.4}$$

which corresponds to the PGF of a Poisson distribution with parameter $\frac{\lambda}{\mu}$. Thus, one might assume that the number of insured cars of a new household follows such a distribution. This choice for the distribution of the number of insured cars is advantageous in that it simplifies the equations of the model because it is a distribution that is not a function of the calendar time $t$.

A problem caused by the use of the Poisson distribution is that it becomes possible for a household to apply for a new insurance contract without having a single car to insure. To address this problem, we change the support of the random variable that models the number of cars per household by supposing that it cannot be less than one. This suggests a transformation where we add 1 to a random variable that follows a Poisson distribution. Consequently, the number of cars per household $\mathcal{N}(t)$ is equal to:

$$\mathcal{N}(t) = \mathcal{N}^*(t) + 1, \tag{3.5}$$

where $\mathcal{N}^*(t) \sim \text{Poisson}(\lambda/\mu)$. For this transformation, however, it is important to note that the interpretation of the parameter $\mu$ changes slightly. Consequently, we obtain the following PGF for the new random variable $\mathcal{N}(t)$:

$$\begin{aligned} P_{\mathcal{N}(t)}(z,t) &= P_{\mathcal{N}^*(t)}(z,t) \times z \\ &= z e^{\frac{\lambda}{\mu}(z-1)}, \end{aligned} \tag{3.6}$$

which means that we have:

$$P_{\mathcal{J}_N(t)}(z,t) = e^{\tau t q_t (e^{\frac{\lambda}{\mu}(z-1)} z - 1)}, \tag{3.7}$$

### 3.2 Old Households

For all $R_0$ households that are already insured, we can base our modeling on the random variable $\mathcal{M}_i(t)$, explained in Section 2.3, which indicates whether household $i$ is still insured at time $t$. With $\mathcal{H}_j(t)$, which represents the number of insured cars insured for household $j$, we have the following relationship:

$$\mathcal{H}_j(t) = \mathcal{M}_j(t) \times \mathcal{N}_j(t)$$

where the variable $\mathcal{N}_j(t)$ will be the same as the one used for the new households, and expressed by the relation (3.5). The $j^{\text{th}}$ household has its number of insured cars modeled by the random variable $\mathcal{H}_j(t)$, which contains $a_j$ the initial number of insured cars, and $c_j$ the time constant previously introduced. Thus, we find the following equation, by composition of generating functions:

$$
\begin{aligned}
P_{\mathcal{H}(t)}(z,t) &= P_{\mathcal{M}(t)}\left(P_{\mathcal{N}(t,a)}(z,t),t\right) \\
&= P_{\mathcal{M}(t)}\left(P_{\mathcal{N}^*(t,a)}(z,t)\times z,t\right) \\
&= 1-e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\left(1-\left((z-1)e^{-\mu t}+1\right)^{a-1}ze^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}\right)
\end{aligned}
$$

where we can observe a change due to the relation $\mathcal{N}(0)=a$, which implies $\mathcal{N}^*(0)=a-1$.

To model $\mathcal{J}_A(t)$, the number of insured cars by old households, we then have:

$$\mathcal{J}_A(t)=\sum_{j=1}^{R_0}\mathcal{H}_j(t),$$

which corresponds to the sum of all insured cars owned by the $R_0$ old households that are still insured at time $t$. By using all the PGF, we have:

$$P_{\mathcal{J}_A(t)}(z,t) = \prod_{i=1}^{R_0}\left[1-e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\left(1-\left((z-1)e^{-\mu t}+1\right)^{a_i-1}ze^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}\right)\right] \tag{3.8}$$

which expresses the PGF of $\mathcal{J}_A(t)$.

## 3.3 All Households

To model the total number of insured vehicles coming from new and old households, we have to find the PGF of the following random variable:

$$\mathcal{J}(t)=\mathcal{J}_A(t)+\mathcal{J}_N(t). \tag{3.9}$$

The PGF for the all households of the portfolio can be expressed as:

$$
\begin{aligned}
P_{\mathcal{J}(t)}(z,t) &= P_{\mathcal{J}_A(t)}(z,t)\times P_{\mathcal{J}_N(t)}(z,t) \\
&= \prod_{i=1}^{R_0}\left[1-e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\left(1-\left((z-1)e^{-\mu t}+1\right)^{a_i-1}ze^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}\right)\right] \\
&\quad \times e^{\tau t q_t(e^{\frac{\lambda}{\mu}(z-1)}z-1)},
\end{aligned} \tag{3.10}
$$

where information from household $i$ is used through $c_i$ and $a_i$.

# 4 Inference

The PGF expressed in equation (3.10) uses several parameters that we can estimate using real insurance data. Indeed, we have to estimate $\tau$, the rate of arrival of new households in the portfolio, $\gamma$ the rate of departure of households from the portfolio, $p$ the probability of annual contract renewal, $\lambda$ the rate of arrival of new cars on a contract and finally $\mu$ the rate of departure of cars from the contract.

## 4.1 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 parameter $\lambda$ is used to model the addition rate;
   - 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);
   - $\Psi_1$ is a random variables that defines the time before the occurrence of an event of this type.
2. Type 2 event, which represents a removal of a car from the insurance policy.
   - The parameter $\mu$ is used to model the removal rate;
   - The random variable $S$ represents the total number of events of this type.
   - $\Psi_2$ is a random variables that defines the time before the occurrence of an event of this type.
3. Type 3 event, which represents a cancellation of the insurance policy at a different time than the anniversary of the policy.

- The parameter $\gamma$ is used to model the cancellation rate;
- The random variable $A$ represents the total number of events of this type.
- $\Psi_3$ is a random variables that defines the time before the occurrence of an event of this type.

4. Type 4 event, which represents a cancellation of the insurance policy at the policy anniversary date. In other words, an event of this type is recorded if there is no contract renewal.
   - The parameter $p$ is used to model the probability of renewal;
   - The random variable $Q$ represents the total number of events of this type.
   - $\Psi_4$ is a random variables that defines the time before the occurrence of an event of this type.
5. Type 5 event, which represents the arrival of a new household.
   - The parameter $\tau$ is used to model the arrival rate;
   - The random variable $B$ represents the total number of events of this type.
   - $\Psi_5$ is a random variables that defines the time before the occurrence of an event of this type.

Thus, the total number of events, noted as $K$, is equal to the sum of all the previous elements, such as $K = E + S + A + Q + B$. We also propose to note:

1. $U$, the total observed time period;
2. $\xi$, the total number of observed households in the database;
3. $T_i$, the number of years household $i$ was insured;
4. $V_i^*$, the sum of the covered insurance time of all vehicles from household $i$. Note that because we use the transformation $\mathcal{N}^*(t) = \mathcal{N}(t) - 1$, we have $V_i = V_i^* - T_i$;

Finally, because we construct the model by event, we also propose the following event-relation notations:

1. $t_j$ is the time of occurrence (in years) of the $j^{\text{th}}$ event affecting the number of insured cars;

2. $\widetilde{t}_j$ is the time period (in years) between the $(j-1)^{\text{th}}$ and the $j^{\text{th}}$ event, where $\widetilde{t}_j = t_j - t_{j-1}$, with $\widetilde{t}_1 = t_1$;

3. $W_j$ is the number of insured households in the portfolio immediately before the $j^{\text{th}}$ event;

4. $J_j$ is the total number of insured cars in the portfolio immediately before the $j^{\text{th}}$ event;

5. $h(\widetilde{t}_j)$ is the total number of contracts renewed in the portfolio during the period $\widetilde{t}_j$.

