Modelling payment frequency for loss reserves based on various dynamic claim scores.

A Preprint

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

Jean-Philippe Boucher
Department of Mathematics
Université du Québec à Montréal
Montreal, QC, Canada
boucher.jean-philippe@uqam.ca

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

July 25, 2022

Abstract

By modelling reserves with micro-level models, individual claims information is better preserved and can be more easily handled in the fitting process. Some of the claim information is available immediately at the report date and remains known until the closure of the claim. However, other useful information changes unpredictably as claims develop, for example, the previously observed number of payments. In this paper, we seek to model payment counts in a discrete manner based on past information both in terms of claim characteristics and previous payment counts. We use a dynamic score that weighs the risk of the claim based on previous claim behaviour and that gets updated at the end of each discrete interval. In this paper's model we will also distinguish between the different types of payments. We evaluate our model by fitting it into a data set from a major Canadian insurance company. We will also discuss estimation procedures, make predictions, and compare the results with other models.

Keywords loss reserving · individual models · BMS · GAMLSS

1 Introduction

In order to accurately predict the cost of future liabilities for open claims, practitioners and researchers have suggested several loss reserving models over the years. Over time, these models have changed a lot due to a significant increase in computing capacity, as well as in the quantity (and quality) of available data. While, in the past, models were always part of a collective framework, i.e. built for a data set aggregated by occurrence and development period (*run-off triangle*), today we see a wide selection of models based on varying granularity of the underlying data set, ranging from raw data (micro-level) to aggregated data (macro-level). The actuarial literature on the subject has grown considerably in recent years and we do not wish to do a detailed review here in order not to lengthen this paper unnecessarily. A review of the literature associated with some of the most important and well-known models, such as the Chain-Ladder model (Mack [24, 23]), can be found in Wüthrich and Merz [37], and England and Verrall [11]. As for individual approaches, let us mention, among others, the literature review in Blier et al. [5] (section 4), and Taylor [32]. It should be noted that the quick development of research in the field, partially explained by the increasing use of techniques derived from machine learning, makes any review of the literature necessarily incomplete on the day of its publication.

In this paper, we made a proposition in line with parametric and semi-parametric models. More specifically, we were inspired by models based on Position Dependent Marked-Poisson Process (PDMPP) to predict the exact time of each of the events of a claim, such as payments and settlements. One of the first papers using this type of model is Haastrup and Arjas [12] and was expanded, in 2014, by a more practical implementation proposed by Antonio and Plat [1], in which a more evidence-based methodology was suggested for both IBNR and RBNS reserves using a data set from an insurance company. Antonio et al. [2] further developed this type of model by including a multi-state approach that allowed the model to transition from one state to another as the claim evolved. Other processes that have been considered for the loss reserving literature include the Cox process (in Avanzi et al. [3]) for which dependence was considered through common shock variables and the Hawkes process with time-varying intensities (see Maciak *et al.* [22]). In contrast to these propositions, other models have been suggested. For example, let us mention, Zhao et al. [40] who have developed a semi-parametric model for IBNR claims, and later incorporated copulae into the model. Moreover, a more hierarchical structures was introduced in Yanez and Pigeon [39], where the development of claims was divided into three components: duration of claims, payment frequency, and severity. Then in 2022, another hierarchical approach was suggested in Okine et al. [26] which included the dependency between payments and settlement date.

Because of their granular structure, micro-level models can include more claim information in the modelling process than their aggregated counterparts. This information takes the form of covariates, which are of three types (see Taylor et al. [31]): static, time dynamic, and unpredictable time dynamic. Although time dynamic covariates change as time passes while static covariates remain fixed, both can be predicted with certainty at any point in time. In contrast, unpredictable time dynamic covariates are, as the name suggests, unpredictable. Thus, both static and time dynamic covariates can often be included in models in a more straightforward manner than unpredictable time dynamic covariates. Despite the uncertainty associated with the latter type of covariates, useful claim information can be extracted from them. Specifically, when modelling RBNS claims these covariates are abundant because a portion of the claim development has already been observed. Furthermore, few models that can handle this information have been implemented, namely Antonio et al. [2], which considered including interchangeable states based on payment counts, and Pigeon et al. [28], which made use of incurred losses. In this paper, we propose a new method that can handle an unpredictable time dynamic covariate in a discrete time interval framework.

For each of the open claims in the portfolio, we suggest using observed payments to improve the prediction of the future payments. Past payments are summarized using a score system that is updated at the end of a given discrete time interval with the new available information. Our discrete time scoring model can be implemented into any individual model that can predict payment counts at discrete intervals and that allows for the inclusion of covariates. This latter element is important because the claim score will be considered as a covariate. In particular, the frequency component in Yanez and Pigeon [39] has both characteristics making it a candidate for the inclusion of this more intricate type of covariate.

The idea of calculating a score based on previous observations is not new to the actuarial literature. In fact, the model in this paper draws inspiration from the bonus-malus scoring system (BMS) developed for claim counts. Such a method was developed in Boucher et al. [6], where the authors summarized previous claim counts into a single numerical claim score. This model was further developed in Boucher and Pigeon [7], where the claim score included linear effects. More recently Verschuren [35] proposed a version of the model that incorporates the claim development of different product lines into the score system. Finally, in Boucher and Pigeon [8] a more compact and straightforward scoring system, called a Kappa-N model, was implemented. In this work, we take inspiration from all these sources to introduce a similar dynamic claim scoring system into the micro-level reserving literature.

The method we suggest offers a solution to the inclusion of past claim information in the modelling process, fully taking advantage of a discrete interval structure. Moreover, we suggest to distinguish between different types of payments in the modelling process. This distinction is particularly relevant in loss reserving because payments occur for a variety of reasons (e.g., medical bills, legal fees, etc...), and their distribution could vary. We illustrate this fact in our numerical analysis. To summarize, this paper has the following objectives:

- to implement a dynamic claim scoring system into a discrete interval payment loss reserve model, and to weight the impact of such covariates in the fitting process;
- to develop a model that considers different types of payments, and to analyze their distribution;
- to outperform models that only make use of static and time dynamic covariates.

This paper is structured as follows. In Section 2, we look at the general framework of the model. In Section 2.6, we discuss the estimation procedure followed by Section 3 where we describe the simulation procedure of payment counts. 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 our findings.

2 Statistical framework

In this section we specify the statistical framework of our approach. We define the notation that we use throughout the paper and we present the construction of the dynamic claim score.

2.1 Introductory notation

We show the typical development of a P&C claim in Figure 1. First, accident i occurs and we identify $t_i^{(o)}$, the occurrence delay, i.e., the delay between the beginning of the accident year and the exact accident date. There is an additional delay between the accident date and the reporting date denoted by $t_i^{(r)}$. After the accident has been reported, several payments may be made – illustrated by dots in Figure 1 – before the claim is closed after a final delay $t_i^{(c)}$. At the valuation date, claims can be split into two categories depending on their development. If the claim has not yet been reported we considered it Incurred But Not Reported, or IBNR, and if it has been reported we consider it Reported But Not Settled, or RBNS. Furthermore, for RBNS claims we can compute $t_i^{(e)}$, the delay between the reporting date and the valuation date.

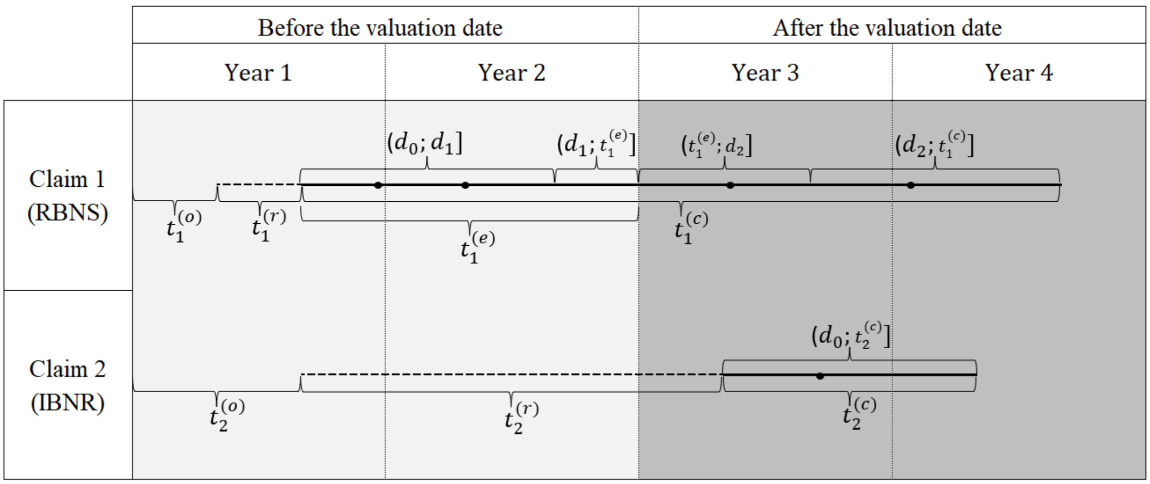

Figure 1: Development of two claims

In a loss-reserving context, we first need to distinguish the status of each of the claims in the portfolio. Let $\mathcal{I} = \mathcal{I}^{(C)} \cup \mathcal{I}^{(O)}$ be the set containing the claims available at the valuation date, where $\mathcal{I}^{(C)}$ and $\mathcal{I}^{(O)}$ are the subsets containing, respectively, the closed and the open (RBNS) claims. Let $\mathcal{I}^*$ be the set containing unreported claims (IBNR), which are unknown at the valuation date.

