

---

# MICRO-LEVEL PARAMETRIC DURATION-FREQUENCY-SEVERITY MODELING FOR OUTSTANDING CLAIM PAYMENTS

---

A PREPRINT

**Juan Sebastian Yanez**  
Department of Mathematics  
Université du Québec à Montréal  
Montreal, QC, Canada  
yanez.juan\_sebastian@uqam.ca

**Mathieu Pigeon**  
Department of Mathematics  
Université du Québec à Montréal  
Montreal, QC, Canada  
pigeon.mathieu.2@uqam.ca

February 1, 2021

## ABSTRACT

Unlike collective models, individual models have the advantage of keeping the attributes of each claim intact. We propose a three-component parametric individual model that uses this information in the form of explanatory variables. The first component predicts the delays between the occurrence, report, and closure of each claim using parametric survival models. For the second (frequency) and third (severity) components, we use generalized linear models and splice models. Moreover, the elapsed time between report and closure of claims is converted into an exposure variable in the count model. Finally, we discuss estimation procedures, make predictions, and compare the results with other models using a data set from a major Canadian insurance company.

**Keywords** loss reserving · individual models · survival · GLM

## 1 Introduction

Non-life insurance companies must control their solvency in order to protect their policyholders. Therefore, a provision or, loss reserve, must be established for claims whose total amount has not been paid or fully paid. Given the importance and the complexity of this task, several models have been proposed in the actuarial literature to predict future payments and to evaluate associated risks. Traditionally, these models can be grouped into two categories, collective and individual, based on the underlying data set. Although collective models have been studied by researchers for a long time and are commonly used by practitioners, individual models have caught the eye of researchers in the more recent years, and are rarely put into practice despite their many advantages. In this paper we aim to provide a parametric framework that can use micro-level information, which in turn may shed light on the advantage of using this information.

Let us begin by looking at the typical development of two claims, as illustrated in Figure 1. When accident  $\ell$  occurs, we can identify the delay between the beginning of the accident year and the exact occurrence date ( $t_\ell^{(o)}$ ). After an additional delay ( $t_\ell^{(r)}$ ), claim  $\ell$  is declared. For several situations (fire, damage to a car, etc.), this second delay may be short, but for other situations (bodily injury, civil liability), a longer period can separate the occurrence and reporting of a claim. Subsequently, one or more payments may be made (illustrated by dots in the figure) before closing the file after a final delay ( $t_\ell^{(c)}$ ). At an valuation date, claims can be separated into several categories according to the information available. For the remainder of this paper, the main categories are as follows:

- if the valuation date is between the date of the accident and the reporting date, the loss is considered not reported (*Incurred But Not Reported* or, IBNR), meaning the actuary has no information about the claim; and
- if the valuation date is between the reporting date and the closing date of the case, the loss is considered reported but not closed (*Reported But Not Settled* or, RBNS) meaning the actuary has only partial information about the claim.

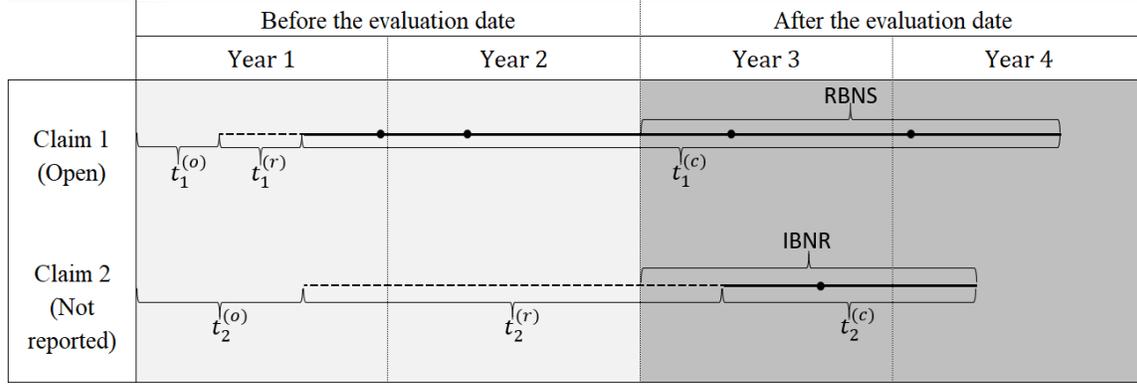


Figure 1: Development of two claims

In the literature, most stochastic collective models for loss reserving are presented in the widely used run-off triangles framework. This representation summarizes claim payments by aggregating them based on the accident and development year. In recent decades, intuitive and popular methods such as the stochastic Chain Ladder model ([20] and [19]) have been developed. Particularly, Wüthrich and Merz [28] and England and Verrall [8] made a compilation of the most widely used models within this collective structure. However, one of the shortcomings of these methods is that the underlying data used to summarize payments are based on the development of many different claims in spite of their individual characteristics, making it very difficult to use micro-level information in the modeling process.

In contrast to this class of models, individual models do not require aggregating payments from different claims; instead each claim is analysed separately. Several of the recently proposed approaches are based on techniques from the field of statistical learning. For examples, Wüthrich [29], Lopez *et al.* [17, 16], Lopez and Milhaud [18], Lopez [15] suggested using regression trees to predict the number of outstanding claim payments. Another implementation for predicting loss reserves was suggested in Baudry and Robert [4], using an algorithm called *ExtraTrees*. Duval and Pigeon [7] recommend using a Gradient Boosting procedure to predict both the IBNR and RBNS reserves. The examples allow the use of micro-level information for their predictions.

In parallel, some authors have put forth parametric models. For example, Haastруп and Arjas [9] proposed a Position Dependent Marked-Poisson Process (PDMPP) to predict claim payments continuously. In 2014, Antonio and Plat [1] presented a similar but more elaborated model, and the authors successfully applied it to a real data set. Although Haastруп and Arjas [9] suggested the possibility of using micro-level information in their model, Antonio and Plat [1] only considered individual covariates for payments severity. Antonio *et al.* [2] suggested an individual multi-state approach and applied it to a real data set. They also mention that micro-level information other than the occurrence and development period can be incorporated in their model. Furthermore, Zhao *et al.* [30] developed a semi-parametric model for individual claims, and Zhao *et al.* [31] incorporated copulae to predict IBNR claims. In contrast, some authors have considered a discrete framework for claim development. For example, Pigeon *et al.* [22] and Pigeon *et al.* [23] considered individual development factors. Verrall *et al.* [26] and, later Huang *et al.* [11], examined modeling claim counts and claim amounts separately, deriving a frequency-severity type structure.

In this paper, we propose an individual parametric model that fully takes advantage of micro-level information to predict outstanding claim payments. In retrospect, we drew inspiration from the two-component framework (frequency-severity), which is often used in ratemaking for Property & Casualty insurance. In this pricing context, general linear models or, GLM, are often used along with micro-level information (contract), to predict each component (for more information, see Ohlsson *et al.* [21]). Moreover, an exposure measure is considered for the frequency, usually in the form of the duration of the insurance contracts. However other forms of exposure may be considered (kilometers driven [5], etc.) In a loss reserving context, we considered a similar frequency-severity GLM framework for outstanding claims using micro-level information. We also establish an exposure measure based on the duration of each claim. However, unlike the contracts in ratemaking models, where the exposure measure is known beforehand, the duration of open claims is unknown at the valuation date. Therefore, an additional model needed to be fitted for this new component. We propose using parametric survival models that use micro-level information for this step, mainly to maintain a fully parametric structure similar to the one used for the aforementioned components, which in turn allows for a similar analysis of the significance of covariates in the fitting process across all steps. Note that in contrast with the studies by Verrall *et al.* [26] and Huang *et al.* [11], which also contain a frequency-severity type of structure, we seek to model single payments instead of the total cost of claims.

Comparatively to the existing literature for parametric individual loss reserving, our model diverges from methods previously suggested, such as models that use development factors, as in Pigeon et al. [22] and Pigeon et al. [23], or models that use Poisson Processes, as in Antonio and Plat [1] and Zhao et al. [30]. Indeed, instead of making use of a Poisson Process to model the delays between payments and closure of claims simultaneously, we suggest a more hierarchical structure in which we model first the duration of the claim, through survival models, and then the frequency, through discrete time parametric modeling for claim payment counts. To the best of our knowledge this is the first individual parametric loss reserving model that suggests modeling these two type of events separately, which in turn allows us to introduce the possibility of modeling payment counts discretely instead of continuously. This new structure offers flexibility in the choice of the distributions for the both the duration and the frequency, all while offering a straightforward way of introducing micro-level information from claims. Furthermore, modeling the cost of single payments have already been studied by Antonio and Plat [1], who recommend Burr, Lognormal and generalized linear models, and Denuit and Trufin [6] who favored a mixture of Gamma and Pareto distribution. However, we make a contribution in the modeling of this component as well by suggesting a new model based on splices, drawing inspiration from Laudagé et al. [13].

In a nutshell, we propose, in this paper

- a new 3-component framework for outstanding claim payments, where the unknown exposure is based on the duration, and where micro-level information can be included at all levels (duration-frequency-severity);
- a methodology to compare our model with either collective approaches based on run-off triangles (e.g., Mack’s model), or with other individual approaches proposed in the literature (e.g., individual model proposed in Antonio and Plat [1]).

In this paper we aim to implement the use of claims covariate information in the prediction of loss reserves. However, we must consider the different kinds of covariates that are available to us. Taylor et al. [25] use this information for their individual loss reserve model and, consider three types of covariates: static, time dynamic and, unpredictable dynamic. Static covariates do not change as time passes, however dynamic covariates will. Furthermore, even though both dynamic types are affected by time, only time dynamic covariates can be predicted with certainty, which in turn makes unpredictable dynamic covariates delicate to work with. An additional model is required in order to predict these uncertain values after the valuation date. In this paper, we will consider only static and time-dependent dynamic covariates because the additional model required for unpredictable dynamic covariates could be very specific, depending on which variable we are looking at.

This paper is structured as follows. In Section 2, we look at the general framework of the three-component model. In Section 3, we discuss the simulation procedure of the IBNR and RBNS reserves. In Section 4, we describe the data set used, followed by the numerical results of both our model and other comparative models. Finally, Section 5 contains concluding remarks and mentions further topics that could be explored based on this work.

## 2 Statistical model

In this section, we define both the individual and collective perspectives of a given portfolio, because we want our model to be interpreted from both perspectives.

On the one hand, in a micro-level structure, let  $\mathcal{L} = \mathcal{L}^{(O)} \cup \mathcal{L}^{(C)}$ , represent a set containing  $L$  reported claims in a portfolio, where  $\mathcal{L}^{(O)}$  and  $\mathcal{L}^{(C)}$  are the subsets containing open and closed (RBNS) claims, respectively. Let  $\mathcal{L}^*$  be the set containing incurred but not reported (IBNR) claims, which is, obviously, unavailable at the valuation date.

