PARAMETRIC OUTSTANDING CLAIM PAYMENT COUNT MODELLING THROUGH A DYNAMIC CLAIM SCORE

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

October 25, 2021

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.

***K*eywords** loss reserving · individual models · BMS · GAMLSS

1 Introduction

Insurance companies must hold loss reserves to cover the cost of future liabilities from open claims. This very important task has been well studied in the loss reserve literature through several propositions. Most of the earlier models are based on loss triangles that summarize payments in terms of development years and accident year. This strand of the literature belongs to the "macro-level" or "collective" family of models. Among them, the most popular is the Chain Ladder model ([20] and [19]), which has been expanded upon by many authors (see Wüthrich and Merz [31] and England and Verrall [8]). These "macro-level" models are appealing to many actuaries because aggregated data in the form of loss triangles allows for concise and summarized data to be used. However, precisely because data is aggregated, these models are often unable to include individual claim information. Thus, some actuaries may prefer more complex approaches that are able to handle this data, if enough individual claim information is readily available. In fact, it was shown in Wang [30] that models that make use of this type of information outperform models that do not.

In contrast with the "macro-level" models, another branch of the literature comprised of what are often called "micro-level" or "individual" models has been suggested in recent years. They seek to model individual claims directly rather than loss triangles and thus are better able to include claim information than their aggregate counterparts. Several authors of individual modelling literature have focused on statistical learning methods. Indeed, interesting propositions have been made by Wülthrich [32], Lopez *et al.* [16, 15], Lopez and Milhaud [17], Lopez [14] by implementing regression trees methods for micro-level loss reserving. Other methods such as the *ExtraTrees* algorithm and the Gradient Boosting procedure have also been implemented (see Baudry and Robert [5] and Pigeon and Duval [24], respectively).

Apart from the statistical learning methods, parametric and semi-parametric methods have also been suggested for micro-level loss reserving. Some authors have developed models using Position Dependent Marked-Poisson Process (PDMPP) to predict the exact time of each of the events of a claim, such as payments and closures. One of the first papers to use this type of model is Haastrup and Arjas [9]. Later, in 2014, a more practical implementation of this approach was 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 developed. In contrast to models based on Haastrup and Arjas [9], other propositions have been made. For instance, Pigeon et al. [22] considered modelling individual development factors, and further expanded the model by including incurred loss information (see Pigeon et al. [23]). Also, Zhao et al. [34] developed a semi-parametric model for IBNR claims, and later incorporated copulae into the model (see Zhao et al. [35]). More recently, a more hierarchical structure was considered in Yanez and Pigeon [33], where the development of claims was divided into three components: duration of claims, payment frequency, and severity.

Because of their granular structure, micro-level models make it easier to include claim information in the modelling process. This information can often be interpreted in the form of covariates, which, based on Taylor et al. [26], are of three types: 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 as time passes. 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. [23], 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 context.

We consider using past observed payment counts to help predict future payment frequency by summarizing the past payment counts using a score system that gets updated at different discrete intervals. At the end of a given interval the score would be updated, and the prediction of the number of payments over the next interval would use the updated score. Our discrete time scoring model can be implemented into any individual model that can predict payment counts at discrete intervals and can incorporate new information at each interval. In Yanez and Pigeon [33], strong emphasis was placed on the ease of implementation of covariate information in each of the hierarchical components (duration, frequency, and severity). Given that, in this work, we aim to incorporate an unpredictable time dynamic covariate into payment count modelling (i.e. frequency), Yanez and Pigeon [33] is an ideal 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. [3], where the authors summarized previous claim counts into a single numerical claim score. This model was further developed in Boucher and Pigeon [4], where the claim score included linear effects. More recently Verschuren [29] proposed a version of the model that incorporates the claim development of different product lines into the score system. Thus, the method we put forward in this paper can be seen as the introduction of a dynamic risk scoring system into the micro-level loss reserve literature.

The method we suggest offers a concrete solution to the inclusion of past claim information in the modelling process, fully taking advantage of a discrete interval structure. We want to make an additional contribution by distinguishing between the 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. To our knowledge, this is the first payment count model that distinguishes between different types of payments, and their difference will be considered in our numerical analysis.

To summarize, our article has the following objectives:

- implement a dynamic risk scoring system into a discrete interval payment loss reserve model, and weight the impact of such covariates in the fitting process;
- develop a model that considers different types of payments, and analyze their distribution;
- 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 3, we discuss the estimation procedure followed by Section 4 where we describe the simulation procedure of payment counts.

In Section 5, we describe the data set used, followed by the numerical results of both our model and other comparative models. Finally, Section 6 contains concluding remarks and mentions further topics that could be explored based on our findings.

2 Statistical framework

In this section we specify the framework of our approach by indicating the notation used and giving the parameters of each model. We then specify the dynamic claim score that allows us to include previous payment counts.