For each claim $i \in \mathcal{I}$, the observation period, i.e., the period between the reporting date and the closure date (if the claim is closed) or the valuation date (if the claim is open), is denoted by $(0; \tau_i]$, where $\tau_i = \min\{t_i^{(c)}, t_i^{(e)}\}$. Afterwards, the observation period, $(0; \tau_i]$, $i \in \mathcal{I}$, can be divided into time intervals based on vector $\boldsymbol{d} = [d_0, d_1, \ldots, d_K]$, where $d_k < d_{k+1}$, $d_0 = 0$ and $d_K > \max_i\{\tau_i\}$. For the sake of simplicity, we can consider an annual framework, i.e., $\boldsymbol{d} = [0, 1, 2, \ldots]$, but one could also consider a monthly or seasonal division. We suggest to base this decision on the expertise within the company, or on a cross-validation technique.

Furthermore, let $N_{i,k}$ be the number of payments for claim i, $i \in \mathcal{I}$, taking place over the interval $(d_k, d_{k+1}]$, and we define $\mathbf{N}_i = [N_{i,0}, N_{i,1}, \ldots, N_{i,K-1}]$. To each $N_{i,k}$, we associate an exposure measure indicating how long claim i has been open over interval $(d_k, d_{k+1}]$. Thus, let $E_{i,k}$ be the exposure measure of the claim i over the interval $(d_k, d_{k+1}]$:

$$E_{i,k} = \max\{\min\{\tau_i, d_{k+1}\} - d_k, 0\},$$

and $\mathbf{E}_i = [E_{i,0}, E_{i,1}, \ldots, E_{i,K-1}]$.

At the reporting date, micro-level information from a claim becomes available in the form of a vector $\mathbf{X}_i = [X_{i,1}, X_{i,2}, \ldots, X_{i,g}]$ of size g containing available static covariates, e.g., the region where the accident occurred, etc. Note that this vector is not available for unreported claims (IBNR).

We can also identify a vector $\mathbf{Z}_{i,k} = [Z_{i,k,1}, Z_{i,k,2}, \ldots, Z_{i,k,h}]$ of size h containing time dynamic covariates available at each interval $(d_k, d_{k+1}]$. In particular, this vector contains at least one covariate indicating the interval k with which $N_{i,k}$ is associated. Thus, this vector exists for reported claims, as well as for those that have not yet been reported (IBNR). For the latter, we define $\mathbf{Z}^*_{i,k} = [\mathbf{d}_k]$.

2.2 *A priori* distribution of the number of payments

2.2.1 *RBNS* claims

For open claims, $i \in \mathcal{I}^{(O)}$, we aim to predict the number of payments $N_{i,k}$, over the unobserved intervals after the valuation date $t_i^{(e)}$. We use the *a priori* information available at the reporting date (vectors $\mathbf{X}_i$ and $\mathbf{Z}_{i,k}$), as well as the exposure $\mathbf{E}_{i,k}$ before $t_i^{(e)}$. Commonly used approaches in a non-life-insurance context can be considered, such as generalized linear models (GLM). The expected value of $N_{i,k}$, conditionally to $\mathbf{X}_i$, $\mathbf{Z}_{i,k}$ and $E_{i,k}$, is given by

$$\mu_{i,k} = \mathrm{E}\left[N_{i,k}|\mathbf{X}_i, \mathbf{Z}_{i,k}, E_{i,k}\right] = (E_{i,k})\, g^{-1}\left(\mathbf{X}_i\boldsymbol{\beta}' + \mathbf{Z}_{i,k}\boldsymbol{\theta}'\right),$$

where $g^{-1}()$ is the inverse of the link function, and $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are, respectively, the parameter vectors of static and time dynamic covariates.

2.2.2 *IBNR* claims

For claims that have occurred but have not been reported, $i \in \mathcal{I}^*$, we again aim to predict the number of payments $N_{i,k}$, however given that the report date occurs after valuation date, predictions must be made for all the intervals. Instead of having access to the information contained in the vectors $\mathbf{X}_i$ and $\mathbf{Z}_{i,k}$ we only have the information contained in $\mathbf{Z}^*_{i,k}$. Thus, the expected value of $N_{i,k}$, knowing $\mathbf{Z}^*_{i,k}$ and $E_{i,k}$, is given by,

$$\mu^*_{i,k} = \mathrm{E}\left[N_{i,k}|\mathbf{Z}^*_{i,k}, E_{i,k}\right] = (E_{i,k})\, g^{-1}\left(\mathbf{Z}^*_{i,k}\boldsymbol{\theta}^{*\prime}\right),$$

where $g^{-1}()$ is defined as previously, and $\boldsymbol{\theta}^*$ is the parameter vector based on time intervals $(d_k, d_{k+1}]$.

2.3 *A posteriori* distribution of the number of payments

We suggested a method to model frequency payments at different intervals based on information from vectors $\mathbf{X}_i$ and $\mathbf{Z}_{i,k}$, that respectively include static and time dynamic covariates. Now, we can now focus on using information from time dynamic through various measures. Let $\epsilon_{i,k}$ and $\eta_{i,k}$ be, respectively, the cumulative number of payments and exposure of claim i over the interval $(d_0, d_k]$:

$$\epsilon_{i,k} = \sum_{j=0}^{k-1} E_{i,j}, \qquad \eta_{i,k} = \sum_{j=0}^{k-1} N_{i,j}.$$

We include the previously observed frequency in the mean parameter of claim i over interval $(d_k, d_{k+1}]$ in the following way:

$$\mu_{i,k} = \mathrm{E}\left[N_{i,k}|\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathcal{H}_{i,k}\right] = (E_{i,k})\, g^{-1}\left(\mathbf{X}_i\boldsymbol{\beta}' + \mathbf{Z}_{i,k}\boldsymbol{\theta}' + \gamma_1\left(\frac{\eta_{i,k}}{\epsilon_{i,k}}\right)\right),$$

where $\mathcal{H}_{i,k}$ is known development of claim i at time d_k, and γ_1 is the parameter associated with the new component.

Then, we want to adjust the expected value of the frequency by incorporating a covariate that identifies payment-free periods in order to distinguish between claims that have been open for a longer or shorter period of time. Thus, as a claim develops, the frequency of payment-free periods may increase or reduce the expected value. This approach is inspired from the Kappa-N structure suggested by Boucher et al. [6]. Let $\kappa_{i,k}$ represent the total payment-free exposure observed over interval $(d_0, d_k]$, such that,

$$\kappa_{i,k} = \sum_{j=0}^{k-1} E_{i,j}\,\mathbb{1}\left(N_{i,j} = 0\right),$$

where $\mathbb{1}()$ is the indicator function.

We can rewrite the mean parameter by incorporating both elements into a single claim score:

$$\begin{aligned}
\mu_{i,k} = \mathrm{E}\left[N_{i,k}|\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathcal{H}_{i,k}\right] &= (E_{i,k})\, g^{-1}\left(\mathbf{X}_i\boldsymbol{\beta}' + \mathbf{Z}_{i,k}\boldsymbol{\theta}' + \gamma_0(-\kappa_{i,k}) + \gamma_1\left(\frac{\eta_{i,k}}{\epsilon_{i,k}}\right)\right) \\
&= (E_{i,k})\, g^{-1}\left(\mathbf{X}_i\boldsymbol{\beta}' + \mathbf{Z}_{i,k}\boldsymbol{\theta}' + \gamma_0\left(-\kappa_{i,k} + \frac{\gamma_1}{\gamma_0}\left(\frac{\eta_{i,k}}{\epsilon_{i,k}}\right)\right)\right) \\
&= (E_{i,k})\, g^{-1}\left(\mathbf{X}_i\boldsymbol{\beta}' + \mathbf{Z}_{i,k}\boldsymbol{\theta}' + \gamma_0\underbrace{\left(-\kappa_{i,k} + \psi\left(\frac{\eta_{i,k}}{\epsilon_{i,k}}\right)\right)}_{\text{claim score } \ell_{i,k}}\right) \\
&= (E_{i,k})\, g^{-1}\left(\mathbf{X}_i\boldsymbol{\beta}' + \mathbf{Z}_{i,k}\boldsymbol{\theta}' + \gamma_0\ell_{i,k}\right),
\end{aligned}$$

where $k > 0$ and ψ is defined as the *jump-parameter*.

With this structure, we summarize past claim experience into a single claim score that will be updated at the end of each interval. Then, the mean parameter can identify claims that have higher chance of producing payments, and riskier claims. Notice that $\kappa_{i,k}$ is multiplied by -1 in order to better accommodate the negative impact that no-payment periods have on the claim score.