On the other hand, in a macro-level structure, let  $i$  and  $j$  be, respectively, the occurrence and the development periods in a run-off triangle, or loss triangle. Also, let  $Y_{i,j}$  be the total paid amount between time  $i - 1 + j$  and  $i + j$  from claims occurring during period  $i$ , where  $i = 1, \dots, I$  and  $j = 0, \dots, (I - 1)$ . For example, Figure 2 illustrates an incremental loss triangle with five occurrence and development periods.

Let us suppose that the insurance company has additional details about the accident, the insured, etc. and, wants to use them in the modeling process. Furthermore, let us suppose that all the covariates become available as soon as the claim is reported. The information regarding one observed claim  $\ell$  can be summarized by  $g$  categorical and/or continuous covariates,

$$\mathbf{c}_\ell = [c_{\ell,1}, c_{\ell,2}, \dots, c_{\ell,g}], \text{ for } \ell \in \mathcal{L}.$$

Having defined these variables, we can now consider the three components of the model in detail. We present the duration component in Subsection 2.1, the frequency component in Subsection 2.2, and the severity component in

		Development Period (j)				
		0	1	2	3	4
Occurrence Period (i)	1	$Y_{1,0}$	$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	$Y_{1,4}$
	2	$Y_{2,0}$	$Y_{2,1}$	$Y_{2,2}$	$Y_{2,3}$	$\hat{Y}_{2,4}$
	3	$Y_{3,0}$	$Y_{3,1}$	$Y_{3,2}$	$\hat{Y}_{3,3}$	$\hat{Y}_{3,4}$
	4	$Y_{4,0}$	$Y_{4,1}$	$\hat{Y}_{4,2}$	$\hat{Y}_{4,3}$	$\hat{Y}_{4,4}$
	5	$Y_{5,0}$	$\hat{Y}_{5,1}$	$\hat{Y}_{5,2}$	$\hat{Y}_{5,3}$	$\hat{Y}_{5,4}$

Figure 2: Incremental loss development triangle

Subsection 2.3. To better illustrate how each of these components could be obtained from a real data set, we provide examples in Appendix A.

### 2.1 Duration component

The duration component can be defined as the delay between the beginning of the occurrence period and the closure of a given claim. Thus, for a claim  $\ell$ , the component is constructed from the three following parts:

- $T_\ell^{(o)}$  the occurrence delay, i.e. the time elapsed between the beginning of the occurrence year and the exact occurrence date;
- $T_\ell^{(r)}$  the declaration delay, i.e. the time elapsed between the exact occurrence date and the reporting date; and
- $T_\ell^{(c)}$  the closure delay, i.e. the time elapsed between the reporting date and the closure date.

Because the claims we consider at are either open, closed or non-reported, Figure 3 represents how the delays are observed at the valuation date for claims having the same occurrence period at these three stages of development in a loss triangle.

	Before the evaluation date			After the evaluation date	
	$Y_{i,0}$	...	$Y_{i,l-i}$	$\hat{Y}_{i,l-i+1}$	...
Claim 1 (Closed)					
Claim 2 (Open)					
Claim 3 (Not reported)	(data not available)				

Figure 3: Observed delays for claims at valuation date in a loss triangle

Recall  $t_\ell^{(o)}$  is the delay from the beginning of the occurrence period. We have covariate information about open claims, and we also know the full extent of the occurrence and declaration delays. Furthermore, for  $\ell \in \mathcal{L}^{(O)}$ , we only have partial information about the closure delay, in the form of a right-censored observation,  $t_\ell^{(e)}$ , the delay between the reporting date and the valuation date. These values coupled with the observed values of closed claims (which are similar except that the closure delay is not a censored observation) constitute the full extent of the training data set. Moreover, not reported claims are undisclosed to the insurer at the valuation date and therefore cannot be included in the training set.

In this paper, we suppose that no claim can reopen after it has been closed for the first time. However, this model can include reopening events by simply adding a variable, the delay between the closure date and a later reopening date. Given the assumption, only open and not reported claims need a reserve to be calculated. Furthermore, the predictions for the future payments of open and not reported claims constitute the full extent of the RBNS and the IBNR reserves

respectively. In fact, the prediction of claims at these two stages of development bring about different challenges in the modeling process of the different delays and are explained separately.

### 2.1.1 Open claims

To begin, we need to find a model for the closure delay of open claims,

$$\left(T_\ell^{(c)} | \mathbf{c}_\ell\right), \text{ for } \ell \in \mathcal{L}^{(O)}. \quad (1)$$

However, at the valuation date these delays have been partial observed. Therefore, we need to find the conditional distribution,

$$\left(T_\ell^{(c)} | T_\ell^{(c)} > t_\ell^{(e)}, \mathbf{c}_\ell\right), \text{ for } \ell \in \mathcal{L}^{(O)}. \quad (2)$$

Nevertheless, depending on the chosen distribution, and the number of covariates, making simulations from the conditional distribution (2) can be simplified by simulating from (1), and then using an acceptance-rejection method to keep only the delays greater than  $t_\ell^{(e)}$ . For this reason, in this subsection we focus on modeling (1), and in Section 3, we further explain the simulation procedure of the conditional delays.

We suggest using parametric survival models for this step. Chapter 1 of the book by Lawless [14] contains some of the distributions that can be considered, among them the Weibull, Log-Logistic, Lognormal and Gamma distributions. Recall that the training set contains non-censored and right-censored observations of the closure delay. For a parametric survival distribution, the likelihood function for the closure delay is

$$\begin{aligned} \Lambda^{(RBNS)}(\Theta_{T^{(c)}}) = & \prod_{\ell \in \mathcal{L}^{(C)}} \left\{ f_{(T_\ell^{(c)} | \mathbf{c}_\ell)} \left( (t_\ell^{(c)} | \mathbf{c}_\ell); \Theta_{T^{(c)}} \right) \right\} \times \\ & \prod_{\ell \in \mathcal{L}^{(O)}} \left\{ S_{(T_\ell^{(c)} | \mathbf{c}_\ell)} \left( (t_\ell^{(e)} | \mathbf{c}_\ell); \Theta_{T^{(c)}} \right) \right\}, \end{aligned} \quad (3)$$

where,  $f()$  and  $S()$ , are the probability density function (pdf) and the survival function of the closure delay for reported claims, respectively and,  $\Theta_{T^{(c)}}$  is the parameter vector.

### 2.1.2 Not reported claims

IBNR claims are more complex to predict because the insurer does not have any information about them. They only know that some probably have occurred, and they will be reported at some point after the valuation date. Therefore, the number of IBNR claims is also unknown and, must be predicted. In Section 3, we suggest a procedure to predict how many not reported claims have occurred at each period  $i$ , however this particular model is better explained after defining the duration component for this type of claim. Thus, in this subsection, we suppose that  $i_\ell$ , the occurrence period of claim  $\ell$ , is known. Moreover, this section concerns the different time measures of the model, i.e. we focus on modeling the delays of a single not reported claim  $\ell^* \in \mathcal{L}^*$ , while supposing that the occurrence period is known.

Unlike open reported claims, we need to predict all three delays, instead of just completing the closure delay. Admittedly, as mentioned in the introduction of this paper, one of the advantages of micro-level models is that we can use individual information better than macro-level models. This is very much true for RBNS claims however its less so for IBNR claims. In spite of lack of micro-level information available when modeling non reported claims, individual models, unlike collective models, have the advantage of being able to predict IBNR and RBNS reserves separately, allowing the insurer to have insight about the weight of not reported claims in the portfolio.

The first delay we need to model is the occurrence delay,  $T_\ell^{(o)}$ , using  $\ell \in \mathcal{L}$ . The key aspect of this delay is that its observation period is  $(0, 1]$ , i.e. a year. Moreover, because we are modeling unreported claims, micro-level information is unknown and thus, we suggest fitting models based on seasonal effects, for example. Nevertheless, this time measure varies greatly depending on the data set, and many models can be considered. Examples of distributions that can be examined are the multinomial distribution, assigning a probability of a claim occurring within a certain time-window (e.g. each month of a year) or, the empirical distribution.

The other delay that needs to be modeled is the reporting delay,  $T_\ell^{(r)}$ , for which many observations generally occur within only a few days after the occurrence date. In order to take into account this large number of short time observations and those that take more time, Antonio and Plat [1] suggest using a mixture of a Weibull distribution with  $D$  degenerate components, where each component  $d$ , represents the number of days that have passed since the occurrence date,  $d = 0, 1, \dots, D - 1$  days. Furthermore, contrarily to the observed closure delay, the reporting delay does not contain censored observations, but the delays are truncated by the valuation date. Thus, the likelihood of the reporting delay can be written as

$$\Lambda^{(IBNR)}(\Theta_{T^{(r)}}^*) = \prod_{\ell \in \mathcal{L}} \frac{f_{T_\ell^{(r)}}^*(t_\ell^{(r)}; \Theta_{T^{(r)}}^*)}{F_{T_\ell^{(r)}}^*(I - i_\ell + 1 - t_\ell^{(o)}; \Theta_{T^{(r)}}^*)},$$

where,  $f^*(\cdot)$  and  $F^*(\cdot)$  are the pdf and the cumulative distribution function of the reporting delay for not reported claims, respectively and,  $\Theta_{T^{(r)}}^*$  is the parameter vector. Having modeled the reporting and the occurrence delay, let  $U_\ell^{(r)}$  be the delay between the beginning of the occurrence period and the report date, thus,

$$U_\ell^{(r)} = T_\ell^{(o)} + T_\ell^{(r)}.$$

By definition, the report date of an IBNR claim must happen after the valuation date. Moreover, for claim  $\ell \in \mathcal{L}^*$ , we can obtain the delay between the beginning of the occurrence year and the valuation date,  $I - i_\ell + 1$ . Thus,

$$U_\ell^{(r)} > I - i_\ell + 1, \text{ for } \ell \in \mathcal{L}^*.$$

This in turn means that we need to find the distribution of  $(U_\ell^{(r)} | U_\ell^{(r)} > I - i_\ell + 1)$ , for  $i = 1, \dots, I$ . We suggest obtaining these distributions numerically through a simulation of both  $T_\ell^{(o)}$  and  $T_\ell^{(r)}$ . More details are given in Section 3.