## 4.2 Likelihood function

We first note that the time before the arrival of new households has an exponential distribution. According to the Markov assumption, these arrivals are independent from additions of cars, from removals of cars and breaches of contract. We can use the properties of exponential distributions to calculate the joint probability that the first event is the addition of a car.

$$\Pr(\Psi_1 = \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1, \Psi_5 > \widetilde{t}_1) = \left(W_1 \lambda e^{-\lambda W_1 \widetilde{t}_1}\right) e^{-J_1 \mu \widetilde{t}_1} e^{-W_1 \gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)} e^{-\tau \widetilde{t}_1}.$$

Similarly, we have the following probabilities:

$$\begin{aligned}
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 = \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1, \Psi_5 > \widetilde{t}_1) &= e^{-\lambda W_1 \widetilde{t}_1} \left(J_1 \mu e^{-J_1 \mu \widetilde{t}_1}\right) e^{-W_1 \gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)} e^{-\tau \widetilde{t}_1} \\
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 = \widetilde{t}_1, \Psi_4 > \widetilde{t}_1, \Psi_5 > \widetilde{t}_1) &= e^{-\lambda W_1 \widetilde{t}_1} e^{-J_1 \mu \widetilde{t}_1} \left(W_1 \gamma e^{-W_1 \gamma \widetilde{t}_1}\right) p^{h(\widetilde{t}_1)} e^{-\tau \widetilde{t}_1} \\
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 = \widetilde{t}_1, \Psi_5 > \widetilde{t}_1) &= e^{-\lambda W_1 \widetilde{t}_1} e^{-J_1 \mu \widetilde{t}_1} e^{-W_1 \gamma \widetilde{t}_1} \left(p^{h(\widetilde{t}_1)-1}(1-p)\right) e^{-\tau \widetilde{t}_1} \\
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1, \Psi_5 = \widetilde{t}_1) &= e^{-\lambda W_1 \widetilde{t}_1} e^{-J_1 \mu \widetilde{t}_1} e^{-W_1 \gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)} \left(\tau e^{-\tau \widetilde{t}_1}\right)
\end{aligned}$$

Because all processes involve exponential distribution, the likelihood function can be computed as the product of all events that occur during the observed time period $U$, such as:

$$\begin{aligned}
&\mathcal{L}(\tau,\gamma,p,\lambda,\mu)\\
&= e^{-\sum_{j=1}^{K}\widetilde{t}_j(\tau+W_j(\lambda+\gamma)+J_j\mu)}p^{\sum_{j=1}^{K}h(\widetilde{t}_j)}\prod_{j=1}^{E}(\lambda W_j)\prod_{j=1}^{S}(\mu J_j)\prod_{j=1}^{A}(\gamma W_j)\prod_{j=1}^{Q}(1-p)\prod_{j=1}^{B}\tau\\
&\propto e^{-(U\tau+\sum_{i=1}^{\xi}T_i(\lambda+\gamma)+\sum_{i=1}^{\xi}V_i\mu)}p^{\sum_{i=1}^{\xi}\lfloor T_i\rfloor-Q}\lambda^{E}\mu^{S}\gamma^{A}(1-p)^{Q}\tau^{B}, \qquad (4.1)
\end{aligned}$$

where constants have been removed because they are not useful for maximum likelihood calculation.

We must adapt the likelihood function to introduce the distribution of the number of insured cars that a new household has when it enters the insurance portfolio. As defined by equation (3.4), this distribution is a function of $\lambda$ and $\mu$, meaning that the distribution of the number of cars is:

$$\begin{aligned}
\Pr(\mathcal{N}(0)=a) &= \Pr(\mathcal{N}^*(0)=a-1)\\
&= \frac{e^{-\frac{\lambda}{\mu}}}{(a-1)!}\left(\frac{\lambda}{\mu}\right)^{a-1}, \qquad (4.2)
\end{aligned}$$

when we add this to the likelihood function (4.1), we obtain the following likelihood function:

$$\begin{aligned}
\mathcal{L}(\tau,\gamma,p,\lambda,\mu) \propto\ & e^{-(U\tau+\sum_{i=1}^{\xi}T_i(\lambda+\gamma)+\sum_{i=1}^{\xi}V_i\mu+\xi\frac{\lambda}{\mu})}p^{\sum_{i=1}^{\xi}\lfloor T_i\rfloor-Q}\\
&\times\lambda^{E+\sum_{i=1}^{\xi}(a_i-1)}\mu^{S-\sum_{i=1}^{\xi}(a_i-1)}\gamma^{A}(1-p)^{Q}\tau^{B}. \qquad (4.3)
\end{aligned}$$

where $a_i$ is the initial number of insured cars for household $i$.

Note that the addition of the distribution of the initial number of insured cars $a_i$ has the effect of creating a more complex equation for estimating the parameters $\mu$ and $\lambda$. Thus, these two parameters cannot be estimated by an explicit formula. However, maximizing equation (4.3), we find the following estimators for the other parameters:

$$\widehat{\gamma} = \frac{A}{\sum_{i=1}^{\xi}T_i}, \qquad (4.4)$$

$$\widehat{p} = \frac{\sum_{i=1}^{\xi}\lfloor T_i\rfloor - Q}{\sum_{i=1}^{\xi}\lfloor T_i\rfloor}, \qquad (4.5)$$

$$\widehat{\tau} = \frac{B}{U}. \qquad (4.6)$$

| New cars $\widehat{\lambda}$ | Removal of cars $\widehat{\mu}$ | Cancellation of contract $\widehat{\gamma}$ | Contract renewal $\widehat{p}$ | New Household $\widehat{\tau}$ |
|---|---|---|---|---|
| 0.0624 (0.0002) | 0.2315 (0.0007) | 0.0918 (0.0003) | 0.9188 (0.0003) | 28,014 (83.68) |

Table 4.1: Parameters Estimators (std. err.)

## 4.3 Estimated Parameters

We estimated parameters of the model with a Canadian car insurance database, already introduced in Section 1.2. Estimators of the parameters can be seen in Table 4.1.

To check the quality of the model, estimated parameters can be compared to data, via empirical estimators. Because the data used is highly censored, we used the Kaplan-Meier estimators and Kaplan-Meier survival plots over up to 5 years. We did not include confidence intervals in our graphes, as it was quite narrow around each curve. However, as we analyzed a dataset observed over a small number of years (5 years), and because new households are constantly added to the portfolio, it is important to understand that most of the observed events happens in the first 1-2 years of each graphes.

For each estimated parameter, we observe the following results:

1. Cancellation ($\gamma$) and renewal ($p$) rates:

   The first interesting comparison between the data and the estimated parameters of the model would be the survival analysis of households in the insurance portfolio, which includes both parameters $\gamma$ and $p$. We see that the annual renewal rate $p$ is about 92% and that the annual probability of cancellation $\gamma$ is approximately 9%. Figure 4.1 compares the fit of the model with the Kaplan-Meier estimators. The fit of the model is very good. We can observe that the model seems to approximate quite well what is observed with real insurance data.

2. The rate of arrival $\lambda$ of new cars on a contract:

   In Table 4.1, we see that the value of $\widehat{\lambda}$ is equal to 0.0624, which means that, for an active contract, at each $0.0624^{-1} = 16.02$ years on average, a new car will be added to the a contract. Figure 4.2 compares the value of $\widehat{\lambda}$ with Kaplan-Meier estimators, from which we can see that the fit is close to the empirical estimators, but seems to constantly, but slightly, overestimate