Note that the mean parameter is unbounded. This can be an issue because upper values of the mean parameter can become excessively large as we are including past frequency in our calculations and outliers are not uncommon. Thus, we suggest considering a maximum value for the claim score to avoid an overestimation of future payment counts. Moreover, the decreasing part of the measure, based on $\kappa_{i,k}$, is bounded by the maximal duration of a claim, and is less prone to impacting excessively the prediction of the mean. Thus, the inclusion of a minimal value for the mean parameter is less suitable. Finally, when we look into new claims, no past history has been previously observed, and we can not include the dynamic claim score measure. Thus, by setting the initial value of the claim score to 0, all predictions of the mean parameter are based only on the other covariates available at the report date. We suggest obtaining a claim score such that:

$$\ell_{i,k} = \begin{cases} \min\left\{\left(-\kappa_{i,k} + \psi\left(\frac{\eta_{i,k}}{\epsilon_{i,k}}\right)\right), \ell_{max}\right\}, & \text{for } k = 1, 2, \ldots \\ 0, & \text{for } k = 0. \end{cases} \tag{1}$$

One should note that the claim score for claim i is updated at the end of each interval k based on information up to the end of the previous interval $k-1$. As such, it is possible to identify which claims are more likely to produce payments derived from past information summarized by the value of the claim score at any given time. We could also expand upon the definition of the claim score by letting $\nu_{i,k}$ be the sum of the previously observed frequencies such that:

$$\nu_{i,k} = \sum_{j=0}^{k-1} \frac{N_{i,j}}{E_{i,j}},$$

and we can then reformulate the value of the risk measure:

$$\begin{aligned}
\ell_{i,k} &= \begin{cases} \min\left\{\left(-\kappa_{i,k} + \psi\nu_{i,k}\right), \ell_{max}\right\}, & \text{for } k = 1, 2, \ldots \\ 0, & \text{for } k = 0. \end{cases} \\
&= \begin{cases} \min\left\{\sum_{j=0}^{k-1}\left(-E_{i,j}\mathbb{1}\left(N_{i,j} = 0\right) + \psi\left(\frac{N_{i,j}}{E_{i,j}}\right)\right), \ell_{max}\right\}, & \text{for } k = 1, 2, \ldots \\ 0, & \text{for } k = 0. \end{cases}
\end{aligned} \tag{2}$$

Then we can obtain a recursive structure reminiscent of the Bonus-Malus structure used for claim count modelling:

$$\ell_{i,k} = \begin{cases} \min\left\{\left(\ell_{i,k-1} - E_{i,j}\mathbb{1}\left(N_{i,j}=0\right) + \psi\left(\frac{N_{i,j}}{E_{i,j}}\right)\right), \ell_{max}\right\}, & \text{for } k = 1, 2, \dots \\ 0, & \text{for } k = 0. \end{cases} \tag{3}$$

In particular, model (3) has the added advantage of being able to compute the value of any risk score $\ell_{i,k}$ just by knowing the value of the previous risk score $\ell_{i,k-1}$ and the information from the current interval $(d_{k-1}, d_k]$. Hence, unlike previous propositions (1 and 2), all information observed over the period $(d_0, d_{k-1}]$ is not mandatory to compute $\ell_{i,k}$.

For the remaining part of this paper, we label these three propositions as models **(M1)**, **(M2)** and **(M3)**, respectively for models based on claim score definitions (1), (2) and (3). Further considerations will be addressed in next section using **(M1)** as an example, however similar results can be obtained for models **(M2)** and **(M3)**.

2.4 Payment categories and IBNR specifications for claim-score modelling

Payments can be divided into several categories, e.g., payments related to medical costs, or administrative costs. Suppose there are A different categories of payments. Also, suppose that we want to incorporate past payment count information in the fitting process from different payment categories as the claims develop using a claim score. For a given payment category, we propose using a dynamic claim score model with two parameters $\left(\psi^{(a)}, \ell_{max}^{(a)}\right)$ where the level of risk associated with the category a, $a = 1, \dots, A$, at the beginning of the interval $(d_k, d_{k+1}]$ is given by

$$\ell_{i,k}^{(a)} = \begin{cases} \min\left\{\left(-\kappa_{i,k}^{(a)} + \psi^{(a)}\left(\frac{\eta_{i,k}^{(a)}}{\epsilon_{i,k}}\right)\right), \ell_{max}^{(a)}\right\}, & \text{for } k = 1, 2, \dots \\ 0, & \text{for } k = 0, \end{cases}$$

where $\psi^{(a)}$ is the *jump-parameter* for category a, $\ell_{max}^{(a)}$ is the *maximum claim score* for category a, and

$$\epsilon_{i,k} = \sum_{j=0}^{k-1} E_{i,j}, \qquad \eta_{i,k}^{(a)} = \sum_{j=0}^{k-1} N_{i,j}^{(a)}, \qquad \kappa_{i,k}^{(a)} = \sum_{j=0}^{k-1} E_{i,j}\mathbb{1}\left(N_{i,j-1}^{(a)} = 0\right).$$

Information from claim scores of each category can then be incorporated into the process. Let $\boldsymbol{\ell}_{i,k} = \left[\ell_{i,k}^{(1)}, \ell_{i,k}^{(2)}, \dots, \ell_{i,k}^{(A)}\right]$ be the vector containing the risk levels associated with the different categories of payments. Then, for RBNS claims, we can obtain the expected value of the number of payments from category a,

$$\mu_{i,k}^{(a)} = \mathrm{E}\left[N_{i,k}^{(a)} | \mathbf{X}_i, \mathbf{Z}_{i,k}, E_{i,k}, \boldsymbol{\ell}_{i,k}\right] = (E_{i,k})\, g^{-1}\left(\mathbf{X}_i'\boldsymbol{\beta}^{(a)} + \mathbf{Z}_{i,k}'\boldsymbol{\theta}^{(a)} + \gamma^{(a)}\ell_{i,k}^{(a)}\right),$$

and we obtain the expected value of the number of payments from category a for IBNR claims:

$$\mu_{i,k}^{*(a)} = \mathrm{E}\left[N_{i,k}^{(a)} | \mathbf{Z}_{i,k}^{*}, E_{i,k}, \boldsymbol{\ell}_{i,k}^{*}\right] = (E_{i,k})\, g^{-1}\left(\mathbf{Z}_{i,k}^{*\prime}\boldsymbol{\theta}^{*(a)} + \gamma^{*(a)}\ell_{i,k}^{*(a)}\right).$$

We include the same restriction that we used in the RBNS claims by setting $\ell_{max}^{*(a)}$ as maximal claim score and by including its respective jump-parameter $\psi^{*(a)}$. Notice that, because information from these type of claims is unknown we can only include covariate vector $\mathbf{Z}_{i,k}^{*}$, in addition to the claim scores $\ell_{i,k}^{*(a)}$.

2.5 Distribution of duration of claims

With pricing models, where BMS models are commonly used to predict claim counts, the duration of contracts is known beforehand. However, in a loss reserve context, when we seek to predict outstanding payment counts the full duration of open or unreported claims is unknown, and thus an additional model is required to predict this value in order to obtain the exposure values after the evaluation date. This problem was fully addressed in Yanez and Pigeon [39], where, for claim i, the duration was divided in three parts modelled by three random variables:

- $T_i^{(o)}$ for the occurrence delay;
- $T_i^{(r)}$ for the reporting delay; and
- $T_i^{(c)}$ for the closure delay.

For RBNS claims, the report and occurrence date are known, and the information contained in the covariate vectors $\mathbf{X}_i$ and $\mathbf{Z}_{i,k}$ is also accessible. Hence, it is only necessary to model the closure delay with the added advantage of having access to micro-level information. In Yanez and Pigeon [39], a variety of distributions are considered from the survival literature, such as the Weibull and the Gamma distribution. It is worth noting that the training set used contains right-censored observations because of the valuation date. For more details, refer to the above-mentioned paper.

For IBNR claims however, it is necessary to model all three parts of the duration, and no individual information is available. In Yanez and Pigeon [39], the occurrence delay is addressed with methods that consider seasonal effects. The reporting delay is based on the paper by Antonio and Plat [1], where a mixture of a Weibull distribution with degenerate components was considered to accommodate the observations that only take a few days to complete. The closure delay was addressed similarly to the RBNS claims without considering individual information. Again, refer to Yanez and Pigeon [39] for more details.

2.6 Parameter estimation

The *a priori* distribution parameters $\boldsymbol{\beta}^{(a)}$, $\boldsymbol{\theta}^{(a)}$, and $\gamma^{(a)}$ for each type of payment $a = 1, \ldots, A$ are estimated by maximizing the likelihood function given by

$$\Lambda = \prod_{i \in \mathcal{I}} \prod_{k=0}^{K-1} \prod_{a=1}^{A} p_{\left(N_{i,k}^{(a)} | \mathbf{X}_i, \mathbf{Z}_{i,k}, E_{i,k}, \boldsymbol{\ell}_{i,k}\right)} \left(n_{i,k}^{(a)} | \mathbf{x}_i, \mathbf{z}_{i,k}, e_{i,k}, \boldsymbol{\ell}_{i,k}\right),$$

where $p()$ is the probability mass function of the number of claim payments over each interval given covariates, dynamic claim score, and exposure. We suggest estimating jump-parameter $\psi^{(a)}$ and the maximal values of claim scores $\ell_{max}^{(a)}$ by looking for the values that generate the best likelihood or the best predictions, based on an out-of-sample analysis.

Because we distinguish between IBNR and RBNS reserves, it is also important to comment on the parameter estimation procedure for IBNR claims. One can follow the same procedure already described, but instead of using micro-level covariate vectors, i.e, $\mathbf{X}_i$ and $\mathbf{Z}_i$), we only have access to the covariate vector $\mathbf{Z}^*_{i,k}$. Thus, the likelihood function is given by

$$\Lambda^* = \prod_{i \in \mathcal{I}} \prod_{k=0}^{K-1} \prod_{a=1}^{A} p^*_{\left(N_{i,k}^{(a)} | \mathbf{Z}^*_{i,k}, E_{i,k}, \boldsymbol{\ell}^*_{i,k}\right)} \left(n_{i,k}^{(a)} | \mathbf{z}^*_{i,k}, e_{i,k}, \boldsymbol{\ell}^*_{i,k}\right),$$

where $p^*()$ is the probability mass function. The procedure for estimating jump-parameters, $\psi^{*(a)}$, and the maximum values of claim scores $\ell_{max}^{*(a)}$ remains similar.

3 Simulation procedure

As stated previously, loss reserves are split into two types: IBNR and RBNS. We have established different modelling procedures for both reserves, and in this section, we must establish the two different simulation procedures. We consider model **(M1)** for these algorithms, however similar algorithms can be constructed for models **(M2)** and **(M3)** by adapting the calculation of **step 5c** (5) for IBNR claims and **steps 3(a and b)** (6 and 7) for RBNS claims.