The final delay that needs to be addressed is the closure delay,  $T_\ell^{(c)}$ , where the same parametric distributions can be used in a similar manner to the the one suggested in Subsection 2.1.1, with the only difference being that most of the explanatory variables are missing, thus the likelihood to model the closure delay of IBNR in the training set is defined as:

$$\Lambda^{(IBNR)}(\Theta_{T^{(c)}}^*) = \prod_{\ell \in \mathcal{L}^{(c)}} \left\{ f_{(T_\ell^{(c)} | i_\ell)}^* \left( (t_\ell^{(c)} | i_\ell); \Theta_{T^{(c)}}^* \right) \right\} \times \prod_{\ell \in \mathcal{L}^{(o)}} \left\{ S_{(T_\ell^{(c)} | i_\ell)}^* \left( (t_\ell^{(e)} | i_\ell); \Theta_{T^{(c)}}^* \right) \right\}, \quad (4)$$

where,  $f^*(\cdot)$  and  $S^*(\cdot)$ , are the pdf and the survival function of the closure delay for not reported claims, respectively and,  $\Theta_{T^{(c)}}^*$  is the parameter vector.

## 2.2 The frequency component

Payments are time-framed by the reporting and closure dates because they can happen only between these two events (closure delay). In this subsection we aim to define a partition of this timeline in order to count the number of payments within each of the sub-intervals. We build this partition with two main goals in mind: (1) allow the structure of a run-off triangle to be easily reconstructed from our model, and (2) capture variations in individual development as a function of the time elapsed since the reporting date of each claim.

Let  $\mathcal{Q} = \{0, 1, \dots, I\}$  be a partition of the time interval between the beginning of the period in which the claim is reported and the maximum development time in a run-off triangle. This first division allows us to easily draw parallels between our results and those obtained using a collective approach based on a triangular structure.

We define a second partition of the same time interval which, this time, will not be constrained by the regularity of the construction of the run-off triangle, i.e. the fact that a loss triangle is generally divided into development periods of one year. We aim to capture the individual development of the frequency of claims from the report date until its closure, and

to identify, precisely, what stage of development each time division is in. This type of division is reminiscent of that used in the position dependent marked-Poisson process model proposed by Antonio and Plat [1].

Thus, let us begin by noting that the observation period of claim  $\ell$ , starting from the reporting date, is  $[0, \tau_\ell]$ , where  $\tau_\ell = \max \{t_\ell^{(c)}, t_\ell^{(e)}\}$ . Let  $\tau$  be the longest possible observation period, such that  $\tau = \max_{\ell \in \mathcal{L}} \{\tau_\ell\}$ . Let  $\mathcal{D} = \{\delta_0, \delta_1, \dots, \delta_{K-1}, \delta_K\}$  a partition of the interval  $[0, \tau]$ , and  $\mathcal{P}_\ell = \{0, \mathcal{D} + t_\ell^{(o)} + t_\ell^{(r)}, I\}$ . We assume that  $\tau < I - t_\ell^{(o)} - t_\ell^{(r)}, \forall \ell$ , but it is easy to adapt this definition if this inequality is not satisfied. Finally, we define the common refinement of  $\mathcal{Q}$  and  $\mathcal{P}_\ell$ , which consists of all the points of  $\mathcal{Q}$  and  $\mathcal{P}_\ell$ :

$$\mathcal{D}_\ell = \mathcal{Q} \vee \mathcal{P}_\ell = \{0, D_{\ell,1}, \dots, D_{\ell, M_\ell - 1}\},$$

where  $M_\ell$  is the number of points in  $\mathcal{D}_\ell$ .

The exposure corresponding to the  $k^{\text{th}}$  sub-interval of  $\mathcal{D}_\ell$  is

$$E_{\ell,k} = \begin{cases} 0, & k = 1 \\ D_{\ell,k} - D_{\ell,k-1}, & k \in \{k : k > 1, D_{\ell,k} \leq \tau_\ell + t_\ell^{(o)} + t_\ell^{(r)}\} \\ \tau_\ell + t_\ell^{(o)} + t_\ell^{(r)} - D_{\ell,k-1}, & k \in \{k : D_{\ell,k-1} < \tau_\ell + t_\ell^{(o)} + t_\ell^{(r)} < D_{\ell,k}\} \\ 0, & \text{elsewhere.} \end{cases}$$

Let  $N_{\ell,k}$  be a random variable which counts the number of payments for the claim  $\ell$  during the  $k^{\text{th}}$  sub-interval. The distribution of each  $N_{\ell,k}$ , is given by

$$(N_{\ell,k} | \mathbf{c}_\ell, \mathcal{D}_\ell) \sim \text{Dist}^{(n)} \left( E_{\ell,k} \cdot \mu_{\ell,k}^{(n)} \left( \boldsymbol{\beta}^{(n)} \right), \cdot \right), \quad \text{for } \ell \in \mathcal{L}, \text{ and} \quad (5)$$

$$(N_{\ell,k} | \mathcal{D}_\ell) \sim \text{Dist}^{*(n)} \left( E_{\ell,k} \cdot \mu_{\ell,k}^{*(n)} \left( \boldsymbol{\beta}^{*(n)} \right), \cdot \right), \quad \text{for } \ell \in \mathcal{L}^*, \quad (6)$$

with  $j = 0, \dots, I - 1$  and  $k = 2, \dots, M_\ell - 1$ .  $\text{Dist}^{(n)}$  and  $\text{Dist}^{*(n)}$  are the distributions of the number of payments from reported and unreported claims, respectively. Also,  $\mu_{\ell,k}^{(n)} \left( \boldsymbol{\beta}^{(n)} \right)$  and  $\mu_{\ell,k}^{*(n)} \left( \boldsymbol{\beta}^{*(n)} \right)$  are the mean parameters for one exposure unit, while,  $\boldsymbol{\beta}^{(n)}$  and  $\boldsymbol{\beta}^{*(n)}$  are the parameter vectors used to predict the mean. We explicitly mention  $j$  in our construction to ease comparison with collective approaches. If this is not the objective, it is possible to simplify the model and to keep only the second partition.

### 2.3 The severity component

With the final component we seek to predict the cost of single payment. Let  $\text{Dist}^{(x)}$  and  $\text{Dist}^{*(x)}$  be the distributions of the cost of single payments from reported and not reported claims respectively. Let  $X_{\ell,k,m}$  be the cost of the  $m^{\text{th}}$  payment from claim  $\ell$  occurring during the  $k^{\text{th}}$  sub-interval of  $\mathcal{D}_\ell$ . We have

$$(X_{\ell,k,m} | \mathbf{c}_\ell, \mathcal{D}_\ell) \sim \text{Dist}^{(x)}, \text{ for } m = 1, \dots, N_{\ell,k} \text{ and } \ell \in \mathcal{L}, \text{ and} \quad (7)$$

$$(X_{\ell,k,m} | \mathcal{D}_\ell) \sim \text{Dist}^{*(x)}, \text{ for } m = 1, \dots, N_{\ell,k}, \text{ and } \ell \in \mathcal{L}^*. \quad (8)$$

Although it is possible to consider generalized linear models, where the distribution is continuous on  $\mathbb{R}^+$ , modeling the severity of payments may require a more complex approach. This is due to the fact that payments may be highly diverse and, a model that can accommodate both large and small payments could be preferable. Antonio and Plat [1], for example, suggest using models such as the Burr and the Lognormal distributions. Alternatively, one may consider a mixture of distributions, for example Denuit and Trufin [6] suggest a discrete mixture of a Gamma and a Pareto distribution.

Finally, a splicing model can also be considered, of which some of the best known are the threshold models. Laudagé et al. [13] advocate using this method in the context of claim severity for rate making and this model can be

accommodated to model claim payments instead. This so-called *Threshold severity model* can better accommodate large payments by fitting the tail and the body of the distributions separately through a splicing point  $u$ , called a threshold. For a given payment,  $(X_{\ell,k,m} | \mathbf{c}_\ell, \mathcal{D}_\ell)$ , let  $h_{(X_{\ell,k,m})}(x_{\ell,k,m}; \Theta_{X^{(h)}})$  and  $g_{(X_{\ell,k,m})}(x_{\ell,k,m}; \Theta_{X^{(g)}})$  be the probability density functions of the bulk and the tail, respectively, with parameter vectors  $\Theta_{X^{(h)}}$  and  $\Theta_{X^{(g)}}$ . Let  $H_{(X_{\ell,k,m})}$  and  $G_{(X_{\ell,k,m})}$  be the respective cumulative distribution functions. The probability of exceeding the splicing point is given by  $q_{(X_{\ell,k,m})}(\Theta_{X^{(g)}})$ , where  $\Theta_{X^{(g)}}$  is the parameter vector. Thus, the pdf of  $(X_{\ell,k,m} | \mathbf{c}_\ell, \mathcal{D}_\ell)$  with parameter vector  $\Theta_X = (\Theta_{X^{(h)}}, \Theta_{X^{(g)}}, \Theta_{X^{(q)}})$  is given by

$$f_{(X_{\ell,k,m})}(x_{\ell,k,m}; \Theta_X) = \begin{cases} 0, & \text{for } x_{\ell,k,m} \leq 0 \\ (1 - q_{(X_{\ell,k,m})}(\Theta_{X^{(g)}})) \frac{h_{(X_{\ell,k,m})}(x_{\ell,k,m}; \Theta_{X^{(h)}})}{H_{(X_{\ell,k,m})}(u; \Theta_{X^{(h)}})}, & \text{for } 0 < x_{\ell,k,m} \leq u \\ q_{(X_{\ell,k,m})}(\Theta_{X^{(g)}}) g_{(X_{\ell,k,m})}(x_{\ell,k,m}; \Theta_{X^{(g)}}), & \text{for } x_{\ell,k,m} > u, \end{cases}$$

for  $m = 1, \dots, N_{\ell,k}$ .

### 3 Loss reserves

In this section we illustrate how to simulate both the IBNR and the RBNS reserves after fitting the three-component model defined in Section 2. For payments that happen after the valuation date, for  $(i + j > I)$ , we have

$$\begin{aligned} Y_{i,j} &= Y_{i,j}^{(RBNS)} + Y_{i,j}^{(IBNR)} \\ &= \sum_{\ell \in \{\ell | \ell \in \mathcal{L}^{(O)}, i_\ell = i\}} \sum_{k \in \{k: j < D_{\ell,k} \leq j+1\}} \sum_{m=1}^{N_{\ell,k}} X_{\ell,k,m} + \sum_{\ell \in \{\ell | \ell \in \mathcal{L}^*, i_\ell = i\}} \sum_{k \in \{k: j < D_{\ell,k} \leq j+1\}} \sum_{m=1}^{N_{\ell,k}} X_{\ell,k,m}. \end{aligned}$$

We recall that  $i_\ell$  is the occurrence period of claim  $\ell$ . Then, we calculate the total reserve:

$$\begin{aligned} R &= R^{(RBNS)} + R^{(IBNR)} \\ &= \sum_{(i+j>I)} Y_{i,j}^{(RBNS)} + \sum_{(i+j>I)} Y_{i,j}^{(IBNR)}. \end{aligned}$$

We can now describe the simulation procedure for both parts of the total reserve.