2.1 Introductory notation

The development of a P&C claim is shown 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 declared, several payments may be made – illustrated by dots in the figure – before the claim is closed after a final delay $t_i^{(c)}$. At the valuation date, claims can be separated into several categories according to the information available.

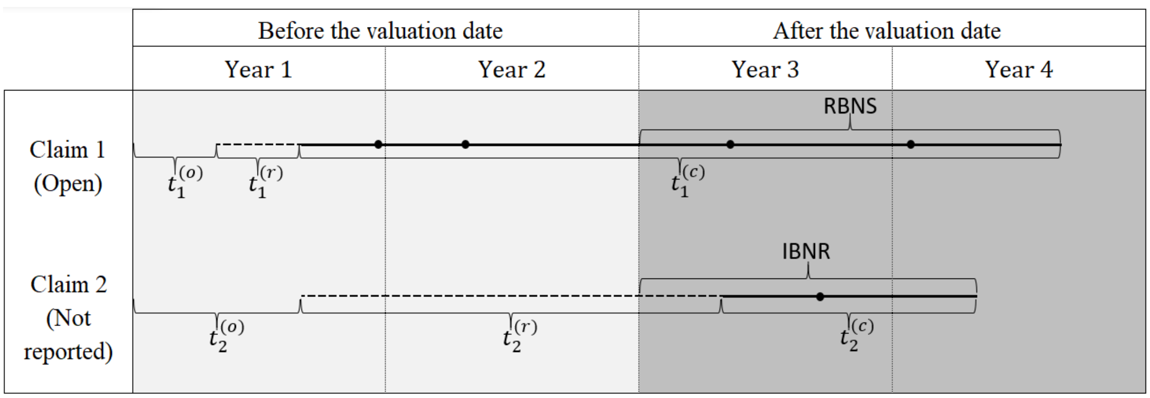

Figure 1: Development of two claims

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

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 $\mathbf{N}_i = [N_{i,0}, N_{i,1}, \ldots, N_{i,K-1}]$. For each $N_{i,k}$ we associate an exposure measure that indicates how long claim i is open in interval $(d_k, d_{k+1}]$. Thus, let $E_{i,k}$ be the exposure measure of the claim i in the interval $(d_k, d_{k+1}]$ so that,

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

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 vector $\mathbf{X}_i$ of size g containing the g available static covariates, which do not vary over time, e.g., the region where the accident occurred. Thus, let $\mathbf{X}_i = [X_{i,1}, X_{i,2}, \ldots, X_{i,g}]$ be the vector containing all static covariates for claim i. Note that this vector is not available for unreported claims (*IBNR*).

We can also identify another vector $\mathbf{Z}_{i,k}$ of size h containing h time dynamic covariates, which vary over time with certainty, e.g., the current age of the insured, available at each interval $(d_k, d_{k+1}]$. Thus, let $\mathbf{Z}_{i,k} = [Z_{i,k,1}, Z_{i,k,2}, \ldots, Z_{i,k,h}]$ be the vector containing all the time dynamic covariates for claim i and interval k. In particular, the covariate that indicates at which interval k, $N_{i,k}$ is situated is considered to be known even if the claim is not declared (*IBNR*). Thus, we can define $\mathbf{Z}^*_{i,k} = \mathbf{d}_k$, as the vector containing the risk factors for IBNR claims.

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 intervals $(d_k, d_{k+1}]$. 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 $\mathbf{E}_{i,k}$. 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 $N_{i,k}$, over the intervals $(d_k, d_{k+1}]$. 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,

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

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

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. Thus, we want to incorporate past payment count information in the fitting process from different payment categories as the claims develop. Thus, for a given payment category we propose using a dynamic risk score model with three parameters $\left(\ell_0, \psi^{(a)}, s^{(a)}\right)$ where the level of risk associated with the category a at the beginning of the interval $(d_k, d_{k+1}]$ is given by

$$L_i^{(a)}(k) = \begin{cases} \min\left\{\max\left\{L_i^{(a)}(k-1) - E_{i,k-1}\mathbb{1}\left(N^{(a)}_{i,k-1} = 0\right) + \psi^{(a)} \frac{N^{(a)}_{i,k-1}}{E_{i,k-1}}, 1\right\}, s^{(a)}\right\}, & \text{for } k = 1, \ldots, K-1 \\ \ell_0, & \text{for } k = 0, \end{cases}$$

where $\psi^{(a)}$ is the *jump parameter*, ℓ_0 is the *initial claim score*, and s is the *maximum claim score*. Hence, risk scores have higher values for claims for which more frequent payments have been observed in the past. This structure can be denoted by $-1/+\psi^{(a)}$, and is directly inspired by bonus-malus pricing systems.