3.1 *IBNR* simulation procedure

The exact number of IBNR claims and their information are unknown at the valuation date. Before we define the simulation procedure for the number of payments, we must perform a few steps. As indicated in Subsection 2.5, for these claims all three delays must be simulated: the occurrence delay, $t_i^{(o)}$, the reporting delay, $t_i^{(r)}$, and the closure delay, $t_i^{(c)}$ (see Figure 1). In this particular context we consider, $u_i^{(r)} = t_i^{(o)} + t_i^{(r)}$, the delay between the beginning of the accident year and the report date of claim i. Moreover, because of the unobserved nature of IBNR claims we must also simulate how many have occurred per accident year. Several propositions have been put forward to predict this value. For instance, in Zhao et al. [40] a semi-parametric methodology was suggested, whereas in Antonio and Plat [1] an approach based on a Poisson process was considered. In this paper we will accommodate the thinned-Poisson model by Pigeon et al. [28] to our simulation procedure, although the aforementioned models can also be considered.

Let $m = 1, \ldots, M$ be the accident year of a given claim, where M is the total number of years considered. We select an approach inspired by the work of Pigeon et al. [28], and we assign a distribution to I_m^*, the number of IBNR claims for each m accident year. By letting m_i be the accident year of claim i, we have:

$$I_m^* \sim \text{Poisson}\left(\theta\omega_m \Pr\Big(U_i^{(r)} \leq M - m_i + 1 | m_i = m\Big)\right), \tag{4}$$

where $\theta\omega_m$ is the occurrence measure, for which ω_m is the total exposure registered for period m. Notice that the occurrence measure is thinned by $\Pr\Big(U_i^{(r)} \leq M - m_i + 1\Big)$. This value represents the probability that the report date occurs before the evaluation date. In order to obtain this value we consider the distribution of the sum of the occurrence delay $T_i^{(o)}$, that is the delay between the beginning of the accident date and the exact accident date, and the report delay $T_i^{(r)}$, the delay between the accident date and the report date. The distributions suggested for each of these two delays are briefly detailed in section 2.5. We can now define the simulation procedure for IBNR payments as follows:

- **Step 1:** Obtain $\widetilde{I}^* = \sum_m \widetilde{I}_m^*$,where $\widetilde{I}_m^*$ is the simulated value of I_m^* for each occurrence period m (see Equation (4)).
- **Step 2:** Obtain $\widetilde{U}_i^{(r)}$, the simulated value of $\Big(U_i^{(r)} | U_i^{(r)} > M - m_i + 1\Big)$, the delay between the beginning of the occurrence period and the exact reporting date of each simulated IBNR claim, where,

  $$\Pr\Big(U_i^{(r)} \leq u | U_i^{(r)} > M - m_i + 1\Big) = \frac{\Pr\Big(M - m_i + 1 < U_i^{(r)} \leq u\Big)}{1 - \Pr\Big(U_i^{(r)} \leq M - m_i + 1\Big)},$$

  for $i = 1, \ldots, \widetilde{I}^*$.
- **Step 3:** Obtain $\widetilde{T}_i^{(c)}$, the simulated value of $\Big(T_i^{(c)} | m_i\Big)$, the closure delay of claim i, for $i = 1, \ldots, \widetilde{I}^*$.
- **Step 4:** Calculate

  $$\widetilde{E}_{i,k} = \begin{cases} d_{i,k+1} - d_{i,k}, & \text{if } d_{i,k+1} \leq \widetilde{T}_i^{(c)} \\ \widetilde{T}_i^{(c)} - d_{i,k}, & \text{if } d_{i,k+1} > \widetilde{T}_i^{(c)} \\ 0, & \text{elsewhere,} \end{cases}$$

  for $k = 0, \ldots, K - 1$ and $i = 1, \ldots, \widetilde{I}^*$.
- **Step 5:** For $i = 1, \ldots, \widetilde{I}^*$, go through each of the following sub-steps.
  - **Step 5a:** Set $k = 0$, the first time interval for which the exposure of claim i is positive and obtain its risk level by setting $\widetilde{\ell}_{i,0}^{*(a)} = 0$ for $a = 1, \ldots, A$.
  - **Step 5b:** Obtain $\widetilde{N}_{i,k}^{(a)}$, a simulated value of $\Big(N_{i,k}^{(a)} | \mathbf{Z}_{i,k}^*, \widetilde{E}_{i,k}, \widetilde{\boldsymbol{\ell}}^*{}_{i,k}\Big)$, for $a = 1, \ldots, A$.
  - **Step 5c:** Calculate the next risk level,

    $$\widetilde{\ell}_{i,k+1}^{*(a)} = \min\left\{ -\sum_{j=1}^{k} \widetilde{E}_{i,j} \mathbb{1}\left(\widetilde{N}_{i,j}^{(a)} = 0\right) + \psi^{*(a)} \frac{\sum_{j=1}^{k} \widetilde{N}_{i,j}^{(a)}}{\sum_{m=1}^{k} \widetilde{E}_{i,j}}, \ell_{max}^{*(a)} \right\} \tag{5}$$

    for $a = 1, \ldots, A$.
  - **Step 5d:**

* If $\widetilde{E}_{i,k+1} > 0$, set $k = k + 1$, the next time interval for which the exposure of claim i is positive. Then return to **Step 5b**.
* If $\widetilde{E}_{i,k+1} = 0$ stop the simulation procedure of claim i.

3.1.1 *RBNS* simulation procedure

With RBNS claims, we have micro-level information in the form of vectors $\mathbf{X}_i$ and $\mathbf{Z}_{i,k}$. Because we are dealing with open claims, a portion of the development has already been observed, so we can use the observed risk level contained in $\boldsymbol{\ell}_{i,k}$ to simulate the unobserved potion of the development. Furthermore, unlike with IBNR claims, the exact number of open claims, $I^{(O)}$, is known beforehand. With these considerations can now describe the simulation procedure,

- **Step 1a:** Set $i = 1$, the first open claim.
- **Step 1b:** Obtain $\widetilde{T}_i^{(c)}$, the simulated value of $\left(T_i^{(c)}|\mathbf{X}_i\right)$, the closure delay of open claim i,
- **Step 1c:** If $\widetilde{T}_i^{(c)} > t_i^{(e)}$, set $i = i + 1$, the next open claim.
- **Step 1.d:**
  - If $i \leq I^{(O)}$, go to **Step 1.b**.
  - If $i = I^{(O)} + 1$, continue.
- **Step 2:** Calculate the exposures after the evaluation date,

$$\widetilde{E}_{i,k} = \begin{cases} d_{i,k+1} - t_i^{(e)}, & k \in \{k : d_{i,k} \leq t_i^{(e)}, d_{i,k+1} \leq \widetilde{T}_i^{(c)}\} \\ \widetilde{T}_i^{(c)} - t_i^{(e)}, & k \in \{k : d_{i,k} \leq t_i^{(e)}, d_{i,k+1} > \widetilde{T}_i^{(c)}\} \\ d_{i,k+1} - d_{i,k}, & k \in \{k : d_{i,k} > t_i^{(e)}, d_{i,k+1} \leq \widetilde{T}_i^{(c)}\} \\ \widetilde{T}_i^{(c)} - d_{i,k}, & k \in \{k : d_{i,k} > t_i^{(e)}, d_{i,k+1} > \widetilde{T}_i^{(c)}\} \\ 0, & \text{elsewhere,} \end{cases}$$

  for $k = 0, \ldots, K - 1$ and $i \in \mathcal{I}^{(O)}$.
- **Step 3:** For each $i \in \mathcal{I}^{(O)}$, go through each of the following sub-steps.
  - **Step 3a:** Set $k = \{k : d_{i,k} \leq t_i^{(e)} < d_{i,k+1}\}$, the first time interval that takes place after the evaluation date and obtain its risk level by calculating

$$\widetilde{\ell}_{i,k}^{(a)} = \begin{cases} \min\left\{-\sum_{j=1}^{k} E_{i,j} \mathbb{1}\left(N_{i,j}^{(a)} = 0\right) + \psi^{(a)} \frac{\sum_{j=1}^{k} N_{i,j}^{(a)}}{\sum_{j=1}^{k} E_{i,j}}, \ell_{max}^{(a)}\right\}, & \text{if } d_{i,k} < t_i^{(e)} \\ \ell_{i,k}^{(a)}, & \text{if } d_{i,k} = t_i^{(e)}, \end{cases} \tag{6}$$

    for $a = 1, \ldots, A$. Note that if a portion of the interval has been observed, i.e., when $d_{i,k} < t_i^{(e)}$, we use the first portion, $(d_{i,k}, t_i^{(e)}]$, to update the risk level of the remainder of the interval. However, if no portion of the interval has been observed, i.e., when $d_{i,k} = t_i^{(e)}$, then the latest information available occurs at the previous time interval $(d_{i,k-1}, d_{i,k}]$, and the risk level is updated based on this information instead.
  - **Step 3b:** Obtain $\widetilde{N}_{i,k}^{(a)}$, a simulated value of $\left(N_{i,k}^{(a)}|\mathbf{X}_i, \mathbf{Z}_{i,k}, \widetilde{E}_{i,k}, \widetilde{\boldsymbol{\ell}}_{i,k}\right)$, for $a = 1, \ldots, A$.
  - **Step 3c:** Calculate the next risk level,

$$\widetilde{\ell}_{i,k+1}^{(a)} = \min\left\{-\sum_{j=1}^{k} \widetilde{E}_{i,j} \mathbb{1}\left(\widetilde{N}_{i,j}^{(a)} = 0\right) + \psi^{(a)} \frac{\sum_{j=1}^{k} \widetilde{N}_{i,j}^{(a)}}{\sum_{j=1}^{k} \widetilde{E}_{i,j}}, \ell_{max}^{(a)}\right\} \tag{7}$$

    for $a = 1, \ldots, A$.
  - **Step 3d:**
    * If $\widetilde{E}_{i,k+1} > 0$, set $k = k + 1$, the next time interval for which the exposure of claim i is positive. Then return to **Step 3b**.
    * If $\widetilde{E}_{i,k+1} = 0$ stop the simulation procedure of claim i.