#### 3.1 IBNR reserve

Before we give the complete simulation procedure, we must predict the number of IBNR claims in the portfolio. Our approach is based on the work of Pigeon et al. [22] and on the distribution of  $U_\ell^{(r)}$  as defined in Subsection 2.1. Let  $L_i$ , the total number of claims occurring during period  $i$ , follow a Poisson distribution with occurrence measure  $\theta\omega_i$ , where  $\omega_i$  is the total exposure registered for period  $i$ , for  $i = 1, \dots, I$ . Because we only observe reported claims, the Poisson distribution should be thinned in the following way

$$L_i \sim \text{Poisson} \left( \theta\omega_i \Pr \left( U_\ell^{(r)} \leq I - i + 1 \right) \right).$$

Thus,  $L_i^*$ , the number of IBNR claim(s) from occurrence period  $i$  follows a Poisson distribution with occurrence measure given by

$$\theta\omega_i \Pr \left( U_\ell^{(r)} > I - i + 1 \right). \quad (9)$$

We can now proceed to the simulation procedure of an IBNR reserve.

- **Step 1:** Obtain  $\tilde{L}^* = \sum_i \tilde{L}_i^*$ , where  $\tilde{L}_i^*$  is the simulated value of  $L_i^*$  for each occurrence period (see Equation (9)).

- **Step 2:** For  $\ell = 1, \dots, \tilde{L}^*$ , go through each of the following sub-step.
  - **Step 2a:** Obtain  $\tilde{U}_{\ell,r}$ , the simulated value of  $\left(U_{\ell}^{(r)} | U_{\ell}^{(r)} > I - i_{\ell} + 1\right)$ , the delay between the beginning of the occurrence period and the exact reporting date, where

$$\Pr\left(U_{\ell}^{(r)} \leq u | U_{\ell}^{(r)} > I - i_{\ell} + 1\right) = \frac{\Pr\left(I - i_{\ell} + 1 < U_{\ell}^{(r)} \leq u\right)}{1 - \Pr\left(U_{\ell}^{(r)} \leq I - i_{\ell} + 1\right)}.$$

- **Step 2b:** Obtain  $\tilde{T}_{\ell,c}$ , the simulated value of  $\left(T_{\ell}^{(c)} | i_{\ell}\right)$ , the closure delay, where (see Equation (8))

$$\left(T_{\ell}^{(c)} | i_{\ell}\right) \sim Dist^{*(t^{(c)})}.$$

- **Step 2c:** Based on  $\tilde{U}_{\ell,r}$ ,  $\tilde{T}_{\ell,c}$  and  $\mathcal{D}$ , calculate  $\tilde{\mathcal{P}}_{\ell}$  and  $\tilde{\mathcal{D}}_{\ell} = \mathcal{Q} \vee \tilde{\mathcal{P}}_{\ell} = \{0, \tilde{D}_{\ell,1}, \dots, \tilde{D}_{\ell, \tilde{M}_{\ell}-1}\}$ .

- **Step 2d:** Obtain  $\tilde{N}_{\ell,k}$ , a simulated value of  $\left(N_{\ell,k} | \tilde{\mathcal{D}}_{\ell}\right)$  using Equation (6).

- **Step 2e:** Obtain  $\tilde{X}_{\ell,k,m}$ , a simulated value of  $\left(X_{\ell,k} | \tilde{\mathcal{D}}_{\ell}\right)$ , for  $m = 1, \dots, \tilde{N}_{\ell,k}$  using Equation (8).

- **Step 3:** Calculate the simulated IBNR reserve:

$$\tilde{R}^{(IBNR)} = \sum_{i+j>I} \tilde{Y}_{i,j}^{(IBNR)} = \sum_{\ell \in \tilde{\mathcal{L}}_i^*} \sum_{k \in \{k: j < \tilde{D}_{\ell,k} \leq j+1\}} \sum_{m=1}^{\tilde{N}_{\ell,k}} \tilde{X}_{\ell,k,m},$$

where  $\tilde{\mathcal{L}}_i^*$  is the set containing all the simulated IBNR claims having occurrence period  $i$ .

### 3.2 RBNS reserve

Let  $L^{(O)}$  be the total number of open claims in the portfolio. We describe the simulation procedure for the RBNS reserve below.

- **Step 1:** Set  $\ell = 1$ , the first open claim in  $\mathcal{L}^{(O)}$ .
  - **Step 1a:** Obtain  $\tilde{T}_{\ell,c}$ , the simulated value of  $\left(T_{\ell}^{(c)} | \mathbf{c}_{\ell}\right)$ , the closure delay of open claim  $\ell$ , where,
$$\left(T_{\ell}^{(c)} | \mathbf{c}_{\ell}\right) \sim Dist^{(t^{(c)})}.$$
  - **Step 1b:** If  $\tilde{T}_{\ell,c} > t_{\ell}^{(e)}$ , set  $\ell = \ell + 1$ , the next open claim.
  - **Step 1c:**
    - \* If  $\ell \leq L^{(O)}$ , go to **Step 1a**.
    - \* If  $\ell = L^{(O)} + 1$ , go to **Step 2**.
- **Step 2:** Based on  $U_{\ell,r}$ ,  $\tilde{T}_{\ell,c}$  and  $\mathcal{D}$ , calculate  $\mathcal{P}_{\ell}$  and  $\tilde{\mathcal{D}}_{\ell} = \mathcal{Q} \vee \mathcal{P}_{\ell} = \{0, \tilde{D}_{\ell,1}, \dots, \tilde{D}_{\ell, \tilde{M}_{\ell}-1}\}$ , for  $\ell = 1, \dots, L^{(O)}$ .
- **Step 3:** Obtain  $\tilde{N}_{\ell,k}$ , a simulated value of  $\left(N_{\ell,k} | \mathbf{c}_{\ell}, \tilde{\mathcal{D}}_{\ell}\right)$  using Equation (5), for  $\ell = 1, \dots, L^{(O)}$ .
- **Step 4:** Obtain  $\tilde{X}_{\ell,k,m}$ , a simulated value of  $\left(X_{\ell,k} | \mathbf{c}_{\ell}, \tilde{\mathcal{D}}_{\ell}\right)$ , for  $m = 1, \dots, \tilde{N}_{\ell,k}$  using Equation (7).
- **Step 5:** Calculate the simulated RBNS reserve:

$$\tilde{R}^{(RBNS)} = \sum_{i+j>I} \tilde{Y}_{i,j}^{(RBNS)} = \sum_{\ell \in \mathcal{L}_i^{(O)}} \sum_{k \in \{k: j < \tilde{D}_{\ell,k} \leq j+1\}} \sum_{m=1}^{\tilde{N}_{\ell,k}} \tilde{X}_{\ell,k,m},$$

where  $\mathcal{L}_i^{(O)}$  be the set containing all the open RBNS claims having occurrence period  $i$ .

## 4 Numerical Analysis

In this section we provide a detailed analysis based on a real data set from a Canadian Property & Casualty insurance company. With this example, we want to (1) illustrate the use of our new 3-component framework, (2) perform a comparison with collective approaches, and (3) perform a comparison with another individual approach. We describe our data set in Subsection 4.1, we adjust our model, as well as various collective models in Subsection 4.2, ... in Subsection 4.3, and finally, we analyze our results in Subsection 4.4.

### 4.1 Data set

The data set we worked on contains transactional information for 57.593 claims occurring between January 1, 2011 and, December 31, 2015. The insurer recorded each important event (payment, case estimates, closure date, etc.), along with micro-level information until December 31, 2017. For our numerical analysis, we set the valuation date to December 31, 2015, where there were 48.855 closed claims, 7.872 open claims and, 866 not reported claims.

Some payments from the data set were not considered based on macro and micro level hypothesis. First, we did not consider the payments that happen after the end of the final development period in a loss triangle, i.e.  $(J = I - 1)$ , for each claim. In order words we did not calculate a reserve for payments that happen after time  $t = i_\ell + (I - 1), \forall \ell$ . This hypothesis is often used for collective approaches based on run-off triangles. Second, we also did not consider any payment after the first closure date of every claim, making the data set consistent with the no reopening hypothesis explained in Section 2. Moreover, these hypotheses were assumed for our individual model and the comparative collective models, in order to have a fair comparison between the results obtained.

For the 57.593 claims, we only consider payments for the Accident Benefits (AB) coverages, i.e. no-fault benefits for accident where the insured or a third party were hurt or killed in a car accident. Furthermore, we have micro-level information regarding each claim, which was used in the three-component model, in the form of categorical static covariates. Table 1 contains a summary of these variables, and Figures 4- 9 contain the percentages of each group among all claims.

Table 1: Categorical variables description

Variable	Label	Number of levels
Gender	Gender of the injured/killed	3
Region	Geographical region where the accident occurred	3
Type of loss	Kind of AB claim	5
Vehicle age	Age of the vehicle, in years, when the accident occurred	6
Injured age	Age of the injured/killed, in years, when the accident occurred	7
Reporting delay	Delay calculated in days	7
Initial Reserve	Reserve at report date	5

All the covariates are static, and some considerations must be explained. First, regarding the type of loss, in some situations a single accident may cause different kinds of losses, therefore, some claims are dependant because they originate from the same casualty, even though most of the covariates could be different. The dependence of related claims introduces an interesting, yet complex, additional problem within the framework we developed in this paper, which will be better explored in a future project. Consequently, for this analysis we assumed independence between claims. In addition, for some claims, some covariates could be missing (NA). We decided to keep these observations as additional categories because the number of claims with at least one unknown value is significant.

Besides the individual micro-level information for each claim, the data set also contains the exact occurrence and reporting dates, as well as the exact date and the cost of each payment up until December 31, 2017. However, at this date there are still 1.135 open claims and thus, some data is missing (e.g. the total paid amount after this date). In order to provide a full comparison between our model and collective triangle-based approaches, we decided to complete the missing values with a Chain Ladder model that uses the latest information available. With these predicted values we can have an estimation of the total payments in 2018 and 2019, allowing us to have the full development triangle for the portfolio. Table 2 contains the development triangle based on the above mentioned hypothesis. The total observed reserves amount is \$188,520,892.