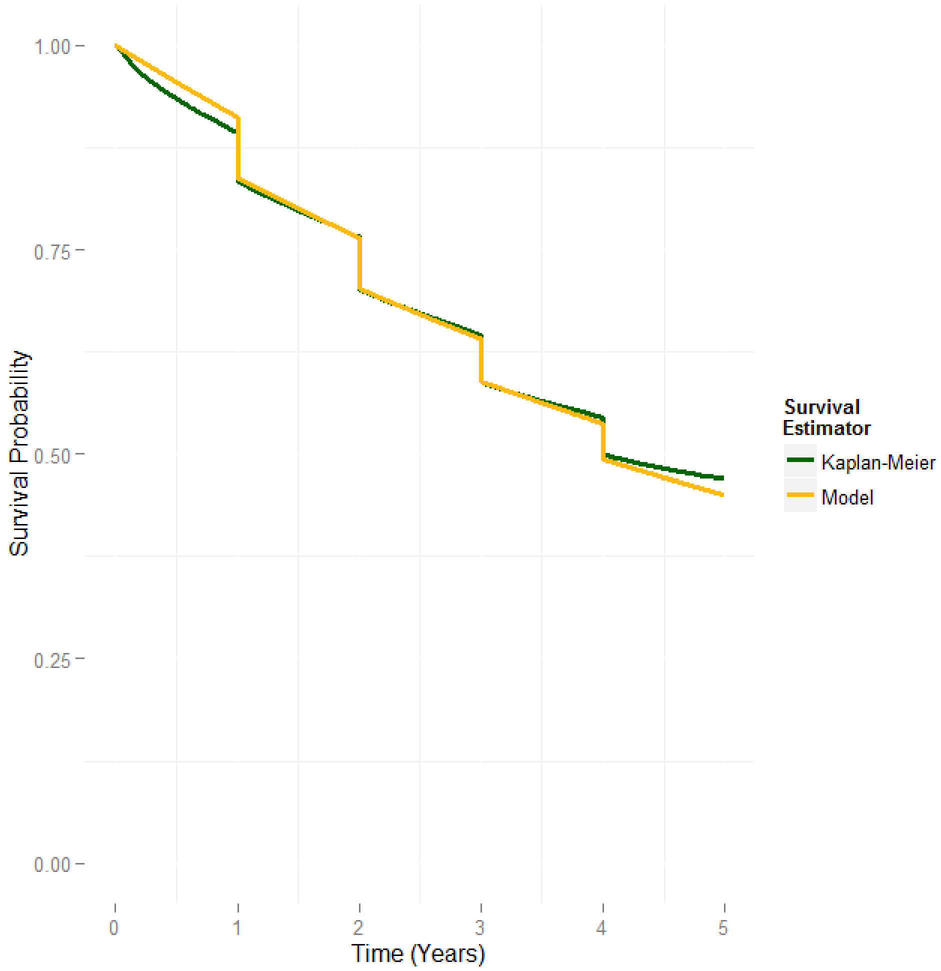


Figure 4.1: Analysis of the cancellation rate of households

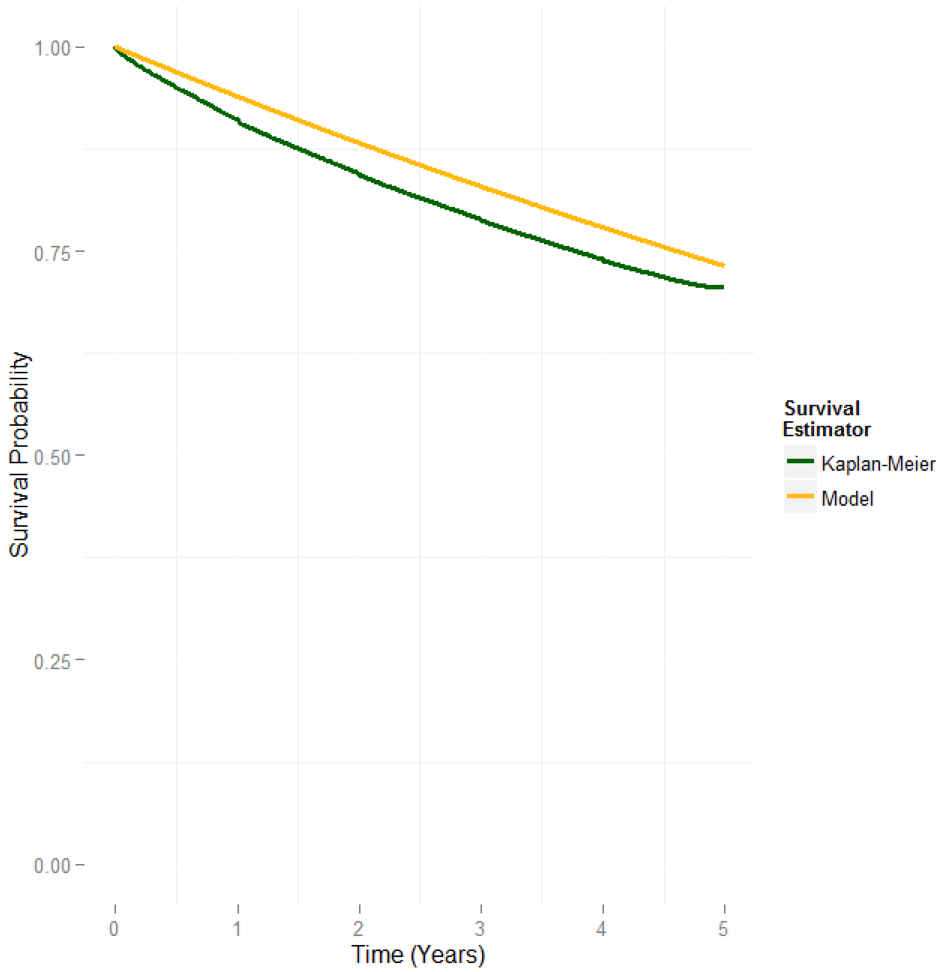


Figure 4.2: Analysis of the arrival rate of new vehicles

it. Note that the number of cars at time 0, as seen in equation (4.2), also depends on $\lambda$ (and $\mu$). We will compare the model with the initial number of insured cars later.

3. The departure rate $\mu$ of cars from the contract:

   The value of $\widehat{\mu}$ means that each car has an average life of $0.2315^{-1} = 4.32$ years into an insurance contract. It is also interesting to note that the arrival rate of cars is not enough to compensate for the departure rate of cars since $\widehat{\lambda} < \widehat{\mu}$.

   We must be careful with the analysis of this parameter, which is used to model the removal rate of vehicles. Indeed, this rate depends on the number of insured cars on the contract, as expressed in the $M/M/\infty$ process, meaning that the departure rate is equal to $(J-1)\mu$, with

$J$ representing the number of insured cars of a specific insurance contract, at a specific time. Figure shows survival functions for contracts with respectively 2, 3 and 4 insured vehicles.

The fit of the model is interesting for contracts with 3 or 4 vehicles, but not for the contracts with 2 vehicles. As explained, we should focus on the first years of the graph to draw conclusions. The model consistently underestimates the departure rate for each situation. This might explains why the arrival rate $\lambda$ was slightly higher than the Kaplan-Meier estimates; as the arrival process tries to correct the effect of underestimation of the departure rate $\mu$.

The model proposed in this paper uses only one parameter ($\mu$) to explain the behavior of all households, without directly considering the number of insured cars on the contract. For example, the departure rate of an insurance contract with 3 insured cars is simply the double of the departure rate of an insurance contract with 2 insured cars, and a contract with 4 vehicles has a departure rate 3 times higher than a contract with 2 cars. This lack of flexibility might explain the bad quality of the fit for this process. Regrettably, it is not easy to correct this property of the model.

Both $\hat{\mu}$ and $\hat{\lambda}$ are also used to estimate the number of cars at time 0, as seen in equation (4.2). Figure 4.6 compares the initial number of insured cars with the model for a new insurance policy, where we can also observe the impact of the transformation $\mathcal{N}^*$. The transformation ensures that the distribution of the initial number of insured cars is closer to what is observed in reality. Interestingly, if we had not changed the model to constrain the initial number of insured cars of a new insurance policy, i.e. using $\mathcal{N}$ instead of $\mathcal{N}^*$, the number of new households without insured cars would have been approximately equal to 40%, which is obviously unrealistic.

Compared with the estimated values of parameters shown in Boucher and Couture-Piché (2015), the main difference in estimators is $\widehat{\mu}$. Note that the transformation $\mathcal{N}^*$ is also responsible for the small change in the parameter estimator $\widehat{\lambda}$, while $\widehat{p}$ remains very similar to the estimators of in Boucher and Couture-Piché (2015). The BCP model assumes an inactive state, while the transition from one insured cars to no insured car is now considered as a cancellation of the contract. Thus, it is likely that the parameter $\widehat{\gamma}$ is higher than the BCP model.