4 Numerical results

4.1 Data Set

For our numerical analysis, we consider a data set from a Canadian insurance company. The data set contains information from 57,593 claims about Accident Benefits (AB) coverage, i.e., no-fault benefits for accidents where the driver, or a third party, was injured or killed in a car accident. Micro-level information is incorporated in the modelling process in the form of categorical static covariates, which are summarized in Table 1. However, some of the covariates contain missing values (NA). We are able to keep these observations in the process by creating a "missing value" category for each of the covariates. We decided not to remove observations with one or more missing values as this would have deprived us of a large amount of information.

The claims considered in our analysis have occurrence dates from 2011 to 2015, and we have information regarding their development until December 31, 2017. In order to evaluate the performance of our model, we chose to set the valuation date to December 31, 2015, splitting the data set into a training and an evaluation set. Payments before the evaluation date are used to fit the models while payments from that date until December 2017 are used for validation. At the valuation date, there were 48,855 closed claims, 7,872 open claims, and 866 unreported claims in our portfolio.

Table 1: Description of covariates

Covariate	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

Diving more deeply into the number of payments from the data set, which is the focus of this paper, we group payments into three categories:

1. **Medical**: all medical payments;
2. **Disability**: recurrent payments such as Disability Income and Caregiver Disability Income; and
3. **Expenses**: all other types of expenses.

We chose these groups based not only on the nature of the payments, as previously described, but also on their empirical distribution. We present, in Table 2, some descriptive statistics of the claim frequency for each category in the training set, such as the Value-at-Risk, or VaR.

Table 2: Claim frequency descriptive statistics for each category

	Mean	Std. dev.	95% VaR	99% VaR
Medical	3.44	9.86	13.70	41.00
Disability	1.01	5.79	4.00	27.00
Expense	1.11	3.60	7.00	17.00
All	5.57	16.81	24.00	74.00

Finally, we make some simplifying assumptions about the possible dependency that may exist in the data set. First, in some situations, it is possible that a casualty may trigger coverages from different claims, and we acknowledge that this situation can cause dependency between these claims. However, we are not going to address this situation in this study because the proposition made in this paper is more geared towards tackling the problem of including past information from the claims themselves rather than the information from other dependent claims. Consequently, we assumed independence between those claims. Second, we do not consider the possible dependency that may exist between different types of payments from the same claim. We believe that this is a more complex issue that would require a full analysis and allow for the use of innovative methods. We postpone this analysis to a future work where we can better deal with this point.

4.2 Fitting the models

In this subsection, we describe the models we considered in our numerical analysis, as well as the choices made regarding estimating parameters, distributions, etc. The choices and thought process for each step are based on Section 2.6. As previously stated, two models are required: one for IBNR claims and one for RBNS claims. We thoroughly describe the procedure for RBNS claims and make some remarks concerning the procedure for IBNR claims.

First, we consider a time division vector with an even yearly division between each period: $\boldsymbol{d} = \{0, 1, 2, 3, 4, 5\}$. We choose this division because it is the easiest to interpret, since many time divisions in the reserving literature are done year-wise, such as the development periods in a loss triangle. Although, as mentioned before, this model does allow for other time divisions. Second, we select the Poisson distribution and the Negative Binomial distribution for our frequency models.

The Negative Binomial (type II) can be described by its mean and variance:

$$\left(N_{i,k}^{(a)}|\mathbf{X}_{i,k}, \mathbf{Z}_{i,k}, E_{i,k}, \boldsymbol{\ell}_{i,k}\right) \sim \text{Neg Bin II}\left(\mu_{i,k}^{(a)}, \sigma\right), \text{ if } E_{i,k} > 0, \text{ for } i \in \mathcal{I},$$

where $\mu_{i,k}^{(a)}$ and σ are such that,

$$\begin{aligned} \text{E}\left[N_{i,k}^{(a)}|\mathbf{X}_{i,k}, \mathbf{Z}_{i,k}, E_{i,k}, \boldsymbol{\ell}_{i,k}\right] &= \mu_{i,k}^{(a)}, \\ \text{Var}\left[N_{i,k}^{(a)}|\mathbf{X}_{i,k}, \mathbf{Z}_{i,k}, E_{i,k}, \boldsymbol{\ell}_{i,k}\right] &= \mu_{i,k}^{(a)}(\sigma + 1). \end{aligned}$$

Note that there is another version of the Negative Binomial distribution (type I), that will not be considered in this numerical analysis[1]

Finally, for our numerical analysis we estimate parameters $\boldsymbol{\beta}^{(a)}$, $\boldsymbol{\theta}^{(a)}$, $\boldsymbol{\theta}^{*(a)}$, $\psi^{(a)}$, $\psi^{*(a)}$, $\ell_{max}^{(a)}$ and $\ell_{max}^{*(a)}$ by maximizing the likelihood function for each distribution (Poisson and Negative Binomial), each type of payment (medical, disability and expenses) and each method to obtain a claim score (**M1**, **M2** and **M3**). A goodness of fit analysis is performed for the models considered in the next section.

4.3 Goodness-of-fit analysis

In order to streamline the impact of a claim score in the modelling process we begin by selecting the best method for computing the claim score among methods **M1**, **M2** and **M3**. This was achieved by comparing the Akaike information criterion (AIC) and the Bayesian (or Schwarz) information criterion (BIC) between these models. Table 3 and Table 4 contain these results, respectively for RBNS and IBNR claims. In these tables, we notice that models **M1** have consistently the lowest value for both criteria. Henceforth, since for this particular data set model **M1** seems to be the most appropriate, future numerical analysis will be done only for this particular model (estimated values of the parameters are available in Appendix A).

Table 3: Likelihood Information Criteria for RBNS models **M1**,**M2** and **M3**

		AIC			BIC		
Distribution	Payment type	**M1**	**M2**	**M3**	**M1**	**M2**	**M3**
NB	Medical	232,720	232,857	232,752	233,085	233,222	233,117
	Disability	81,099	81,379	81,233	81,464	81,745	81,597
	Expenses	122,865	122,920	122,888	123,230	123,285	123,252
POI	Medical	331,810	332,793	332,342	332,165	333,148	332,698
	Disability	203,585	205,819	204,226	203,940	206,174	204,581
	Expenses	160,369	160,608	160,505	160,725	160,963	160,861

[1] For this distribution the variance is Var $\left[N_{i,k}^{(a)}|\mathbf{X}_{i,k}, \mathbf{Z}_{i,k}, E_{i,k}, \boldsymbol{\ell}_{i,k}\right] = \mu_{i,k}^{(a)} + \left(\mu_{i,k}^{(a)}\right)^2 \sigma$.

Table 4: Likelihood Information Criteria for IBNR models **M1**,**M2** and **M3**

		AIC			BIC		
Distribution	Payment type	**M1**	**M2**	**M3**	**M1**	**M2**	**M3**
NB	Medical	238,606	238,774	238,613	238,707	238,875	238,713
	Disability	82,481	82,951	82,698	82,581	83,056	82,798
	Expenses	130,755	131,289	130,954	130,855	131,390	131,054
POI	Medical	348,062	349,250	348,669	348,153	349,341	348,760
	Disability	217,424	221,672	219,138	217,515	221,764	219,229
	Expenses	185,596	187,235	186,242	185,687	187,326	186,333

Our main goal in this subsection is to assess the performance of the inclusion of the claim score $\boldsymbol{\ell}_{i,k}$ into frequency models, in terms of goodness-of-fit. Hence, we suggest to compare the AIC and BIC of two versions of our models. The first version will include $\boldsymbol{\ell}_{i,k}$ as a covariate and the second version will not. We present these results in Table 5 and Table 6. As shown in these tables, the inclusion of the claim scores provides better results in terms of BIC and AIC across all models and all types of payments.

Table 5: AIC and BIC of RBNS models with and without the claim score

		AIC		BIC	
Model	Payment type	with	without	with	without
NB	Medical	232,720	236,794	233,085	237,149
	Disability	81,099	84,576	81,464	84,931
	Expenses	122,865	123,964	123,230	124,320
POI	Medical	331,810	358,342	332,165	358,688
	Disability	203,585	240,730	203,940	241,077
	Expenses	160,370	164,481	160,725	164,828

Table 6: AIC and BIC of IBNR models with and without the claim score

		AIC		BIC	
Model	Payment type	with	without	with	without
Model	Payment	with	without	with	without
NB	Medical	238,606	243,361	238,707	243,453
	Disability	82,481	86,557	82,581	86,648
	Expenses	130,755	132,828	130,855	132,920
POI	Medical	348,062	380,520	348,153	380,602
	Disability	217,424	263,932	217,515	264,014
	Expenses	185,596	193,395	185,687	193,477

With the same goal in mind, we perform a likelihood ratio test between the models that use it and those that do not. We present results in Table 7. Given low p-values, we can confidently reject all restricted models, i.e., models that do not include a claim score.

Then, we perform t-tests specifically for the parameter of the dynamic claim score, $\gamma^{(a)}$, for each RBNS model. Results are in Table 8. Again, with very low p-values, we can determine that the claim score is significant as a covariate.

Having assessed the increase in terms of goodness of fit, through the AIC, the BIC, the likelihood ratio test and the Student t-test, we can also observe how changes in the dynamic claim score affect the mean of payment counts by plotting its relativity, i.e.,

$$\exp\left(\gamma^{(a)}\ell\right), \text{ for } -5 < \ell \leq \ell_{max}^{(a)},$$

Table 7: Likelihood Ratio (L. R.) test RBNS and IBNR models with and without the dynamic claim score.