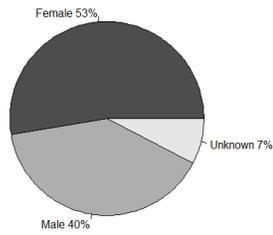


Figure 4: Gender

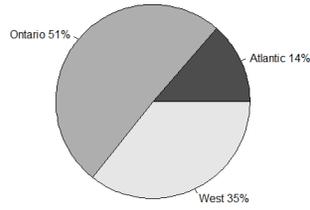


Figure 5: Region

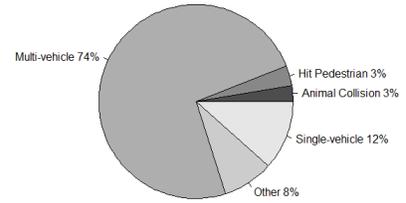


Figure 6: Type of loss

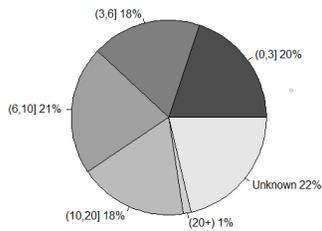


Figure 7: Vehicle age

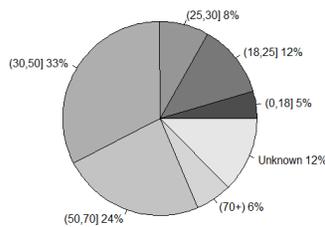


Figure 8: Injured age

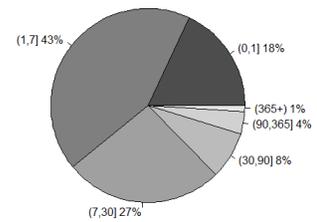


Figure 9: Reporting delay

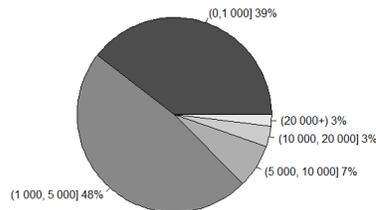


Figure 10: Initial Reserve

Table 2: Full development triangle of the observed total cost

	0	1	2	3	4
1 (2011)	17,749,045	25,449,306	19,499,061	12,186,259	8,914,255
2 (2012)	15,050,987	27,422,359	19,162,367	18,114,766	<b>8,353,106</b>
3 (2013)	16,322,509	33,692,522	23,145,859	<b>18,673,999</b>	<b>12,541,295</b>
4 (2014)	19,451,913	34,563,968	<b>28,254,849</b>	<b>12,360,982</b>	<b>12,180,220</b>
5 (2015)	20,899,092	<b>41,120,193</b>	<b>26,399,552</b>	<b>15,939,656</b>	<b>12,697,041</b>

## 4.2 Fitting the models

The training set consists of all the 56.727 closed and open claims information up to the valuation date. Based on this information, we fit various collective models (see Subsection 4.2.1), our 3-component model (see Subsection 4.2.2), and an individual model based on a Poisson process and suggested by Antonio and Plat [1] (see Subsection 4.2.3).

### 4.2.1 Collective models

We consider four collective models based on the run-off triangle illustrated in Table 2. We consider two classes of approaches: stochastic Chain Ladder model, or Mack's model (see [20] and [19]), and Generalized Linear Model, or GLM, for reserves. For the first, we obtain the predictive distribution through a bootstrap procedure proposed by England and Verrall [8] and based on the quasi-Poisson distribution (Model **Ia**) and the Gamma distribution (Model **Ib**). For the second, we use occurrence and the development periods as covariates. We considered the quasi-Poisson distribution (Model **IIa**) and the Gamma distribution (Model **IIb**) for these models as well. Since these four models are well known in the literature, we do not detail the estimation procedure further and we present the results in Subsection 4.4.

### 4.2.2 The three-component model

We adjust the 3-component model introduced in Section 2 (**Model III**). Regarding the closure delay, we test the Gamma, log-logistic and Weibull distributions and, based on the AIC and the BIC, we chose the Weibull distribution. Thus,

$$\left(T_\ell^{(c)}|\mathbf{c}_\ell\right) \sim \text{Weibull}\left(\lambda, \gamma_\ell\left(\boldsymbol{\beta}^{(t^{(c)})}\right)\right), \text{ for } \ell \in \mathcal{L},$$

where,  $\lambda$  is the *shape parameter*, and  $\gamma_\ell(\cdot)$  is the *scale parameter*. Also,  $\boldsymbol{\beta}^{(t^{(c)})}$  is the parameter vector used to predict the scale parameter.

Next, we fit the frequency component (testing the Poisson, Negative Binomial type I, and Negative Binomial II distributions), considering the following piece-wise development of claims:

$$\mathcal{D} = \{0, 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 5\},$$

where the Negative Binomial type II was chosen, again based on the AIC and the BIC. Thus,

$$(N_{\ell,k}|\mathbf{c}_\ell, \mathcal{D}_\ell) \sim \text{Neg Bin}\left(E_{\ell,k} \cdot \mu_{\ell,k}^{(n)}\left(\boldsymbol{\beta}^{(n)}\right), \sigma\right), \text{ if } E_{\ell,k} > 0, \text{ for } \ell \in \mathcal{L},$$

where  $\mu_{\ell,k}^{(n)}(\cdot)$  and  $\sigma$  are such that,

$$\begin{aligned} \mathbb{E}[N_{\ell,k}|\mathbf{c}_\ell, \mathcal{D}_{\ell,k}] &= E_{\ell,k} \cdot \mu_{\ell,k}^{(n)}\left(\boldsymbol{\beta}^{(n)}\right), \\ \text{Var}[N_{\ell,k}|\mathbf{c}_\ell, \mathcal{D}_{\ell,k}] &= E_{\ell,k} \cdot \mu_{\ell,k}^{(n)}\left(\boldsymbol{\beta}^{(n)}\right) + \sigma\left(E_{\ell,k} \cdot \mu_{\ell,k}^{(n)}\left(\boldsymbol{\beta}^{(n)}\right)\right)^2. \end{aligned}$$

Subsequently, the severity component is fitted with a splice model similar to the one suggested by Laudagé et al. [13]. Regarding the choice of the threshold, we used a 5-fold cross validation procedure to find the value of  $u$  that minimizes the mean square error between the predicted and the observed values of the out-of-sample sum of all payments. Regarding the choice of the distributions, we used a logit model to predict the probability of exceeding the threshold, and we chose the Gamma distribution for both the bulk and the value that exceeds the threshold. It is worth noting that in order to fit the bulk model  $H_{(X_{\ell,k,m})}$ , which is a right-truncated parametric model, we used the `gamlss.tr` package of the statistical software R; for more information about this package we recommend the book by Stasinopoulos et al. [24]. Also, Laudagé et al. [13] mentioned using GLM for the tail distribution  $G_{(X_{\ell,k,m})}$  is generally problematic because extreme values are rare. However in our particular problematic the data set is much larger because we are modeling payments instead of total losses, therefore we believe the data set of payments over the threshold is large enough to circumvent the problem found by Laudagé et al. [13]. For reference, there are only 56.727 claims in our training set, from which 315.527 payments originate. Furthermore, with the optimal threshold,  $u = 5, 433$ , the subset of the training set containing the payments exceeding the threshold contains 5, 987 observations.

Finally, regarding the IBNR claims, the occurrence delay is fitted with a multinomial distribution with 12 outcomes (one for each month of the year). Afterwards, the reporting delay is fitted with the mixture model suggested by Antonio and Plat [1], using 8 degenerate components, i.e.  $D = 7$ . Then, based on 100.000 simulations of both  $T_\ell^{(o)}$  and  $T_\ell^{(r)}$ , the distribution of  $U_\ell^{(r)}$  is obtained. Subsequently, the model of  $L_i^*$ , the number of IBNR claims for each occurrence year ( $i = 1, \dots, 5$ ) is fitted. It is worth noting that our data set did not contain the yearly registered exposures, thus we wrote  $\omega_i = 1$ , for  $i = 1, \dots, 5$ . Afterwards, the closure delay, the frequency and, the severity are fitted with the same distributions used for the RBNS claims, though no information from vectors  $\mathbf{c}_\ell$  was used in the fitting process.

In order to highlight the significance of micro-level information, Table 3 contains the estimated values of the parameters for covariates introduced in Table 1, and the  $p$ -value of their respective  $t$ -tests. We can observe, based on these tests, that most categories are significant in the fitting process. This in turn shows that this underlying information has an impact on the prediction of reserves. Therefore, individual models that can handle this kind of data could be attractive to insurers that have access to it.

Table 3: Estimated values

Variable	Category	Models							
		Duration		Frequency		Severity (bulk)		Severity (prob)	
		$\hat{\beta}$	p-value	$\hat{\beta}$	p-value	$\hat{\beta}$	p-value	$\hat{\beta}$	p-value
Type of loss	Single vehicle	0.17	< 0.01	0.37	< 0.01	0.12	< 0.01	0.32	0.09
	Multi vehicle	0.35	< 0.01	0.08	0.24	0.07	< 0.01	-0.04	0.85
	Hit Pedestrian	0.74	< 0.01	0.32	< 0.01	0.07	< 0.01	0.09	0.64
	Other	0.77	< 0.01	0.14	0.07	0.02	0.31	0.01	0.99
Injured Gender	Male	-0.10	< 0.01	0.15	< 0.01	0.13	< 0.01	0.20	< 0.01
	Unknown	-0.45	< 0.01	0.42	< 0.01	0.34	< 0.01	0.30	0.04
Region	Ontario	0.36	< 0.01	0.77	< 0.01	0.61	< 0.01	1.66	< 0.01
	West	-0.73	< 0.01	0.54	< 0.01	0.43	< 0.01	1.07	< 0.01
Injured Age	(18, 25]	0.01	0.71	0.25	< 0.01	0.12	< 0.01	0.54	< 0.01
	(25, 30]	0.18	< 0.01	0.23	< 0.01	0.16	< 0.01	0.56	< 0.01
	[30, 50]	0.25	< 0.01	0.26	< 0.01	0.14	< 0.01	0.60	< 0.01
	(50, 70]	0.29	< 0.01	0.26	< 0.01	0.10	< 0.01	0.63	< 0.01
	(70, $\infty$ )	0.10	< 0.01	0.25	< 0.01	0.01	0.76	0.89	< 0.01
	Unknown	-0.23	< 0.01	-0.07	0.34	-0.09	< 0.01	0.25	0.10
Vehicle age	(3, 6]	0.02	0.30	0.02	0.51	-0.01	0.87	0.01	0.89
	(6, 10]	0.03	0.06	-0.01	0.57	-0.02	< 0.01	0.03	0.60
	(10, 20]	0.08	< 0.01	0.06	0.01	0.01	0.35	0.10	0.05
	(20, $\infty$ )	0.14	< 0.01	0.23	< 0.01	0.02	0.40	0.32	0.01
	Unknown	0.39	< 0.01	0.05	0.05	0.06	< 0.01	0.29	< 0.01
$t_{\ell}^{(r)}$	(1, 7]	0.08	< 0.01	0.07	< 0.01	0.05	< 0.01	0.18	< 0.01
	(7, 30]	0.07	< 0.01	0.13	< 0.01	0.11	< 0.01	0.25	< 0.01
	(30, 90]	-0.12	< 0.01	0.38	< 0.01	0.16	< 0.01	0.51	< 0.01
	(90, 180]	-0.09	0.01	0.46	< 0.01	0.26	< 0.01	0.84	< 0.01
	(180, 365]	0.07	0.22	0.89	< 0.01	0.38	< 0.01	1.44	< 0.01
	(365, $\infty$ )	0.25	< 0.01	0.82	< 0.01	0.37	< 0.01	1.61	< 0.01
Initial Reserve	(1000, 5000]	0.02	0.1	-0.24	< 0.01	0.04	< 0.01	0.08	0.03
	(5000, 10000]	0.63	< 0.01	-0.02	0.18	0.07	< 0.01	0.22	< 0.01
	(10000, 20000]	0.77	< 0.01	0.05	< 0.01	0.13	< 0.01	0.20	< 0.01
	(20000, $\infty$ )	1.18	< 0.01	0.38	< 0.01	0.17	< 0.01	0.24	< 0.01
Time intervals	(0.25, 0.5]	-	-	0.59	< 0.01	0.17	< 0.01	-0.10	0.13
	(0.5, 0.75]	-	-	0.42	< 0.01	0.21	< 0.01	0.10	0.15
	(0.75, 1]	-	-	0.33	< 0.01	0.25	< 0.01	0.68	< 0.01
	(1, 1.5]	-	-	0.41	< 0.01	0.34	< 0.01	1.64	< 0.01
	(1.5, 2]	-	-	0.46	< 0.01	0.40	< 0.01	1.90	< 0.01
	(2, 2.5]	-	-	0.55	< 0.01	0.44	< 0.01	2.16	< 0.01
	(2.5, 3]	-	-	0.58	< 0.01	0.52	< 0.01	2.45	< 0.01
	(3, 4]	-	-	0.75	< 0.01	0.52	< 0.01	2.69	< 0.01
	(4, 5]	-	-	1.1	0.01	0.57	< 0.01	3.27	< 0.01
Intercept		4.50	< 0.01	5.08	< 0.01	5.34	< 0.01	-7.62	< 0.01
$\ln(1/\lambda)$		0.10	< 0.01	-	-	-	-	-	-