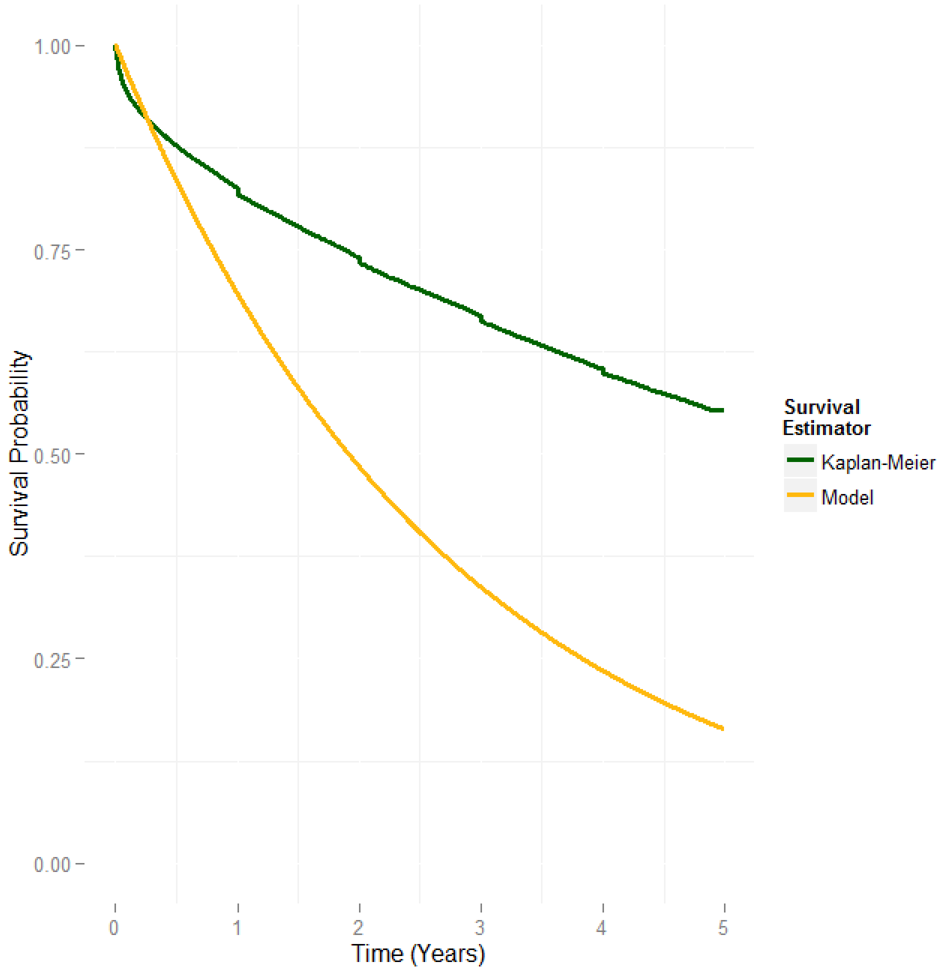


Figure 4.3: Analysis of the removal rate of vehicles, for contracts with 2 insured vehicles

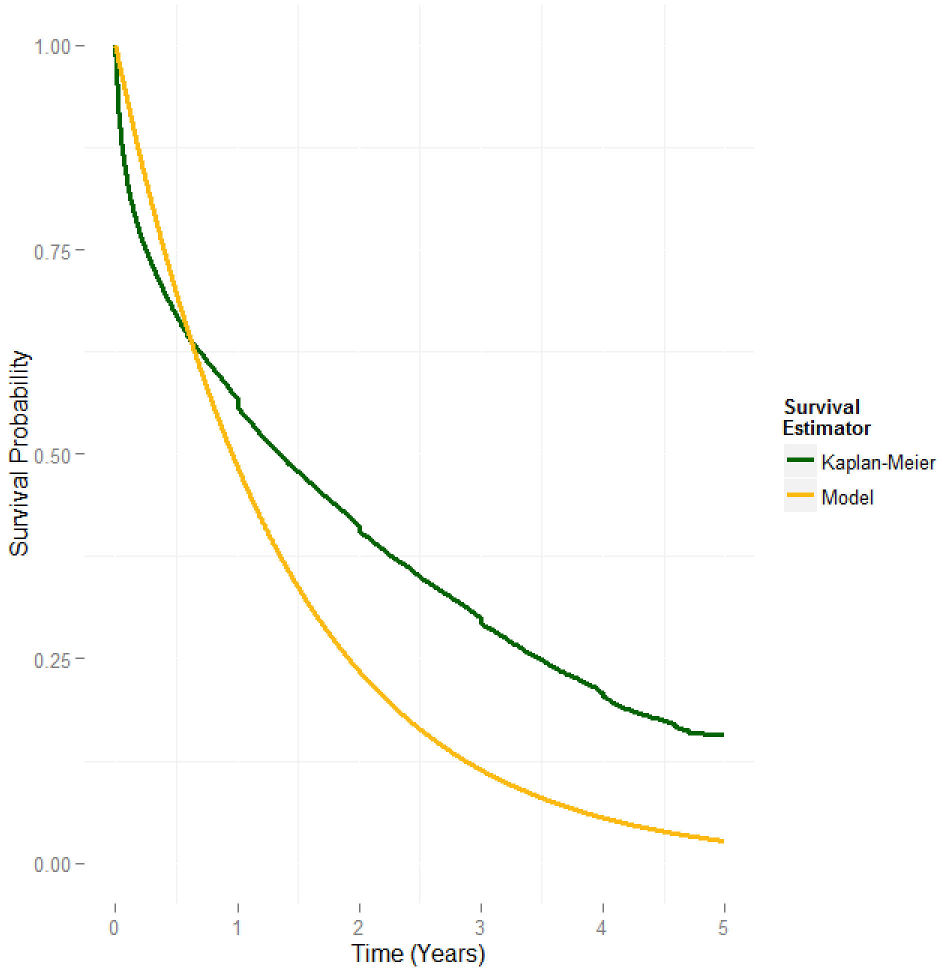


Figure 4.4: Analysis of the removal rate of vehicles, for contracts with 3 insured vehicles

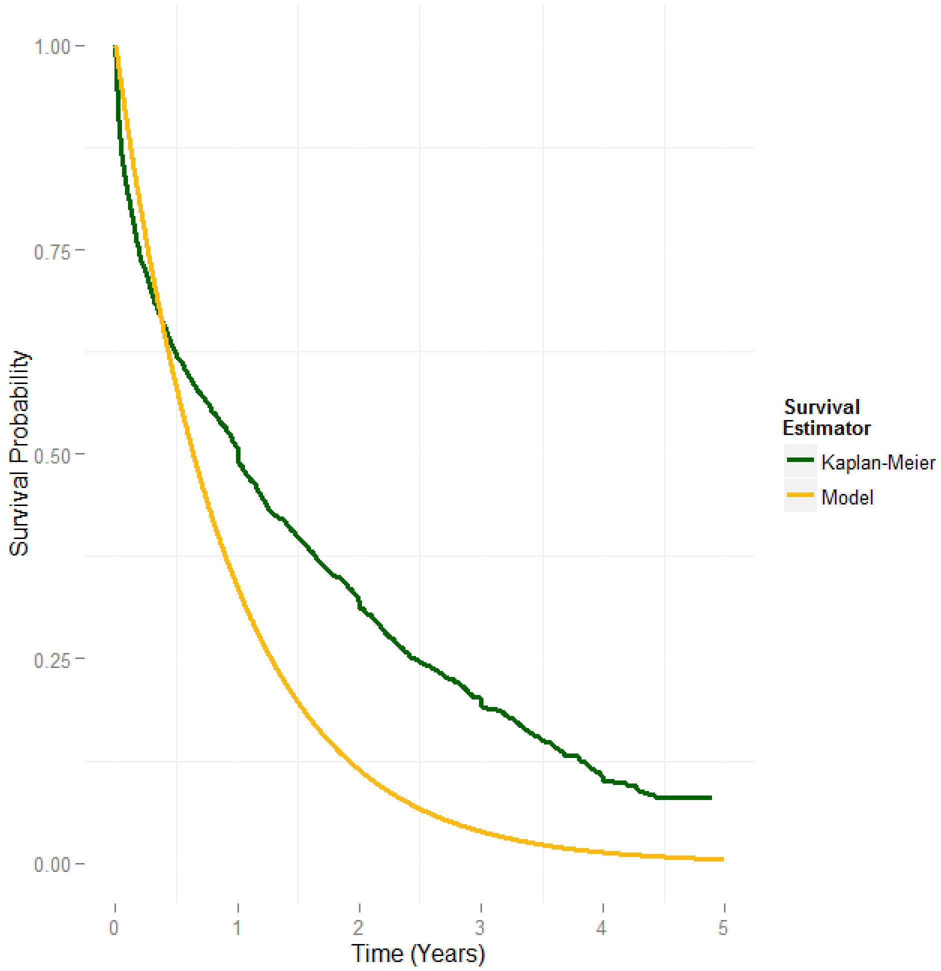


Figure 4.5: Analysis of the removal rate of vehicles, for contracts with 4 insured vehicles