Model	Payment	Restricted model covariates	Unrestricted model covariates	L.R. test statistic	p-value
NB	Medical	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \boldsymbol{\ell}_{i,k}$	4075.42	< 0.01
	Disability	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \boldsymbol{\ell}_{i,k}$	4085.04	< 0.01
	Expenses	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \boldsymbol{\ell}_{i,k}$	26,534.17	< 0.01
POI	Medical	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \boldsymbol{\ell}_{i,k}$	1100.91	< 0.01
	Disability	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \boldsymbol{\ell}_{i,k}$	945.97	< 0.01
	Expenses	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \boldsymbol{\ell}_{i,k}$	4113.61	< 0.01
NB	Medical	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \boldsymbol{\ell}^*_{i,k}$	4757.13	< 0.01
	Disability	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \boldsymbol{\ell}^*_{i,k}$	4232.36	< 0.01
	Expenses	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \boldsymbol{\ell}^*_{i,k}$	32,459.69	< 0.01
POI	Medical	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \boldsymbol{\ell}^*_{i,k}$	2075.74	< 0.01
	Disability	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \boldsymbol{\ell}^*_{i,k}$	1593.22	< 0.01
	Expenses	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \boldsymbol{\ell}^*_{i,k}$	7801.23	< 0.01

Table 8: Student's t-test for parameter $\gamma^{(a)}$ for RBNS models with the dynamic claim score.

Distribution	Negative Binomial			Poisson		
Payment category	Medical	Disability	Expenses	Medical	Disability	Expenses
t-value	77.03	62.88	33.74	167.86	156.24	63.74
p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

for the suggested distributions. Figures 2, 3 and 4 depict these results for RBNS payments. We notice that the dynamic claim score has an important impact on the mean parameter, particularly in the extremes. For instance, the lowest increase of the mean parameter for a claim that has reached its maximum score comparatively to a claim with no past history, i.e., having a score equal to zero, is 3.44 times as high (by considering the Negative Binomial for expense payments), whereas the highest comparative increase is 12.79 times as high (by considering the Poisson for disability payments).

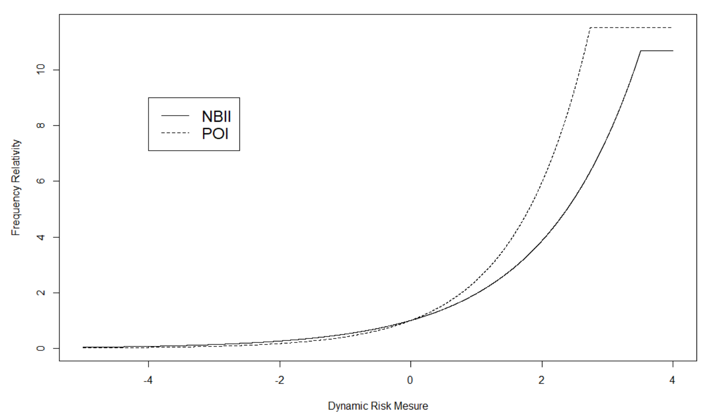

Figure 2: Relativity of the dynamic risk score to the mean of medical RBNS payments

4.4 Simulation analysis

We continue our numerical analysis by simulating the number of outstanding payments for each of the claims. By repeating algorithms described in Section 3 10,000 times, we obtain predicted values for the frequency of payments for

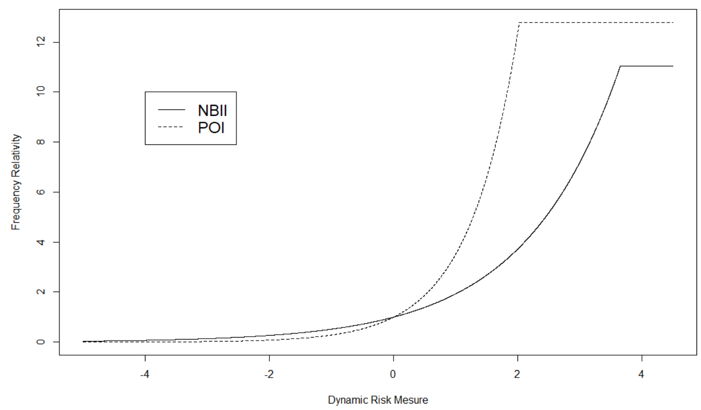

Figure 3: Relativity of the dynamic risk score to the mean of disability RBNS payments

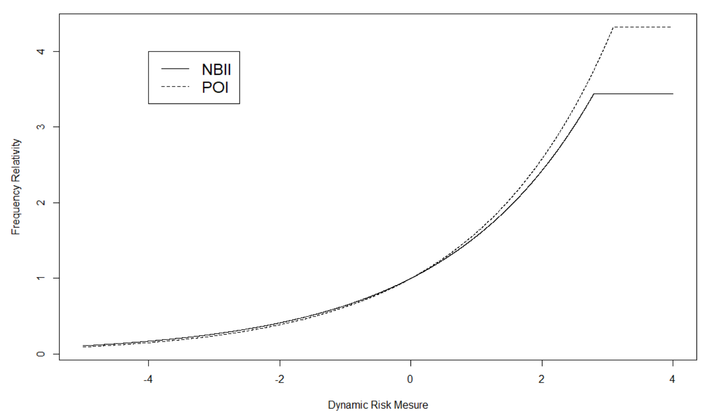

Figure 4: Relativity of the dynamic risk score to the mean of expense RBNS payments

all our models. We summarize our results for IBNR, RBNS, and total reserves in Tables 9 and 10. These tables contain results for models that use the dynamic claim score and those that do not.

Regarding the exposure, we see that it is very well adjusted to the observed value of the RBNS claims: both the mean and the values-at-risk are close to it. Furthermore, when considering the total reserve, we include the IBNR claims, which reduces the accuracy of the exposure predictions, where 99% VaR is slightly under the observed value. We can infer that the model is less accurate when handling IBNR claims. These results can be explained by the lack of information from IBNR claims, in terms of covariates and past history.

Next, we focus on frequency models. For these results we want to compare the results between frequency models that include the dynamic claim score to ones that do not. This analysis is not only done for the total number of payments but also for each type of payment. We will begin by looking at the results from medical payments, which represent the majority of the total. For these payments, the inclusion of the claim score significantly bring the 95% and 99% VaR and mean values closer to observed value, indicating a significant improvement. Next, in terms of RBNS disability payments, the inclusion of the claim score in the Negative Binomial model allows for the 95% and 99% VaR to be over the observed value, a result that does not occur when the claim score is not included. However, we do not see this improvement when considering the Poisson distribution. Finally, regarding the expense payments, both models without and with claim score provide 95% and 99% VaR over the observed value, however the latter models tend to be more conservative with higher results in terms of mean and VaR. Overall, all types of payments are not impacted in the same manner, but their combined value is greatly improved when the claim score is included, without it the Values-at-Risk considerably fall below the observed value.

Table 9: Simulation results for RBNS outstanding payment counts from models with and without claim scores

Claim Score	Dist.	Payment	Mean	Std. dev.	95% VaR	99% VaR	Observed
	Weibull	Exposure	5893	52.43	5979	6015	5889
with	NB	Medical	48,941	837.23	50,309	50,938	51,565
		Disability	21,087	813.54	22,419	23,028	20,601
		Expenses	22,905	425.36	23,599	23,902	16,653
		Total	92,932	1451.76	95,299	96,303	88,819
	POI	Medical	50,749	686.03	51,888	52,384	51,565
		Disability	16,727	433.92	17,430	17,713	20,601
		Expenses	20,607	282.73	21,075	21,259	16,653
		Total	88,084	1125.81	89,945	90,669	88,819
without	NB	Medical	38,426	587.57	39,384	39,801	51,565
		Disability	18,519	625.93	19,563	20,018	20,601
		Expenses	20,820	359.09	21,405	21,688	16,653
		Total	77,765	1088.22	79,554	80,344	88,819
	POI	Medical	42,420	441.76	43,141	43,444	51,565
		Disability	17,464	243.50	17,862	18,028	20,601
		Expenses	18,277	229.26	18,657	18,805	16,653
		Total	78,161	796.56	79,487	79,970	88,819

Table 10: Simulation results for the total outstanding payment counts from models with and without claim scores

Claim Score	Dist.	Payment	Mean	Std. dev.	95% VaR	99% VaR	Observed
	Weibull	Exposure	6275	57.38	6369	6409	6454
with	NB	Medical	51,922	869.71	53,360	53,973	54,986
		Disability	21,885	825.47	23,248	23,843	21,620
		Expenses	24,054	440.34	24,780	25,085	18,080
		Total	97,861	1500.83	100,308	101,291	94,686
	POI	Medical	53,473	706.31	54,641	55,137	54,986
		Disability	17,360	436.99	18,066	18,354	21,620
		Expenses	21,570	290.65	22,051	22,236	18,080
		Total	92,403	1157.50	94,314	95,067	94,686
without	NB	Medical	41,480	629.11	42,501	42,951	54,986
		Disability	19,529	648.38	20,612	21,102	21,620
		Expenses	22,031	375.11	22,647	22,928	18,080
		Total	83,039	1156.66	84,940	85,735	94,686
	POI	Medical	45,353	479.29	46,143	46,481	54,986
		Disability	18,237	250.12	18,645	18,816	21,620
		Expenses	19,296	240.68	19,695	19,855	18,080
		Total	82,886	851.75	84,288	84,843	94,686

After analyzing the frequency models, we can now compare the best performing model (the one that uses the Negative Binomial distribution) to other models in the literature. However, because most models directly predict the total cost of payments rather than payment counts, we choose to compare this value instead. Thus, we are required to add a severity model to our dynamic score frequency model. We test popular distributions such as the Gamma, log-Normal and inverse normal. We find that fitting each type of payment separately and including the claim score as a covariate is satisfactory, and the Gamma distribution was chosen for this numerical analysis. As for the comparative distributions, we chose two collective generalized linear models,based on the quasi-Poisson distribution and the Gamma distribution (for more details see Wüthrich and Merz [37]). We also consider the individual model by Yanez and Pigeon [39], which serves as a comparative baseline for the inclusion of dynamic claim scores. Table 11 contains the results of $10,000$ simulations of each described model, and Figure 5 displays the results.