### 4.2.3 Individual model based on a Poisson Process

We have also adjusted the model suggested by Antonio and Plat [1] (**Model IV**) in order to provide a more complete analysis and not to rely solely on collective approaches, which are by definition much simpler. We choose to keep as many similarities as possible to have more fair comparison between the models, while making minor adjustments to fit this model. Specifically, we used the the same time intervals (from  $\mathcal{D}$ ) to delimit the intervals of the events Poisson Process which includes payments, closure and closure with payment. Finally, we used the severity component's distribution for the cost of single payments. However, in Antonio and Plat [1] no methodology is indicated to include covariate information in the events Poisson Process, therefore micro-level information was not included for this step.

Table 4 contains the fitted (from events *before* the valuation date) and Table 5 contains the observed (from events *after* the valuation date) intensity of the Poisson process for each type of event, noted as  $h_p$  for Payments,  $h_{sep}$  for closure, or SEttlement with Payments, and  $h_{se}$  for SEttlement without payment. We notice that  $h_p$  is higher for the observed values compared to the fitted ones across all time intervals. Thus, the fitted  $h_p$  is does not represent accurately the intensity of the number of payments that will occur after the valuation date, and will likely result in an underestimation of the number of payments. This difference can be explained by a discrepancy between the claims considered from the training set (which contains closed and open claims) and the test set (which contains only open and unreported claims). This problem is further emphasised in Table 6, where the predicted number of payments after the valuation for this model was obtained by multiplying the total observed exposure by intervals to the fitted intensity of payments and closures with payments. Table 6 also contains the predicted value of the chosen frequency component model (Section 4.2.2), and the predictions based on Section 2.2, using two simple models (Poisson and Negative Binomial) that only consider  $\mathcal{D}$  as a covariate. For the latter three models, the exposure is also considered known. We notice that although all predictions are lower than the observed value considering covariate information considerably reduces the gap between the predicted and observed values, further indicating that, for this particular data set, considering covariate information is vital for better addressing the discrepancy between the training and the test data sets.

Table 4: Fitted intensity of the event Poisson process

	(0, 0.25]	(0.25, 0.5]	(0.5, 0.75]	(0.75, 1]	(1, 1.5]	(1.5, 2]	(2, 2.5]	(2.5, 3]	(3, 4]	(4, 5]
$\hat{h}_{se}$	2.05	1.33	0.90	0.86	0.89	0.68	0.63	0.63	0.53	0.54
$\hat{h}_p$	8.44	13.38	13.02	12.11	12.72	13.17	12.79	12.73	12.60	14.98
$\hat{h}_{sep}$	0.49	0.76	0.25	0.20	0.26	0.28	0.26	0.25	0.23	0.40

Table 5: Observed intensity of the event Poisson process

	(0, 0.25]	(0.25, 0.5]	(0.5, 0.75]	(0.75, 1]	(1, 1.5]	(1.5, 2]	(2, 2.5]	(2.5, 3]	(3, 4]	(4, 5]
$h_{se}$	2.04	1.26	0.80	0.80	0.78	0.69	0.71	0.66	0.47	0.48
$h_p$	11.12	15.38	15.50	14.08	14.08	14.43	14.71	15.18	16.97	22.12
$h_{sep}$	0.70	0.66	0.22	0.16	0.27	0.26	0.27	0.26	0.35	0.38

Table 6: Predicted and observed number of payments after the valuation date

Model	Covariates	Exposure	Exp. value of the number of payments
Poisson Process	$\mathcal{D}$	known	76,236
simple Poisson	$\mathcal{D}$	known	77,425
simple Neg. Bin.	$\mathcal{D}$	known	78,455
Neg. Bin.	$\mathcal{D}$ and $c$	known	84,306
Observed value			88,405

### 4.3 Goodness of fit analysis

The structure of our approach makes it possible to analyze the impact of micro-level covariates. We compare each of the components of our model (duration, frequency and severity) using micro-level information, with the corresponding model at the macro level, i.e., with a model that only uses the occurrence year ( $i$ ) and the development year ( $j$ ) as covariates. For the frequency component and for the severity component, we also compare a model using only the

individual time intervals ( $\mathcal{D}$ ) as covariates. The main objective is to weight the information provided by the individual time frame provided by  $\mathcal{D}$  and the individual characteristics of each claim separately.

Table 7 contains the AIC and BIC criteria of all three kinds of models for their respective component. We can clearly observe that just by considering the individual time frame for the frequency and severity models, we obtain better results in terms of both criteria. We also notice that introducing individual claim information improves the performance of all models. Furthermore, we performed likelihood ratio tests between restricted and unrestricted models across all components. Yet again, through these tests we want to determine whether including micro-level information, in the form of time intervals or characteristics of claims, is preferable than omitting them in the modelling process. Table 8 contains the results of the aforementioned Likelihood ratio tests, where we can reject the restricted models with an error of at most 0.01 %. Thus, we draw the same conclusion that was drawn from Table 7, i.e. that micro-level information improves the goodness of fit of models from each component.

Table 7: AIC and BIC criteria for models with different covariates across all components

		Macro only	Time intervals only	All micro
AIC	Duration	608,652		591,831
	Frequency	524,531	521,567	515,104
	Severity	4,756,217	4,752,714	4,737,359
BIC	Duration	608,705		592,144
	Frequency	524,630	521,716	515,581
	Severity	5,387,217	5,383,684	5,368,159

Table 8: Likelihood Ratio ( $L.R.$ ) test for models with different covariates across all components

		restricted model covariates	unrestricted model covariates	L.R. test statistic	$p$ -value
Duration model	$i$		all covariates	9,502	< 0.01
Frequency model	$i$ and $j$		$i, j$ and $\mathcal{D}$	2,974	< 0.01
	$i, j$ and $\mathcal{D}$		all covariates	6,528	< 0.01
Severity model	$i$ and $j$		$i, j$ and $\mathcal{D}$	3,829	< 0.01
	$i, j$ and $\mathcal{D}$		all covariates	15,229	< 0.01

The structure of our model also allows us to perform a residual analysis based on cells like in a run-off triangle. We drew inspiration from the residual analysis performed by Avanzi, et al. [3] for their suggested collective model. Thus, we calculate residuals as ratios of observed values to the fitted values of cumulative triangles, and then we obtained heat maps based on these residuals. For our individual model, it is worth noting that the fitted cumulative exposure was obtained through simulation of the duration component while the fitted cumulative number of payments and the fitted cumulative cost were directly obtained from the expected value of each observation. Figure 11 contains the heat maps for all three components, while Figure 12 contains the heat map for the collective Gamma model. Comparatively, the severity component provides better results than the Gamma model for the first three development years, and overall seems to provide a better fit in spite of the worst results observed at the last two development years. Regarding the heat map of the duration component, accident years 2 and 5 have worst values than years 1, 3, 4 and 5 but, overall no extreme value is observed. As for the heat map of the frequency component, residuals are more variable than for the other components but overall the fitted values are close to the observed ones.

Again, taking inspiration from Avanzi, et al. [3], we also plot residuals in terms of accident years, development years and calendar years. This time residuals are calculated as difference between the sum of the observed values and the sum of fitted values for all cells in that year, divided by the sum of fitted values. Figures 13, 14 and 15 plot the residuals of each component. Residuals are close to 0 for all components and all type of years indicating that the goodness-of-fit is overall reasonable.

#### 4.4 Outstanding reserve discussion

After the models were fitted, we proceeded to simulate the distribution of the loss reserves. For the 3-component model, this was accomplished by performing 10,000 times the simulation procedure suggested in Section 3. As for the collective models and the Poisson Process model, more details about the simulation procedures are given by England and Verrall [8], Wüthrich and Merz [28] and Antonio and Plat [1]. For all the fitted models we obtained the distribution

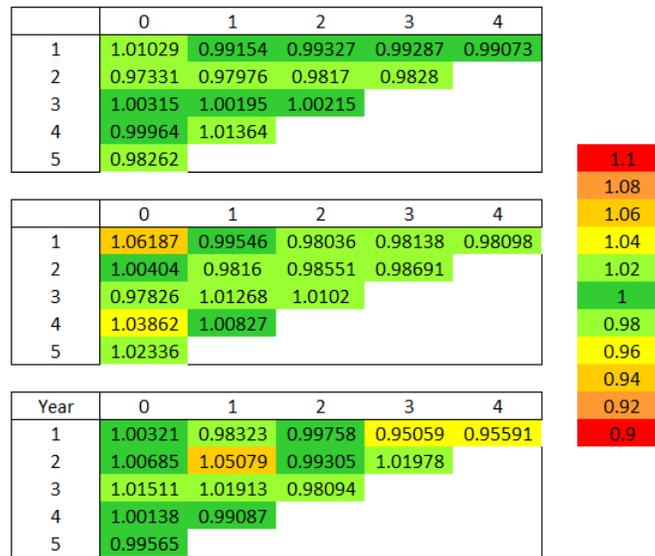


Figure 11: Heat maps of ratios of observed cumulative values to fitted cumulative values for the three component model (in order, from up to down, the exposure, the number of payments and the cost).