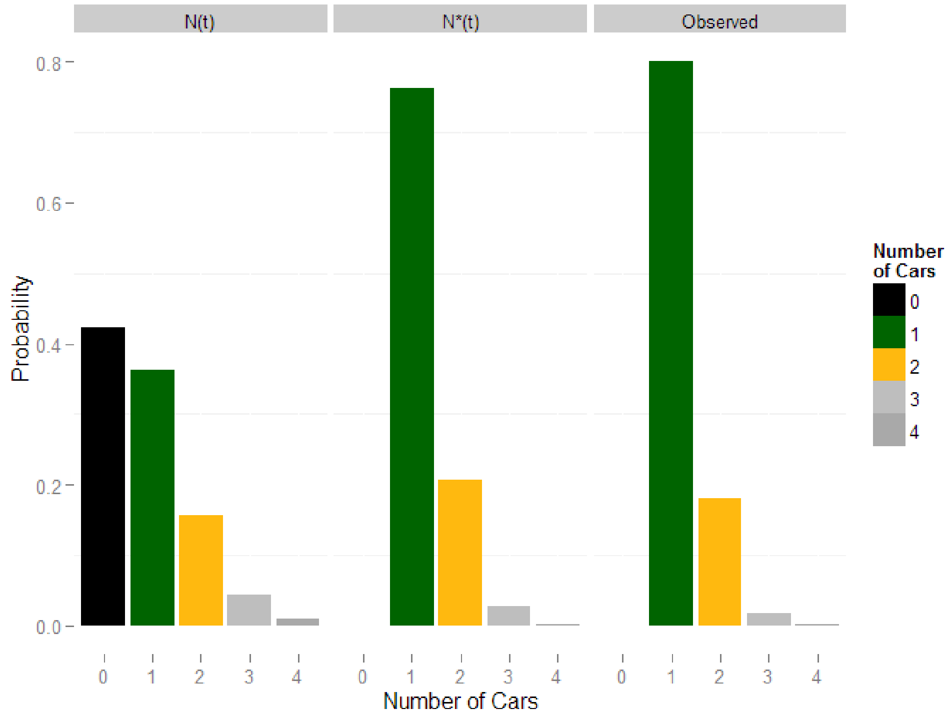


Figure 4.6: Initial number of insured cars for a new insurance policy

4. The rate of arrival $\tau$ of new households in the portfolio:

   Finally, we can note that $\widehat{\tau}$ is equal to 28,014. This value means that there would be approximately $\frac{28{,}014}{365.25} = 76.7$ new policyholders per day, or between $76.7 \times 30 = 2301$ and $76.7 \times 31 = 2377$ new households per month in the insurer's portfolio, depending on the number of days per month. As mentioned at the end of Section 2.2, the insurance data we used show seasonality, as illustrated in Figure 1.1. The number of days per month does not fully explain the observed seasonality.

## 4.4 Covariates

We know that some household profiles are more likely to be insured by a specific insurance company, and are also more or less likely to add or remove cars from their insurance contract. Similarly, some profiles may cancel more than others and certain types of policyholders may have lower or higher renewal rates. We still want to use the log-likelihood function expressed in equation (2.3), but think that the addition of covariates into each parameter $\lambda, \mu, \gamma, p$ and $\tau$ of our queuing process model is

| 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 |

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

| Parameter | $\boldsymbol{\beta}_\lambda$ | $\boldsymbol{\beta}_\mu$ | $\boldsymbol{\beta}_\gamma$ | $\boldsymbol{\beta}_p$ |
|---|---|---|---|---|
| $\beta_0$ | -2.5788 (0.0045) | -1.7056 (0.0043) | -2.7137 (0.0058) | 2.5989(0.0066) |
| $\beta_1$ | -0.0868 (0.0074) | 0.1562 (0.0071) | 0.2492 (0.0074) | -0.2541(0.0090) |
| $\beta_2$ | -0.2403 (0.0106) | 0.3735 (0.0103) | 0.0717 (0.0100) | 0.0784(0.0127) |
| $\beta_3$ | -0.4222 (0.0076) | 0.6885 (0.0071) | 0.5453 (0.0071) | -0.2628(0.0087) |

Table 4.3: Estimated parameters and standard errors (in brackets) for the process with covariates

justified.

Covariates selected to define the vector $\boldsymbol{X}_i$ of each household are provided in Table 4.4. Some of the covariates available in our database refer to calendar date, for example the fact that the effective date of the contract is in July, or the fact that the effective date of the insurance contact is on the first day of a month. We cannot use those covariates because this link with the calendar date would require us to change basic assumptions of our $M/G/\infty$ model. Thus, three explanatory variables are used, which create 8 different types of profiles, given that there are three binary explanatory variables (so $2^3 = 8$ types of insureds).

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 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 use of explanatory variables included in parameter $\tau$, corresponding to the arrival rate of new households in the portfolio, is more difficult. Indeed, parameter $\tau$ can be seen as a measure of the time between the arrival of two new insureds. Thus, it is not easy to introduce explanatory variables without changing the nature or the interpretation of the process. To solve this problem, because our numerical example uses very few covariates and thus few types of households, we use an estimator $\hat{\tau}_j, j = 1, ..., 8$, that is different for each type of household. This means that we assume 8 different arrival processes for the number of new households. To estimate the parameters, we

| Covariates | | | Parameters | | | | |
|---|---|---|---|---|---|---|---|
| $X1$ | $X2$ | $X3$ | $\widehat{\lambda}$ | $\widehat{\mu}$ | $\widehat{\gamma}$ | $\widehat{p}$ | $\widehat{\tau}$ |
| 1 | 1 | 1 | 0.0359 | 0.6142 | 0.1576 | 0.8966 | 133 |
| 1 | 1 | 0 | 0.0547 | 0.3085 | 0.0914 | 0.9186 | 111 |
| 1 | 0 | 1 | 0.0456 | 0.4228 | 0.1467 | 0.8891 | 2060 |
| 1 | 0 | 0 | 0.0696 | 0.2124 | 0.0851 | 0.9125 | 985 |
| 0 | 1 | 1 | 0.0391 | 0.5254 | 0.1229 | 0.9179 | 1187 |
| 0 | 1 | 0 | 0.0597 | 0.2639 | 0.0712 | 0.9357 | 1880 |
| 0 | 0 | 1 | 0.0497 | 0.3617 | 0.1144 | 0.9118 | 9596 |
| 0 | 0 | 0 | 0.0759 | 0.1817 | 0.0663 | 0.9308 | 12098 |

Table 4.4: Parameter values for each profile

took the likelihood equation (4.3) as a function to maximize for each household type. We obtain a likelihood function for the parameter $\tau_j$ from which we find the estimator $\hat{\tau}_j = \frac{B_j}{U}$, where $B_j$ is the number of new households of type $j = 1, \ldots, 8$.

Finally, the parameters for each type of insured appear in Table 4.4. We clearly see that the arrival rate varies greatly depending on the type of household.

## 4.5 Discussion

The proposed model seems to include all the possible client movement in the insurance industry: from arrivals and departures of households, to arrivals and departures of vehicles in each annual contract of each household. However, as we observed in Section 4.3 in the estimated parameters analysis, the $M/G/\infty$ model has some limitations. The model seems to approximate the departures of households and the arrivals of new vehicles on the contracts correctly, but the modeling of departures of vehicles and the absence of seasonality in the arrivals of households highlight some defects of the model. Obviously, the addition of covariates in the model improves the fit. However, we analyzed the model with covariates similarly to what has been done in Section 4.3, and the conclusions were approximately the same for each process of the model. Those conclusions about the quality of the approximation are based on a single insurer's portfolio. Depending on which dataset we analyze, the conclusions can be different and the quality of the fit could have been better. Nevertheless, future research that generalizes the approach and improves the approach and improve the flexibility of the model could be interesting. Such studies should focus on queuing models other than the $M/G/\infty$.

The proposed model is one of the first approaches proposed in the actuarial sciences literature to model client's movement in the insurance industry. The proposed approach opens the way to interesting generalization, such as the inclusion of ratemaking in the analysis or the inclusion of other insurance products. The model we built is an obvious approximation of what happens in an insurance portfolio, but allows us to compute several interesting values, such as the customer lifetime value and the global value of the insurance portfolio by using the customer equity concept (see next section). We estimate that the general fit of the model is good enough to draw general observations of what happens in the analyzed insurance portfolio. However, because of the approximation, we should draw conclusions prudently.

# 5 Analysis

In this section, applications are presented using the estimated parameters found by regression in the previous section and shown in Table 4.4. Thus, for example, even if we are working with $\widehat{\lambda}$, $\widehat{\gamma}$, $\widehat{\mu}$, $\widehat{p}$ or $\widehat{\tau}$, for simplicity we will suppose that those parameters correspond to $\lambda$, $\gamma$, $\mu$, $p$ and $\tau$.

We can compute different results using the PGFs found previously in Sections 2 and 3. Our first analysis will involve calculating the number of insured cars at time $t$, but this time for the entire portfolio of insureds. We will then discount the future profits generated by future insured cars, which allows us to value the portfolio.