Lets discuss the results from Table 11 and Figure 5. We notice that all the models yield satisfactory results in terms of the 95 % and the 99 % VaRs as the values are higher than the observed value. The two collective models (Gamma and over-dispersed Poisson) have a mean that is lower than the observed value, but their standard deviation is higher than the individual models. Furthermore, because the 95 % and the 99 % Values-at-Risk of individual models are lower than

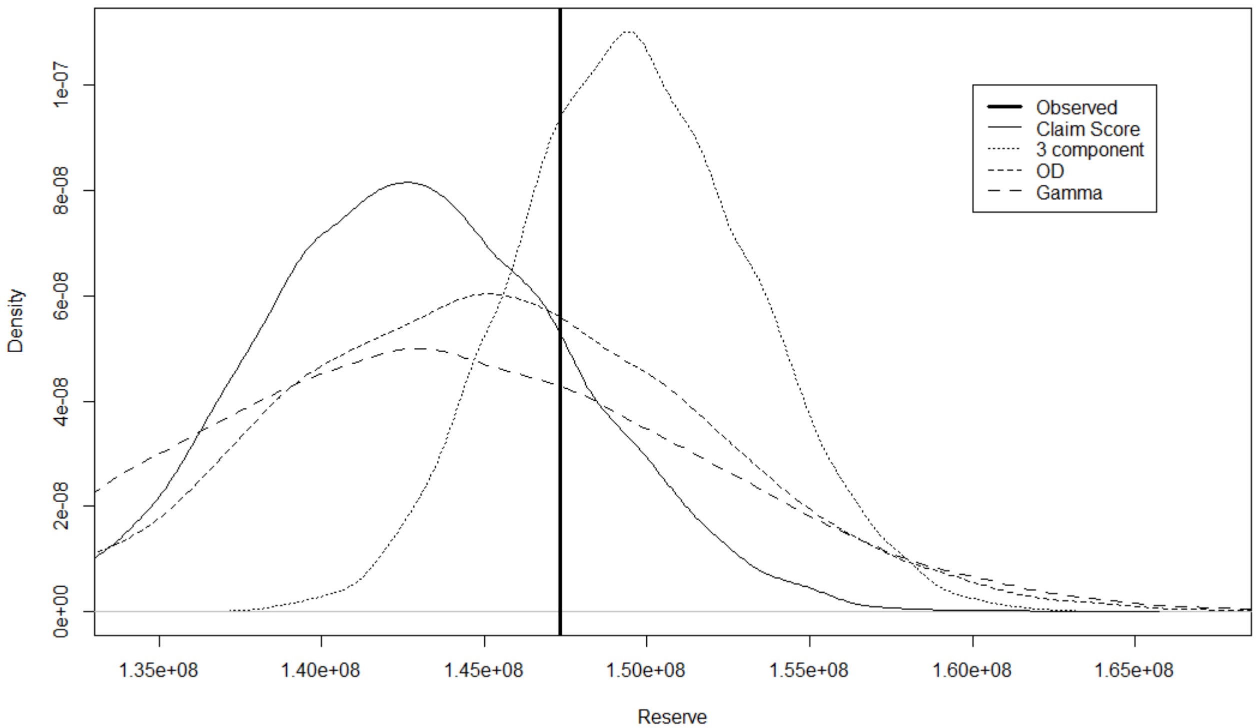

Figure 5: Total reserves for selected models

the collective models but higher than the observed value, the latter approaches are preferable. As for the comparison between both individual approaches, we notice that the mean of the total reserve is lower for the dynamic score model however through a higher standard deviation, the 95 % and the 99 % VaRs become lower than the model that does not make use of the claim score. This further increases the utility of the model by providing values higher than the observed reserve but lower than the other predictions. Again, this shows an overall numerical preference for the model in this paper over the one suggested in Yanez and Pigeon [39].

Table 11: Results of the total reserve predictions

	Mean	Std. dev.	75% VaR	95% VaR	99% VaR
GLM Gamma	143,604,545	7,969,902	148,973,525	156,696,768	162,534,340
GLM ODP	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 total	149,620,225	3,678,054	152,066,762	155,830,382	158,291,786
Dynamic Score RBNS	137,509,168	4,785,344	140,729,680	145,451,829	148,969,071
Dynamic Score total	142,852,107	4,842,791	146,120,166	150,950,931	154,342,708
Observed RBNS	141,830,856				
Observed total	147,308,364				

5 Conclusion

In this paper, we introduced an innovative dynamic claim score to the loss reserve literature. This score allows for the inclusion of past individual claim development in the fitting process of outstanding payment counts. Through an interval-based approach we could feed this score information at the end of each interval and use this updated information for the next interval. We applied this new method to the model by Yanez and Pigeon [39] because of the discrete nature of its payment count modelling and the ease of covariate implementation it allows. However, any model that can predict payment counts at different time development states may incorporate the claim score introduced in this paper. Furthermore, we expanded the scope of payment count modelling by proposing a structure that can consider different payment types.

In our numerical analysis, we applied the aforementioned model to a data set and were able to show that the inclusion of a dynamic claim score improves the performance of traditional count models (such as the Poisson and Negative Binomial models) in terms of goodness-of-fit. Then, we compared the predictions of outstanding payment counts between models that use this new score and models that do not, and we obtained an overall improvement of the predictions. Finally, we showed that our new approach yields better results than collective and individual models available in the literature.

As mentioned before, in this paper we introduced claim scores to the micro-level loss reserving literature. Thus, given pioneering nature of our work, it can branch out into many extensions for a variety of contexts. For example, we could consider a claim score that is based on both the number of payments and their cost, or even a claim score based on the previously observed total cost. We could also consider the correlation between different payment categories of the same claim.

A Appendix

Table 12: Estimated values for the Negative Binomial (type II) Model (RBNS)

Variable	Category	With the claim score			Without the claim score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Intercept		1.66	-0.56	-1.65	1.75	-0.55	-1.75
Type of loss	Single vehicle	0.08	0.44	0.46	0.09	0.53	0.48
	Multi vehicle	0.22	0.07	0.33	0.23	0.06	0.34
	Hit pedestrian	0.32	0.67	0.72	0.40	0.80	0.79
	Other	0.36	0.47	0.46	0.41	0.57	0.50
Injured gender	Male	-0.16	0.12	0.10	-0.18	0.08	0.10
	Unknown	-0.06	0.87	0.94	-0.08	0.93	0.96
Region	Ontario	-0.12	0.30	1.88	-0.19	0.37	1.99
	West	0.50	0.44	0.48	0.45	0.51	0.57
Injured age	$(18, 25]$	0.06	0.54	0.30	0.07	0.63	0.32
	$(25, 30]$	0.20	0.62	0.36	0.20	0.70	0.37
	$[30, 50]$	0.23	0.59	0.38	0.25	0.67	0.40
	$(50, 70]$	0.30	0.60	0.50	0.33	0.67	0.51
	$(70, \infty)$	0.36	0.56	0.65	0.41	0.64	0.70
	Unknown	-0.11	-0.66	-0.32	-0.09	-0.61	-0.32
Vehicle age	$(3, 6]$	0.01	0.07	-0.01	0.00	0.09	-0.01
	$(6, 10]$	0.03	0.12	0.06	0.04	0.15	0.06
	$(10, 20]$	0.05	0.27	0.16	0.05	0.32	0.17
	$(20, \infty)$	-0.00	0.47	0.15	-0.01	0.59	0.15
	Unknown	0.00	0.00	-0.01	-0.03	-0.05	-0.03
$t_\ell^{(r)}$	$(1, 7]$	-0.03	0.14	0.11	-0.03	0.14	0.12
	$(7, 30]$	-0.12	-0.06	0.08	-0.14	-0.13	0.06
	$(30, 90]$	-0.30	-0.35	0.00	-0.36	-0.50	-0.04
	$(90, 180]$	-0.69	-0.57	-0.07	-0.79	-0.85	-0.15
	$(180, 365]$	-0.74	-0.57	-0.14	-0.85	-0.83	-0.28
	$(365, \infty)$	-1.10	-0.74	-0.02	-1.25	-1.07	-0.09
Initial reserve	$(1000, 5000]$	-0.10	-0.25	-0.39	-0.12	-0.27	-0.40
	$(5000, 10000]$	-0.05	0.17	-0.11	-0.08	0.18	-0.12
	$(10000, 20000]$	-0.07	0.26	0.01	-0.11	0.24	0.01
	$(20000, \infty)$	0.00	0.55	0.12	-0.03	0.74	0.13
Time intervals	$(1, 2]$	-0.68	0.17	-0.24	-0.10	0.45	0.03
	$(2, 3]$	-1.07	-0.14	-0.28	-0.18	0.43	0.16
	$(3, 4]$	-1.33	-0.16	-0.35	-0.25	0.37	0.18
	$(4, 5]$	-1.42	0.28	-0.10	-0.21	1.08	0.48
$\sigma^{(a)}$		1.46	2.70	1.08	1.59	2.93	1.13
$\gamma^{(a)}$		0.44	0.61	0.37			
$\psi^{(a)}$		0.14	0.16	0.18			
$\ell_{max}^{(a)}$		4.70	3.77	3.12			

Table 13: Estimated values for the Poisson Model (RBNS)