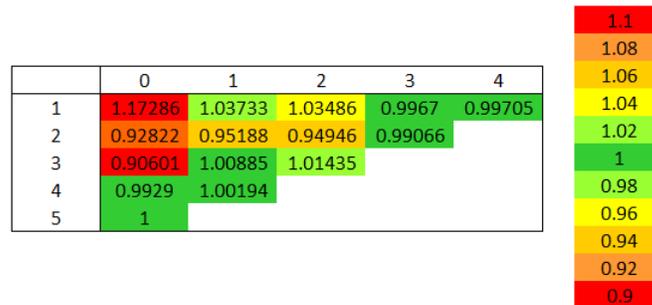


Figure 12: Heat maps of ratios of observed values to fitted values for the gamma collective model.

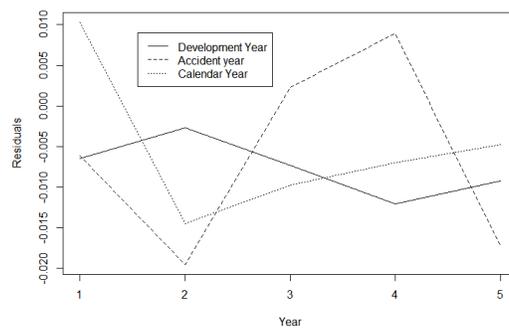


Figure 13: Plot of residuals of the cumulative exposure by accident, development and calendar year.

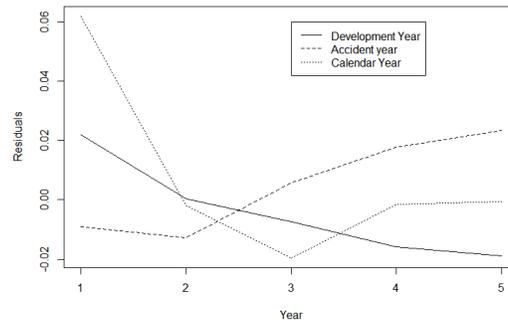


Figure 14: Plot of residuals of the cumulative number of payments by accident, development and calendar year.

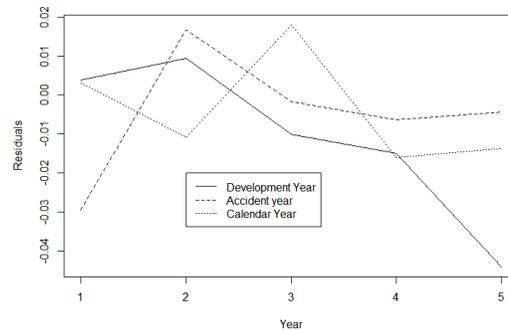


Figure 15: Plot of residuals of the cumulative cost by accident, development and calendar year.

of the loss reserves up to the latest available date (December 31, 2017). These results are depicted in Figure 16 and summarized in Table 9. Then, we obtained these results including the missing data (from January 1, 2018 to December 31, 2019). These results are depicted in Figure 17 and summarized in Table 10.

Table 9: Results of the total reserve predictions until December 31, 2017

	Mean	SD	75% VaR	95% VaR	99% VaR
Mack ODP <b>(Ia)</b>	145,814,301	13,959,646	154,784,259	170,304,205	181,728,307
Mack Gamma <b>(Ib)</b>	146,025,032	13,919,184	155,068,347	169,926,224	180,890,690
GLM Gamma <b>(IIb)</b>	143,604,545	7,969,902	148,973,525	156,696,768	162,534,340
GLM ODP <b>(IIa)</b>	145,171,862	6,565,836	149,603,156	156,112,224	161,073,565
3-component RBNS	145,459,940	3,636,952	147,915,838	151,546,231	154,130,897
3-component IBNR	4,160,285	488,219	4,475,441	5,000,940	5,386,198
3-component total <b>(III)</b>	149,620,225	3,678,054	152,066,762	155,830,382	158,291,786
Poisson Process RBNS	119,353,625	2,832,597	121,325,622	124,050,034	126,042,438
Poisson Process IBNR	3112,515	245,308	3,251,440	3,507,248	3,576,390
Poisson Process total <b>(III)</b>	122,466,140	2,865,310	124,418,939	127,173,916	129,428,485
Observed	147,703,974				

Let us analyze the obtained results. For the results until December 31, 2017, we can observe that the 95 % and the 99 % Values-at-Risk of all models is higher than the observed value, except for the Poisson Process micro-level model. The underestimation was foreseen in Section 4.2.3, where we noticed discrepancy between the observed and fitted payment intensities of the model. The rest of the models provide acceptable results, however the three-component model has the lowest standard-deviation and the lowest 95 % and 99 % Values-at-Risk across all models. This in turn suggests that it

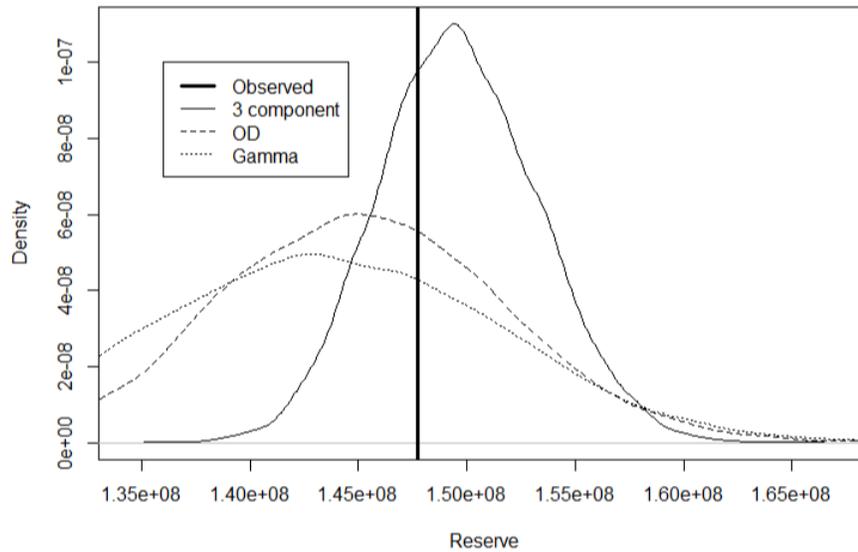


Figure 16: Total reserve distributions until December 31, 2017

Table 10: Results of the total reserve predictions until December 31, 2019

	Mean	SD	75% VaR	95% VaR	99% VaR
Mack ODP ( <b>Ia</b> )	191,065,473	20,320,106	203,802,682	226,475,099	243,132,282
Mack Gamma ( <b>Ib</b> )	190,766,649	20,350,205	203,570,574	226,012,897	244,050,402
GLM Gamma ( <b>IIb</b> )	187,544,147	8,768,697	193,365,173	202,128,221	208,508,242
GLM ODP ( <b>IIa</b> )	189,778,665	7,523,786	194,826,666	202,619,428	207,853,996
3-component RBNS	189,110,346	5,159,041	192,504,713	197,909,943	201,649,986
3-component IBNR	5,875,177	805,072	6,394,193	7,246,948	7,958,991
3-component total ( <b>III</b> )	194,985,523	5,233,211	198,439,003	203,929,434	207,493,721
Poisson Process RBNS	142,493,480	3,426,575	145,323,641	147,299,916	149,601,293
Poisson Process IBNR	5,101,843	677,510	5,579,292	6,162,167	6,446,124
Poisson Process total	147,595,322	3,408,263	149,974,462	152,745,279	154,836,687
Observed*	188,520,892				

\*includes predicted values

outperforms its counterparts by providing a narrower distribution all while providing the insurer a with a loss reserve that covers the observed outstanding payments.

Finally, for the loss reserves until December 31, 2019, the results are narrower. Even though the three-component model still has the lowest standard-deviation among all models, the 95 % and the 99 % Values-at-Risk of the Gamma and the Over-dispersed Poisson have similar results. However, despite how close the Values-at-Risk are, the three-component model has the additional benefit of providing insurers with the loss reserve distribution of individual claims, as we will analyze in the next section. Finally, we can yet again notice that the Poisson Process micro-level model provides underestimated results, and thus is outclassed by our model for this particular data set.

Claims in a portfolio have different risk levels, which can be more easily identified by using their available information. Our 3-component model has the advantage of using this data in the form of explanatory variables to predict outstanding payments for each claim individually, instead of predicting the total reserve directly, as it is done for collective models. Moreover, through the simulation procedure described in Section 3 it is possible to estimate the distribution of each individual reserve at the end of each development year. As an example, we chose four claims from our data set with different characteristics and through 100,000 simulations we estimated the distribution of the cumulative payments. These values are summarized in Table 11.