## 5.1 Expected value and variance of the number of insured cars at time $t$

Below we calculate the expected number of insured cars at time $t$. Although it is possible to find this result by deriving the PGF of equation (3.10), the use of the conditional expectation leads to the same result in a simpler way. Starting with the expectation of equation (3.9), we develop:

$$
\begin{aligned}
& \mathbb{E}(\mathcal{J}(t)) \\
= \quad & \mathbb{E}\left[\sum_{i=1}^{R_0} \mathcal{H}_i(t) + \sum_{i=1}^{\mathcal{K}(t)} \mathcal{N}_i(t)\right] \\
= \quad & \mathbb{E}\left[\sum_{i=1}^{R_0} \mathcal{H}_i(t)\right] + \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{\mathcal{K}(t)} \mathcal{N}_i(t) | \mathcal{K}(t)\right]\right] \\
= \quad & \sum_{i=1}^{R_0} \mathbb{E}(\mathcal{H}_i(t)) + \mathbb{E}(\mathcal{K}(t)\mathbb{E}(\mathcal{N}(t))) \\
= \quad & \sum_{i=1}^{R_0} \mathbb{E}(\mathcal{H}_i(t)) + \frac{\lambda}{\mu}\mathbb{E}(\mathcal{K}(t)) \\
= \quad & \sum_{i=1}^{R_0} e^{-\gamma t} p^{\lfloor t+c_i \rfloor} \left((a_i - 1)e^{-\mu t} + 1 + \left(1 - e^{-\mu t}\right) \frac{\lambda}{\mu}\right) + \tau t q_t \left(\frac{\lambda}{\mu} + 1\right). \qquad (5.1)
\end{aligned}
$$

The variance or the standard deviation can also be computed using a similar method, from which we obtain:

$$
\begin{aligned}
& \mathbb{V}\mathrm{ar}(\mathcal{J}(t)) \\
= \; & \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{R_0}\mathcal{H}_i(t)+\sum_{i=1}^{\mathcal{K}(t)}\mathcal{N}_i(t)\right] \qquad (5.2) \\
= \; & \sum_{i=1}^{R_0}\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\sum_{j=1}^{\mathcal{M}_i(t)}\mathcal{N}_i(t,a_i)\middle|\mathcal{M}(t)\right]\right]+\sum_{i=1}^{R_0}\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\sum_{j=1}^{\mathcal{M}_i(t)}\mathcal{N}_i(t,a_i)\middle|\mathcal{M}(t)\right]\right] \\
& +\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\sum_{i=1}^{\mathcal{K}(t)}\mathcal{N}_i(t)\middle|\mathcal{K}(t)\right]\right]+\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{\mathcal{K}(t)}\mathcal{N}_i(t)\middle|\mathcal{K}(t)\right]\right] \\
= \; & \sum_{i=1}^{R_0}\mathbb{V}\mathrm{ar}\left[\mathcal{M}_i(t)\mathbb{E}[\mathcal{N}_i(t,a_i)]\right]+\sum_{i=1}^{R_0}\mathbb{E}\left[\mathcal{M}_i(t)\mathbb{V}\mathrm{ar}[\mathcal{N}_i(t,a_i)]\right] \\
& +\mathbb{V}\mathrm{ar}\left[\mathcal{K}(t)\mathbb{E}[\mathcal{N}(t)]\right]+\mathbb{E}\left[\mathcal{K}(t)\mathbb{V}\mathrm{ar}[\mathcal{N}(t)]\right] \\
= \; & \sum_{i=1}^{R_0}\mathbb{V}\mathrm{ar}\left[\mathcal{M}_i(t)\left[(a_i-1)e^{-\mu t}+1+\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\right]\right] \\
& +\sum_{i=1}^{R_0}\mathbb{E}\left[\mathcal{M}_i(t)\left[(a_i-1)e^{-\mu t}\left(1-e^{-\mu t}\right)+\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\right]\right] \\
& +\mathbb{V}\mathrm{ar}\left[\mathcal{K}(t)\left(\frac{\lambda}{\mu}+1\right)\right]+\mathbb{E}\left[\mathcal{K}(t)\frac{\lambda}{\mu}\right] \\
= \; & \sum_{i=1}^{R_0}e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\left(1-e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\right)\left((a_i-1)e^{-\gamma t}+1+\left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right)^2 \\
& +\sum_{i=1}^{R_0}e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\left((a_i-1)e^{-\mu t}\left(1-e^{-\mu t}\right)+\left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right) \\
& +\left(\frac{\lambda}{\mu}+1\right)^2\tau t q_t+\frac{\lambda}{\mu}\tau t q_t. \qquad (5.3)
\end{aligned}
$$

We computed those values using the insurer's database, and projected the number of insured cars after 1 year, 5 years and 10 years, as shown in Table 5.1. We also added a variation of plus or minus the standard deviation, thus showing the potential variability of results. Results shown depend on the model (meaning that model specification errors should be considered), but also on other errors, such as the variability of the estimators, which was not considered in this calculation.

By taking the limit of the expected value or by using the PGF when $t \to \infty$, we could see that the number of expected insured cars converges to a specific value when $t$ increases. Even if the expected number of new cars per year is high, we can observe that the standard deviation is low. We can see the expected number of insured cars per year in Figure 5.1, where the standard deviation

| $t$ | 1 | 5 | 10 |
|---|---|---|---|
| $\mathbb{E}(\mathcal{J}(t)) + \sqrt{\mathbb{V}\text{ar}(\mathcal{J}(t))}$ | 240,414 | 228,915 | 225,932 |
| $\mathbb{E}(\mathcal{J}(t))$ | 240,075 | 228,379 | 225,355 |
| $\mathbb{E}(\mathcal{J}(t)) - \sqrt{\mathbb{V}\text{ar}(\mathcal{J}(t))}$ | 239,737 | 227,843 | 224,779 |

Table 5.1: Expected value of the future number of insured cars

on the number of vehicles increases only slightly over time. Our prediction does not include the estimation error, and we think that the addition of this kind of error in the predictions would be more realistic. This element should be analyzed in future research.

The second graph of Figure 5.1 shows the expected number of insured cars from new households ($\mathcal{J}_N$) and the expected number of insured cars from old households ($\mathcal{J}_A$). Obviously, the sum of the two variables gives us the total number of cars insured, as shown by the equation $\mathcal{J}(t) = \mathcal{J}_N(t) + \mathcal{J}_A(t)$. In the case of the insurer analyzed in this paper, the total number of insured cars remains essentially the same over the years. However, we see that the old households are quickly replaced by new ones. Thus, after only five years, half of the insurer's portfolio will be composed of new households. This empirical analysis shows the importance of constantly conducting promotional campaigns to attract new customers.

It may be interesting to analyze whether this behavior is the same for all types of insureds. Table 5.2 provides an overview of $\mathcal{J}(t)$ for $t = 0$ and for $t \to \infty$. Figure 5.2 shows the evolution of the number of insured cars by risk profile.

We can see that the portfolio composition changes significantly over time. Because of substantial changes in the proportions of certain profiles in the portfolio, we can reasonably assume that the arrival rate of some households' profiles was different in the past. Thus, if those insured households are profitable, a marketing analysis could identify what changes to company policy have generated this difference.

## 5.2 Value of the Insurance Portfolio

It is possible to calculate the lifetime value of an insurance portfolio by discounting the future profits of each household. Indeed, for illustration, we will suppose that the insurance company makes a \$1 profit for each one-year car exposure. In Boucher and Couture-Piché (2015), the authors used