Variable	Category	With the claim score			Without the claim score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Intercept		1.72	-0.66	-1.78	1.76	-0.79	-1.90
Type of loss	Single vehicle	0.26	0.68	0.47	0.33	0.81	0.49
	Multi vehicle	0.26	0.25	0.29	0.27	0.25	0.28
	Hit pedestrian	0.61	0.92	0.80	0.77	1.10	0.87
	Other	0.44	0.62	0.49	0.50	0.75	0.52
Injured gender	Male	-0.14	0.09	0.09	-0.16	0.06	0.09
	Unknown	0.12	1.17	0.82	0.11	1.27	0.82
Region	Ontario	-0.17	0.76	2.22	-0.24	0.89	2.35
	West	0.23	0.36	0.60	0.16	0.48	0.70
Injured age	$(18, 25]$	0.07	0.17	0.29	0.08	0.18	0.32
	$(25, 30]$	0.20	0.27	0.36	0.21	0.25	0.38
	$[30, 50]$	0.23	0.29	0.37	0.25	0.27	0.40
	$(50, 70]$	0.29	0.34	0.47	0.31	0.29	0.49
	$(70, \infty)$	0.37	0.41	0.64	0.40	0.42	0.67
	Unknown	-0.22	-0.95	-0.38	-0.21	-1.04	-0.37
Vehicle age	$(3, 6]$	0.03	0.15	-0.00	0.03	0.17	-0.00
	$(6, 10]$	0.05	0.18	0.07	0.06	0.24	0.07
	$(10, 20]$	0.07	0.32	0.19	0.09	0.38	0.21
	$(20, \infty)$	0.12	0.53	0.15	0.14	0.73	0.16
	Unknown	-0.03	-0.04	-0.02	-0.07	-0.12	-0.03
$t_\ell^{(r)}$	$(1, 7]$	-0.04	0.06	0.04	-0.04	0.09	0.03
	$(7, 30]$	-0.15	-0.13	-0.04	-0.19	-0.19	-0.07
	$(30, 90]$	-0.33	-0.50	-0.17	-0.42	-0.67	-0.22
	$(90, 180]$	-0.69	-0.71	-0.30	-0.85	-1.07	-0.41
	$(180, 365]$	-0.68	-0.74	-0.35	-0.86	-1.10	-0.49
	$(365, \infty)$	-0.58	-0.74	-0.18	-0.74	-1.07	-0.27
Initial reserve	$(1000, 5000]$	-0.18	-0.24	-0.31	-0.21	-0.28	-0.32
	$(5000, 10000]$	-0.04	0.17	-0.04	-0.07	0.12	-0.06
	$(10000, 20000]$	0.00	0.30	0.10	-0.04	0.22	0.10
	$(20000, \infty)$	0.23	0.70	0.29	0.28	0.90	0.32
Time intervals	$(1, 2]$	-0.81	-0.31	-0.50	0.03	0.45	-0.07
	$(2, 3]$	-1.25	-0.69	-0.66	0.03	0.39	0.01
	$(3, 4]$	-1.57	-0.52	-0.77	-0.11	0.58	0.04
	$(4, 5]$	-1.68	-0.25	-0.70	-0.17	0.97	0.12
$\gamma^{(a)}$		0.65	1.19	0.39			
$\psi^{(a)}$		0.10	0.09	0.20			
$\ell_{max}^{(a)}$		3.54	2.09	3.59			

Table 14: Estimated values for the Negative Binomial (type II) Model (IBNR)

Variable	Category	With the claim score			Without the claim score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Intercept		1.93	0.54	0.68	1.95	0.69	0.72
Time intervals	$(1, 2]$	-0.83	0.35	0.08	-0.23	0.51	0.32
	$(2, 3]$	-0.83	0.42	0.40	-0.34	0.52	0.50
	$(3, 4]$	-0.85	0.60	0.54	-0.42	0.44	0.54
	$(4, 5]$	-0.71	0.97	0.94	-0.31	1.24	0.86
$\sigma^{(a)}$		1.53	2.71	1.33	1.68	2.99	1.41
$\gamma^{(a)}$		0.62	0.74	0.70			
$\psi^{(a)}$		0.12	0.13	0.12			
$\ell_{max}^{(a)}$		3.37	3.57	2.11			

Table 15: Estimated values for the Poisson Model (IBNR)

Variable	Category	With the claim score			Without the claim score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Intercept		1.94	0.67	0.81	1.91	0.64	0.82
Time intervals	$(1, 2]$	-0.90	0.01	-0.30	-0.01	0.62	0.17
	$(2, 3]$	-0.87	-0.02	-0.03	-0.03	0.60	0.30
	$(3, 4]$	-0.91	0.23	0.08	-0.18	0.80	0.33
	$(4, 5]$	-0.87	0.41	0.32	-0.19	1.24	0.39
$\gamma^{(a)}$		0.85	1.32	0.73			
$\psi^{(a)}$		0.09	0.07	0.15			
$\ell_{max}^{(a)}$		2.88	2.23	2.47			

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., Wong, B., and Yang, X. (2021). On the modelling of multivariate counts with Cox processes and dependent shot noise intensities. Insurance: Mathematics and Economics, 99, 9-24.

[4] 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.

[5] Blier-Wong, C., Cossette, H., Lamontagne, L., and Marceau, M. (2021). Machine Learning in P&C Insurance: A review for Pricing and Reserving. *Risks*, 9, 4.

[6] Boucher, J. P., and Inoussa, R. (2014). A posteriori ratemaking with panel data. ASTIN Bulletin: The Journal of the IAA, 44(3), 587-612.

[7] Boucher, J. P., and Pigeon, M. (2019). A Claim Score for Dynamic Claim Counts Modelling, Canadian Institute of Actuaries.

[8] Boucher, J. P. (2021). Bonus-Malus Scale Models: Creating Artificial Past Claims History. To be published in Annals of Actuarial Science.

[9] Baudry, M., and Robert, C. Y. (2017). Non parametric individual claim reserving in insurance. *Preprint*..

[10] Denuit, M., Trufin, J. (2017). Beyond the Tweedie reserving model: The collective approach to loss development. *North American Actuarial Journal*, 21(4), 611-619.

[11] England, P. D., and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3), 443-518.

[12] Haastrup, S., and Arjas, E. (1996). Claims reserving in continuous time; a nonparametric Bayesian approach. *ASTIN Bulletin*, 26(2), 139-164.

[13] 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.

[14] 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.

[15] 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.

[16] Laudagé, C., Desmettre, S., and Wenzel, J. (2019). Severity modeling of extreme insurance claims for tariffication. *Insurance: Mathematics and Economics*, 88, 77-92.

[17] Lopez, O. (2019). A censored copula model for micro-level claim reserving. *Insurance: Mathematics and Economics*, 87(3), 1-14.

[18] Lopez, O., Milhaud, X., and Thérond, P. E. (2016). Tree-based censored regression with applications in insurance. *Electronic Journal of Statistics*, 10(2), 2685-2716.

[19] Lopez, O., Milhaud, X., and Thérond, P. E. (2016). A tree-based algorithm adapted to microlevel reserving and long development claims. *ASTIN Bulletin*, 49(3), 741-762.

[20] Lopez, O., and Milhaud, (2020). Individual reserving and nonparametric estimation of claim amounts subject to large reporting delays *Scandinavian Actuarial Journal*, in press.

[21] Lawless, J. F. (2011). *Statistical models and methods for lifetime data* (Vol. 362). John Wiley & Sons.

[22] Maciak, M., Okhrin, O., and Pešta, M. (2021). Infinitely stochastic micro reserving. Insurance: Mathematics and Economics, 100, 30-58.

[23] Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*: 23(02), 213-225.

[24] 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.

[25] Ohlsson, E., and Johansson, B. (2010). Non-life insurance pricing with generalized linear models (Vol. 2). *Berlin: Springer*.

[26] Okine, A. N. A., Frees, E. W., and Shi, P. (2022). Joint model prediction and application to individual-level loss reserving. ASTIN Bulletin: The Journal of the IAA, 52(1), 91-116.

[27] Pigeon, M., Antonio, K., and Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin*, 43(3), 399-428.

[28] Pigeon, M., Antonio, K., and Denuit, M. (2014). Individual loss reserving using paid–incurred data. *Insurance: Mathematics and Economics*, 58, 121-131.

[29] Pigeon, M., and Duval, F. (2019). Individual Loss Reserving Using a Gradient Boosting-Based Approach. *Risks*, 7(3), 79.

[30] Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., De Bastiani, F. (2017). *Flexible regression and smoothing: using GAMLSS in R*. Chapman and Hall/CRC.

[31] 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.

[32] Taylor, G. (2019). Loss Reserving Models: Granular and Machine Learning Forms. *Risks*, 7, 82.

[33] 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.

[34] Verrall, R., and Wüthrich, M. (2016). Understanding reporting delay in general insurance. *Risks*, 4(3), 25.

[35] Verschuren, R. (2019). Predictive Claim Scores for Dynamic Multi-Product Risk Classification in Insurance. arXiv preprint arXiv:1909.02403.

[36] Wang, Z., Wu, X., and Qiu, C. (2021). The impacts of individual information on loss reserving. *ASTIN Bulletin: The Journal of the IAA*, 51(1), 303-347.

[37] Wüthrich, M. V., and Merz, M. (2008). Stochastic claims reserving methods in insurance (Vol. 435). *John Wiley and Sons*.

[38] Wüthrich, M. V. (2018). Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, 2018(6), 465-480.

[39] Yanez, J. S., and Pigeon, M. (2021). Micro-level parametric duration-frequency-severity modeling for outstanding claim payments. *Insurance: Mathematics and Economics*, 98, 106-119.

[40] 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.

[41] Zhao, X., and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, 46(2), 290-299.