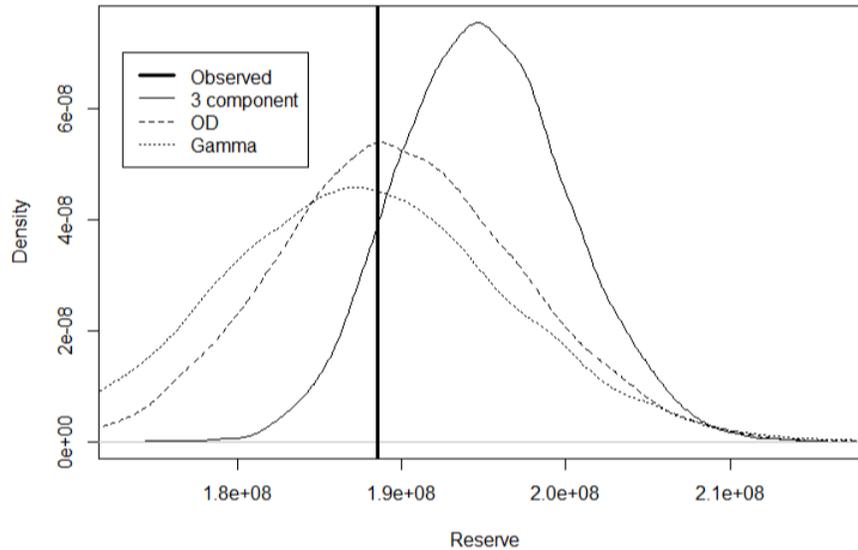


Figure 17: Total reserve distributions until December 31, 2019

Table 11: Outstanding cumulative payments of four claims at the end of each development year ( $j$ )

Claim	$j$	Mean	SD	75% VaR	95% VaR	99% VaR	Observed
1	2	23,108	33,863	32,512	91,900	155,409	53,676
	3	35,415	51,086	49,737	138,745	232,676	93,288
	4	42,023	62,375	57,621	168,511	288,355	n/a
2	2	78,346	96,769	112,934	272,974	432,691	34,184
	3	141,807	161,049	204,967	459,337	720,945	365,183
	4	194,478	217,788	285,711	631,308	959,678	n/a
3	1	6,949	9,411	9,320	25,817	44,102	42,580
	2	10,108	15,819	12,295	41,674	75,956	68,713
	3	11,466	20,401	12,677	48,744	101,116	n/a
	4	11,964	22,567	12,713	50,763	113,601	n/a
4	1	6,527	9,375	8,458	25,076	44,293	7,317
	2	9,668	15,556	11,457	41,169	74,348	13,181
	3	11,051	20,003	11,876	48,416	99,160	n/a
	4	11,534	21,979	11,916	50,452	111,078	n/a

We can see that all the observed values are situated under the 99 % VaR, thus the predictions provide high enough values to meet the required loss reserves under this risk measure. Further, claim 1 and 2 represent higher risks than claim 3 and 4 because the mean, standard deviations, and VaRs have larger values. Therefore, the chosen covariates have an important impact on the prediction of outstanding payments, because the distribution of cumulative outstanding payments changes based on these values. Also, we see how an individual model that uses individual information, such as the one presented in this paper, can provide the insurer with some insight regarding the risk associated with claims. This information could be useful in the pricing process or to enable better reinsurance choices, for example.

## 5 Conclusion

Compared with macro-level models, micro-level models are capable of handling individual covariate information much easily. However, in spite of all their shortcomings, collective type models are still very popular in the industry due to their simplicity and easy to understand structure. In this paper, we suggested a model that can be interpreted in both a macro and micro level structure, while also being able to handle individual claim information. Furthermore, we derived

frequency-severity structure with exposure, which is fairly similar to the one used by Property & Casualty actuaries for price-making predictions, making it even more accessible to the general public. We also put forth a fully parametric approach and proposed some models that can be considered across all components in our real data set analysis.

We managed to show that covariate information is significant in the fitting process across all components and, we even showed that utilizing this information allows insurers to make a more precise prediction of the total reserve compared with conventional models. This indicates that utilizing micro-level information can improve loss reserve predictions, thus making our model appealing to insurers that have access to precise information regarding their claims.

Moreover, this three-component structure opens the door for further research topics, albeit at the cost of interpretability. One may consider the dependence between the frequency and severity as the outstanding claim develops or, even forgoing the piece-wise development triangle structure entirely to predict the total number of payments of each claim directly using more complex offset methods for the exposure, such as splines.

## A Database examples

Table 12 contains the occurrence ( $T_\ell^{(o)}$ ), the reporting ( $T_\ell^{(r)}$ ), and the closure ( $T_\ell^{(c)}$ ) delays. Notice that, for open claims, the closure delay ( $T_\ell^{(c)}$ ) is censored by the valuation date ( $T_\ell^{(e)}$ ). We also have covariate information about the claims, such as the region.

Table 12: Example of a duration training set

$\ell$	$i$	Gender	...	Region	$T_\ell^{(o)}$	$T_\ell^{(r)}$	$T_\ell^{(e)}$	$T_\ell^{(c)}$	Status (1 if closed)
1	3	Female	...	Atlantic	10/365	20/365	-	800/365	1
2	4	Male	...	Ontario	200/365	35/365	-	365/365	1
3	5	Female	...	West	100/365	30/365	235/365	-	0

Table 13 contains the number of payments ( $N_{\ell,k}$ ) based on the development year ( $j$ ) and the time interval vector ( $\mathcal{D} = \{0, 1, 2, 3, 4, 5\}$ ). It also contains their respective exposures  $D_{\ell,k}$  and the same covariate information from Table 12. Note that we can use the intervals from  $\mathcal{D}$  as additional categorical covariates.

Table 13: Example of a frequency training set (with  $\mathcal{D} = \{0, 1, 2, 3, 4, 5\}$ )

$\ell$	$j$	$k$	$i$	Gender	...	$T_\ell^{(r)}$	$N_{\ell,k}$	$D_{\ell,k}$	$E_{\ell,k}$
1	0	0	3	Female	...	20/365	1	335/365	335/365
1	1	0	3	Female	...	20/365	1	35/365	365/365
1	1	1	3	Female	...	20/365	2	330/365	365/365
1	2	1	3	Female	...	20/365	0	35/365	100/365
1	2	2	3	Female	...	20/365	0	65/365	100/365
2	0	0	4	Male	...	35/365	0	130/365	130/365
2	1	0	4	Male	...	35/365	2	235/365	365/365
2	1	1	4	Male	...	35/365	0	130/365	365/365
3	0	0	5	Female	...	30/365	1	235/365	235/365

Table 14 contains the cost of single payments ( $X_{\ell,k,m}$ ) and their respective covariate information. Here again, intervals from  $\mathcal{D}$  can be used as additional covariates.

Table 14: Example of a severity training set (with  $\mathcal{D} = \{0, 1, 2, 3, 4, 5\}$ )

$\ell$	$j$	$k$	$m$	$i$	Gender	...	$T_\ell^{(\tau)}$	$X_{\ell,k,m}$
1	0	0	1	3	Female	...	20/365	\$100
1	1	0	1	3	Female	...	20/365	\$200
1	1	1	1	3	Female	...	20/365	\$550
1	1	1	2	3	Female	...	20/365	\$900
2	1	0	1	4	Male	...	35/365	\$200
2	1	0	2	4	Male	...	35/365	\$300
3	0	0	1	5	Female	...	30/365	\$100

**Acknowledgment:** This research was financially supported by The Co-operators Research Chair in Actuarial Risk Analysis (CARA) .

## References

- [1] Antonio, K., and Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014(7), 649-669.
- [2] Antonio, K., Godecharle, E., and Van Oirbeek, R. (2016). A multi-state approach and flexible payment distributions for micro-level reserving in general insurance. Available at SSRN 2777467.
- [3] Avanzi, B., Taylor, G., Vu, P. A., and Wong, B. (2020). A multivariate evolutionary generalised linear model framework with adaptive estimation for claims reserving. *Insurance: Mathematics and Economics*.
- [4] Baudry, M., and Robert, C. Y. (2019). A machine learning approach for individual claims reserving in insurance. *Applied Stochastic Models in Business and Industry*, 35, 1127-1155.
- [5] Ayuso, M., Guillen, M., and Nielsen, J. P. (2019). Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. *Transportation*, 46(3), 735-752.
- [6] Denuit, M., and Trufin, J. (2017). Beyond the Tweedie reserving model: The collective approach to loss development. *North American Actuarial Journal*, 21(4), 611-619.
- [7] Duval, F., and Pigeon, M. (2019). Individual Loss Reserving Using a Gradient Boosting-Based Approach. *Risks*, 7(3), 79.
- [8] England, P. D., and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3), 443-518.
- [9] Haastруп, S., and Arjas, E. (1996). Claims reserving in continuous time; a nonparametric Bayesian approach. *ASTIN Bulletin*, 26(2), 139-164.
- [10] Huang, J., Qiu, C., and Wu, X. (2015a). Stochastic loss reserving in discrete time: Individual vs. aggregate data models. *Communications in Statistics - Theory and Methods*, 44(10), 2180-2206.
- [11] Huang, J., Qiu, C., Wu, X., and Zhou, X. (2015b). An individual loss reserving model with independent reporting and settlement. *Insurance: Mathematics and Economics*, 64, 232-245.
- [12] Huang, J., Wu, X., and Zhou, X. (2016). Asymptotic behaviors of stochastic reserving: Aggregate versus individual models. *European Journal of Operational Research*, 249(2), 657-666.
- [13] Laudagé, C., Desmettre, S., and Wenzel, J. (2019). Severity modeling of extreme insurance claims for tariffication. *Insurance: Mathematics and Economics*, 88, 77-92.
- [14] Lawless, J. F. (2011). *Statistical models and methods for lifetime data* (Vol. 362). John Wiley & Sons.
- [15] Lopez, O. (2019). A censored copula model for micro-level claim reserving. *Insurance: Mathematics and Economics*, 87(3), 1-14.
- [16] Lopez, O., Milhau, X., and Théron, P. E. (2016). Tree-based censored regression with applications in insurance. *Electronic Journal of Statistics*, 10(2), 2685-2716.
- [17] Lopez, O., Milhau, X., and Théron, P. E. (2016). A tree-based algorithm adapted to microlevel reserving and long development claims. *ASTIN Bulletin*, 49(3), 741-762.
- [18] Lopez, O., and Milhau, (2020). Individual reserving and nonparametric estimation of claim amounts subject to large reporting delays *Scandinavian Actuarial Journal*, in press.

- [19] Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN bulletin*, 23(02), 213-225.
- [20] Mack, T. (1999). The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor. *ASTIN Bulletin*, 29(02), 361-366.
- [21] Ohlsson, E., and Johansson, B. (2010). *Non-life insurance pricing with generalized linear models* (Vol. 2). Berlin: Springer.
- [22] Pigeon, M., Antonio, K., and Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin*, 43(3), 399-428.
- [23] Pigeon, M., Antonio, K., and Denuit, M. (2014). Individual loss reserving using paid–incurred data. *Insurance: Mathematics and Economics*, 58, 121-131.
- [24] Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., and De Bastiani, F. (2017). *Flexible regression and smoothing: using GAMLSS in R*. Chapman and Hall/CRC.
- [25] Taylor, G., McGuire, G., and Sullivan, J. (2008). Individual claim loss reserving conditioned by case estimates. *Annals of Actuarial Science*, 3(1-2), 215-256.
- [26] Verrall, R., Nielsen, J. P., and Jessen, A. H. (2010). Prediction of RBNS and IBNR claims using claim amounts and claim counts. *ASTIN Bulletin*, 40(2), 871-887.
- [27] Verrall, R., and Wüthrich, M. (2016). Understanding reporting delay in general insurance. *Risks*, 4(3), 25.
- [28] Wüthrich, M. V., and Merz, M. (2008). *Stochastic claims reserving methods in insurance*. John Wiley and Sons.
- [29] Wüthrich, M. V. (2018). Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, 2018(6), 465-480.
- [30] Zhao, X. B., Zhou, X., and Wang, J. L. (2009). Semiparametric model for prediction of individual claim loss reserving. *Insurance: Mathematics and Economics*, 45(1), 1-8.
- [31] Zhao, X., and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, 46(2), 290-299.