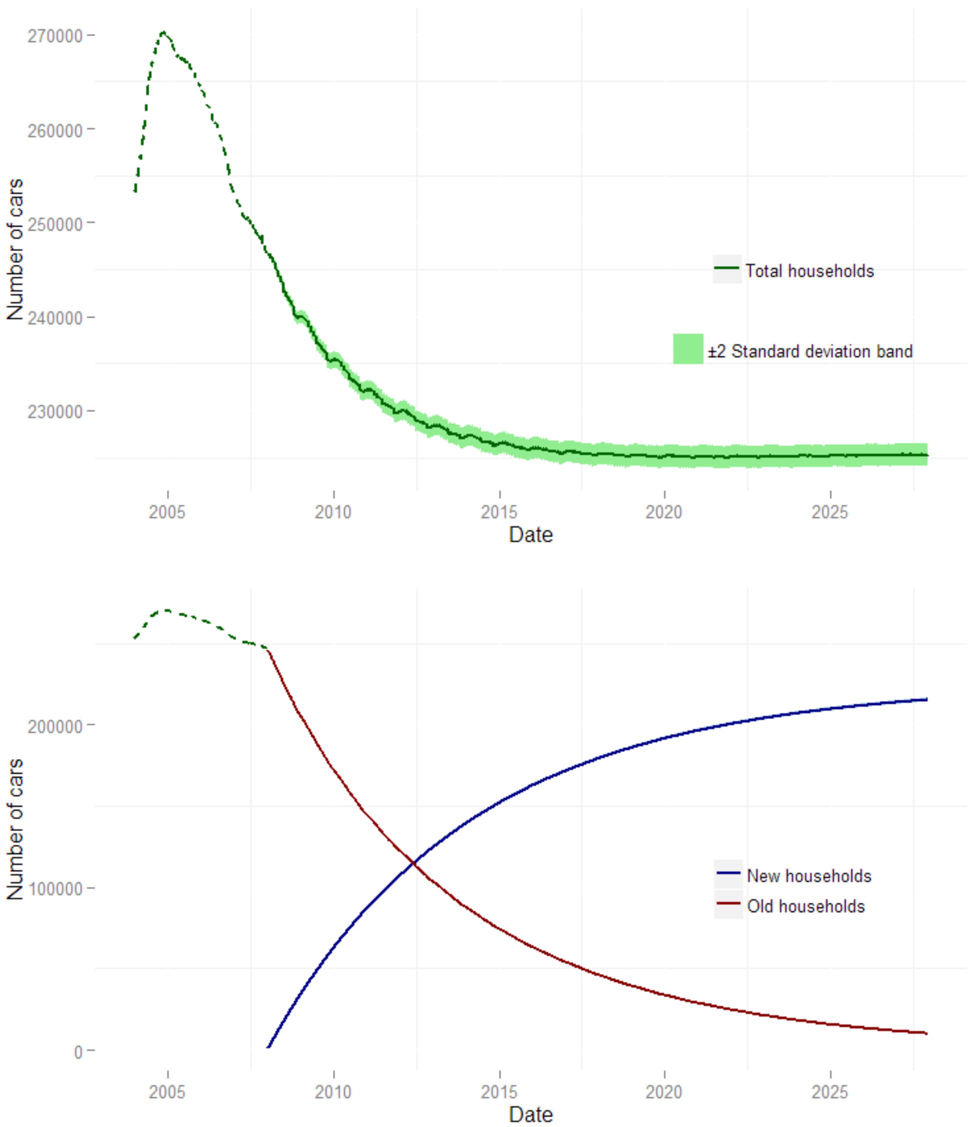


Figure 5.1: Expected number of insured cars over time (dotted line represents observed data)

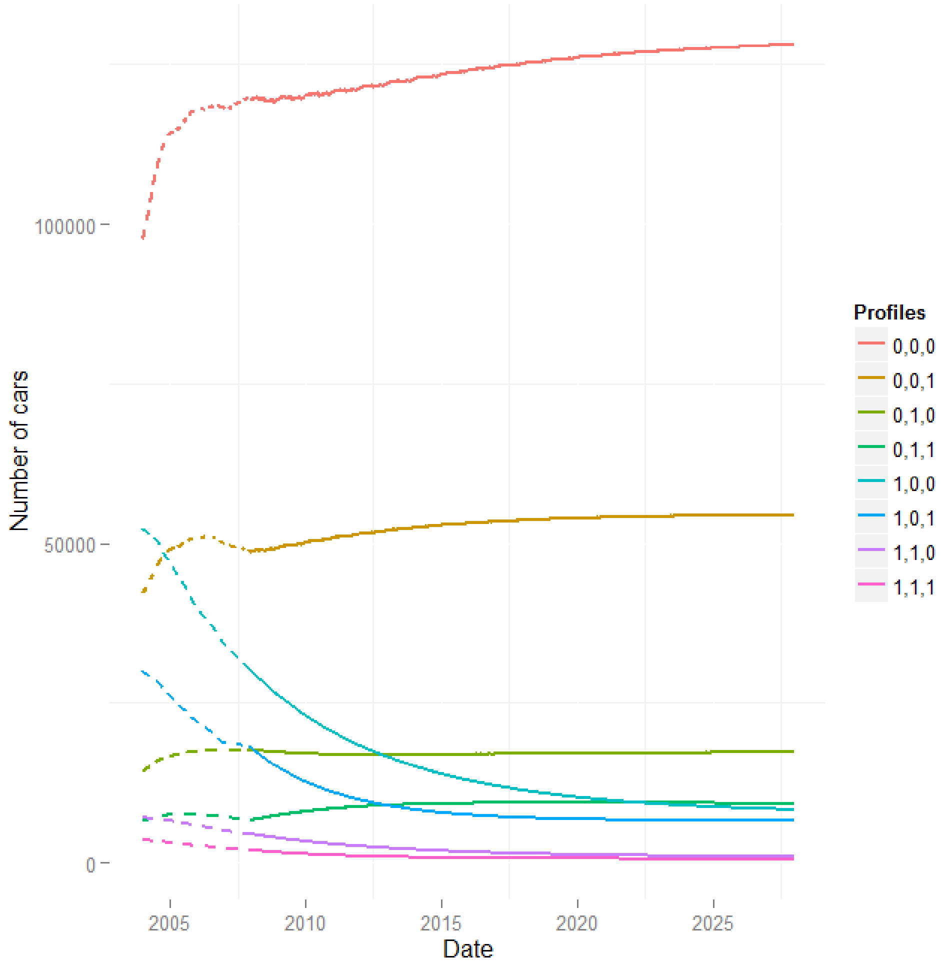


Figure 5.2: Expected number of insured cars over time by risk profile (dotted lines represent observed data)

| Covariates | | | Insured Cars at 12/31/2007 | | Insured cars at stationary level | |
|---|---|---|---|---|---|---|
| $X1$ | $X2$ | $X3$ | Number | Percentage | Number | Percentage |
| 1 | 1 | 1 | 1,929 | 0.8% | 5,56 | 0.2% |
| 1 | 1 | 0 | 4,432 | 1.7% | 7,73 | 0.3% |
| 1 | 0 | 1 | 17,977 | 7.6% | 9,141 | 4.0% |
| 1 | 0 | 0 | 30,040 | 12.3% | 7,741 | 3.4% |
| 0 | 1 | 1 | 6,579 | 2.6% | 6,375 | 2.8% |
| 0 | 1 | 0 | 17,657 | 6.8% | 17,302 | 7.7% |
| 0 | 0 | 1 | 48,605 | 20.0% | 55,221 | 24.5% |
| 0 | 0 | 0 | 119,445 | 48.2% | 128,735 | 57.0% |
| Total | | | 246,664 | 100.0% | 225,845 | 100.0% |

Table 5.2: Expected number of insured cars, and proportion, over time by risk profile

this method. We will apply the same method to all clients who are currently insured, but also to those who will be insured by the insurance company in the future. This allows us to obtain the total present value of the profit of the insurance portfolio. This value is sometimes called customer equity (see Rust et al., 2004 for an analysis of the links between this method and customer lifetime value), and can be seen as a measure of the health of the insurance company.

By noting $\Phi$ the random variable representing the value of this portfolio, we can find the expected value as follows:

$$\begin{aligned}
\mathbb{E}(\Phi) &= \int_0^\infty \mathbb{E}(\mathcal{J}(t))e^{-\delta t}dt \\
&= \int_0^\infty \sum_{i=1}^{R_0} \mathbb{E}(H_i(t))e^{-\delta t} + \left(\frac{\lambda}{\mu}+1\right)\tau t q_t e^{-\delta t}dt \\
&= \sum_{i=1}^{R_0} \int_0^\infty \mathbb{E}(H_i(t))e^{-\delta t}dt + \left(\frac{\lambda}{\mu}+1\right)\int_0^\infty \tau t q_t e^{-\delta t}dt
\end{aligned} \tag{5.4}$$

The BCP model was used to compute $\int_0^\infty \mathbb{E}(H_i(t))e^{-\delta t}dt$. The same computation cannot be used directly in our model. Indeed, some basic assumptions about the model have changed, such as the one that prohibits the possibility of having an insured household without insured cars. Consequently, we have:

$$\begin{aligned}
&\int_0^\infty \sum_{i=1}^{R_0} \mathbb{E}(H_i(t))e^{-\delta t}dt \\
= &\sum_{i=1}^{R_0} \left(a_i - 1 - \frac{\lambda}{\mu}\right) e^{-(1-c_i)(\gamma+\delta+\mu)} \frac{pe^{-(\gamma+\delta+\mu)}\left(1-e^{-(\gamma+\delta+\mu)}\right)}{\left(1-pe^{-(\gamma+\delta+\mu)}\right)(\gamma+\delta+\mu)} \\
&+ \left(\frac{\lambda}{\mu}+1\right) e^{-(1-c_i)(\gamma+\delta)} \frac{pe^{-(\gamma+\delta)}\left(1-e^{-(\gamma+\delta)}\right)}{\left(1-pe^{-(\gamma+\delta)}\right)(\gamma+\delta)} \\
&+ \left(a_i - 1 - \frac{\lambda}{\mu}\right) \frac{1-e^{-(1-c_i)(\gamma+\delta+\mu)}}{\gamma+\delta+\mu} \\
&+ \left(\frac{\lambda}{\mu}+1\right) \frac{1-e^{-(1-c_i)(\gamma+\delta)}}{\gamma+\delta}.
\end{aligned} \tag{5.5}$$

| Covariates | | | Old Households | | New Households | | Total | |
|---|---|---|---|---|---|---|---|---|
| $X1$ | $X2$ | $X3$ | Value | % | Value | % | Value | % |
| 1 | 1 | 1 | 6,481 | 0.5% | 25,867 | 0.3% | 32,349 | 0.3% |
| 1 | 1 | 0 | 22,156 | 1.7% | 35,630 | 0.4% | 57,786 | 0.5% |
| 1 | 0 | 1 | 62,223 | 4.7% | 426,630 | 4.3% | 488,852 | 4.3% |
| 1 | 0 | 0 | 149,871 | 11.4% | 342,266 | 3.4% | 492,137 | 4.3% |
| 0 | 1 | 1 | 27,754 | 2.1% | 299,241 | 3.0% | 326,995 | 2.9% |
| 0 | 1 | 0 | 105,349 | 8.0% | 757,722 | 7.6% | 863,071 | 7.6% |
| 0 | 0 | 1 | 210,379 | 16.0% | 2,486,015 | 24.8% | 2,696,394 | 23.8% |
| 0 | 0 | 0 | 730,249 | 55.6% | 5,644,101 | 56.3% | 6,374,350 | 56.3% |
| Total | | | 1,314,462 | 100% | 10,017,472 | 100% | 11,331,934 | 100% |

Table 5.3: Customer Equity for each risk profile

The second integral of (5.4) is calculated as follows:

$$
\begin{aligned}
& \int_0^\infty \tau t q_t e^{-\delta t} dt \\
= & \int_0^\infty \frac{\tau}{\gamma} \left[ \left(1 - e^{-\gamma}\right) \frac{1 - \left(pe^{-\gamma}\right)^{\lfloor t \rfloor}}{1 - pe^{-\gamma}} + p^{\lfloor t \rfloor} \left( e^{-\gamma \lfloor t \rfloor} - e^{-\gamma t} \right) \right] e^{-\delta t} dt \\
= & \sum_{i=0}^{\infty} \int_i^{i+1} \frac{\tau}{\gamma} \left[ \left(1 - e^{-\gamma}\right) \frac{1 - \left(pe^{-\gamma}\right)^{i}}{1 - pe^{-\gamma}} + p^{i} \left( e^{-\gamma i} - e^{-\gamma t} \right) \right] e^{-\delta t} dt \\
= & \sum_{i=0}^{\infty} \frac{\tau}{\gamma} \left[ \left(1 - e^{-\gamma}\right) \frac{1 - \left(pe^{-\gamma}\right)^{i}}{1 - pe^{-\gamma}} \frac{e^{-\delta i}\left(1 - e^{-\delta}\right)}{\delta} \right. \\
& \left. + \left(pe^{-(\gamma+\delta)}\right)^i \left( \frac{1 - e^{-\delta}}{\delta} - \frac{1 - e^{-(\delta+\gamma)}}{\delta + \gamma} \right) \right] \\
= & \frac{\tau}{\gamma} \left[ \frac{1 - e^{-\gamma}}{\delta\left(1 - pe^{-\gamma}\right)} \left( 1 - \frac{1 - e^{-\delta}}{1 - pe^{-(\gamma+\delta)}} \right) + \frac{1 - e^{-\delta}}{\delta\left(1 - pe^{-(\gamma+\delta)}\right)} \right. \\
& \left. - \frac{1 - e^{-(\delta+\gamma)}}{(\gamma+\delta)\left(1 - pe^{-(\gamma+\delta)}\right)} \right]
\end{aligned}
\tag{5.6}
$$

Using a value of $\delta$ equal to 0.02, the results of the discounted profits appear in Table 5.3 for each type of household. We separated the future profit by old households and new households.

From Table 5.3, we observe that the total value of the insurance portfolio is approximately \$11.3 million. Only \$1.3 million of this value comes from old households, meaning that only a small proportion of actual insureds participate in the long-term profits of the company. Consequently, our analysis can be used to justify that more efforts have to be made to attract new insureds, compared to the efforts made to keep actual insureds. It is important to understand that this analysis is based on the strong assumption that the insurer can make a $1 profit for each one-year car exposure. As an anonymous referee pointed out, however, this conclusion ignores the expenses associated with underwritting new business. Indeed, new business implies greater administrative expenses, and commissions, compared with existing business. A more precise model that could be based on what we propose on the paper would suppose that profit can depend on the characteristics of the household.

However, the analysis does not specify what is a new insured household is. For example, if an insured household cancels its current contract, it may want to be covered again by the same insurance company a few years later. The database used in our analysis cannot differentiate genuine new households from past clients that simply come back. Consequently, efforts to keep current clients should not be minimized following our analysis.

## 6 Conclusion

In this paper, we presented a generalization of the BCP model, whereby we model the number the number of insured households in an insurance company. We based our work on queuing theory models. We assume that the number of insured cars of each insured follows a similar process as the one proposed in the BCP model. However, we also added a new process that models the number of insured households in a portfolio. Because these two mathematical models are similar, models have been easily nested to form a complete model that emulates the total number of insured cars for an insurance company.

The model proposed in this paper requires five parameters: the arrival rate of new households in the portfolio, the departure rate of households from the portfolio, the probability of annual contract renewal, the arrival rate of new cars on a contract and finally the departure rate of cars from the contract. The model has been generalized to allow the use of explanatory variables in each of the model parameters. We compared the estimated parameters with the data to verify if the proposed model correctly approximates what happens in an insurance portfolio. We conclude that the general modeling is reasonable. More precisely, the model seems to correctly approximate the departures of households and the arrivals of new vehicle on the contract, but the modeling of the departure of vehicles and the absence of seasonality in the arrival of households highlight some flaws of the model. Those conclusions about the quality of the approximation are based on a single insurer portfolio, and it is possible that the model proposed in this paper fits better with other insurer's portfolios. Nevertheless, we think that the proposed model should be improved in the future to add flexbility, and to model some specific characteristics of the insurance process more accurately.

Parameters calculated from the model were used to generate various key statistics for an insurance company. The concept of customer equity was explored. This allowed us to note that some types of insureds, even if they represent a large proportion of the current insurance portfolio, represent only a small proportion of future profits of the insurance company. An insurer could therefore direct its marketing policies according to the model we proposed.

## Acknowledgements

The authors would like to thank anonymous referees for their careful reading of the manuscript and the resulting comments that greatly helped to improve the paper. Jean-Philippe Boucher would like to thank the financial support from the Natural Sciences and Engineering Research Council of Canada.

## References

[1] J.-P. Boucher, G. Couture-Piché (2015) Modeling the Number of Insureds' Cars Using Queuing Theory, *Insurance: Mathematics and Economics*, 64, 67-76.

[2] S. G., Eick, W. A. Massey, and W. Whitt (1993) $M_t$/G/$\infty$ queues with sinusoidal arrival rates, *Management Science*, 39(2), 241-252.

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

[4] 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.

[5] Guelman, L., Guillén, M. and Pérez-Marín, A.M. (2014) *A survey of personalized treatment models for pricing strategies in insurance* Insurance: Mathematics and Economics, 58(1), 68-76.

[6] Guelman, L., Guillén, M. and Pérez-Marín, A.M. (2015) *Uplift Random Forests* Cybernetics & Systems Special issue on Intelligent Systems in Business and Economics, in press.

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

[8] R. T. Rust, K. Lemon and V. A. Zeithaml (2004) Return on marketing: using customer equity to focus marketing strategy *Journal of Marketing, 68, 1, 109-127.*