The information from the risk levels of each category can then be incorporated into the process. First, let $\mathbf{L}_i(k) = [L_i^{(1)}(k), L_i^{(2)}(k), \ldots, L_i^{(A)}(k)]$ 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,

$$\mu^{(a)}_{i,k} = \mathrm{E}\left[N^{(a)}_{i,k} | \mathbf{X}_i, \mathbf{Z}_{i,k}, E_{i,k}, \mathbf{L}_i(k)\right] = (E_{i,k})\, g^{-1}\left(\mathbf{X}_i' \boldsymbol{\beta}^{(a)} + \mathbf{Z}_{i,k}' \boldsymbol{\theta}^{(a)} + f^{(a)}\left(L_i^{(a)}(k)\right)\right).$$

Moreover, by setting $f^{(a)}\left(\ell_i^{(a)}(k)\right) = 0$ when $\ell_i^{(a)}(k) = \ell_0$, we decide that for new claims, only covariate vectors $\mathbf{X}_i'$ and $\mathbf{Z}_{i,k}'$ are considered for predictions. Furthermore, past claim behaviour, in the form of claim scores $\mathbf{L}_i(k)$, modifies the baseline mean from new claims. Although it is possible to consider linear functions for functions $f^{(a)}()$, it is also possible to consider functions that better capture changes in the relationship between the risk scores and the number of payments. A piecewise division with break points (or knots) splines COMMENTAIRES DU CORRECTEUR: ** this is unclear to me - there are two plurals in a row with nothing joining them. Do the knots form splines?, is an interesting solution that captures these changes. In this paper we suggest two types of splines: 1) cubic spline smoothers, and 2) penalized cubic basis splines (cubic P-splines). According to Stasinopoulus et al. [25], there are two main differences between these two types of splines. First, P-splines have equidistant knots while cubic smoothing splines have variable values for the knots. Second, smoothness of the fitted function is achieved in two different ways. P-splines penalize the smoothing parameter while cubic splines penalize the second derivative of the function.

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

$$\mu_{i,k}^{*(a)} = \mathrm{E}\left[N_{i,k}^{(a)} | \mathbf{Z}_{i,k}^{*}, E_{i,k}, \mathbf{L}_i(k)\right] = (E_{i,k})\, g^{-1}\left(\mathbf{Z}_{i,k}^{*\prime}\boldsymbol{\theta}^{*(a)} + \sum_{b=1}^{A} f^{*(a)}\left(L_i^{(a)}(k)\right)\right).$$

We include the same restriction that we used in the RBNS claims by setting $f^{(a)}\left(\ell_i^{(a)}(k)\right) = 0$ when $\ell_i^{(a)}(k) = \ell_0$. However, 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 $\mathbf{L}_i(k)$.

In order to better illustrate the dynamic measure, in Figure 2, we provide a graphical toy example of the impact of its incorporation in the modelling process. In this example we show a claim that is riskier than average, where the mean parameter ($\mu_{i,k}$) is lower than the observed number of claims ($N_{i,k}$) between intervals $0 < k \leq 4$. We can modify the mean parameters, $\mu_{i,k}$, into parameters that incorporate past information with a very simple formula:

$$\mu_{i,k}^{*} = \mu_{i,k} \cdot \frac{\sum_{j<k} N_{i,j}}{\sum_{j<k} \mu_{i,j}}.$$

Although the calculation method used in this toy example is very simplistic we can see the value of using previous information in the modelling process, obtaining new mean parameters closer to the observed values.

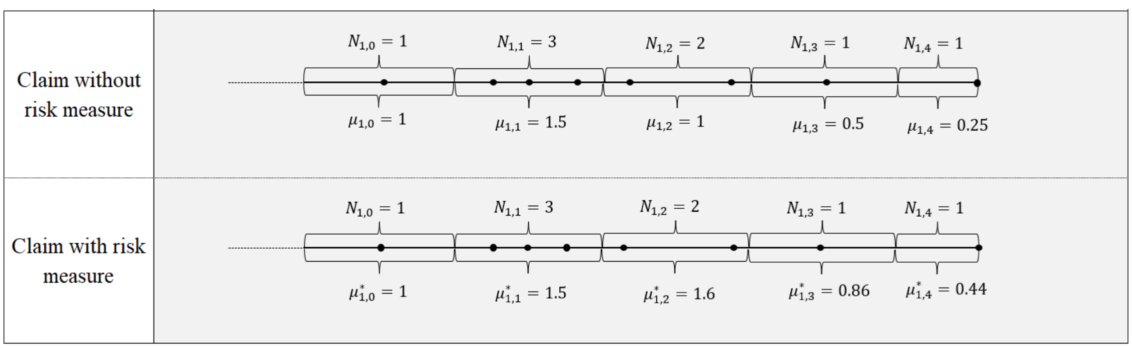

Figure 2: Mean parameters of a claim with and without a dynamic risk measure

2.4 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 [33], 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 [33], 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 [33], 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 [33] for more details.

3 Parameter estimation

We can summarize the steps to follow to estimate the parameters of the model for RBNS claims. First, the actuary must select or estimate the following components:

- the division of time intervals $\boldsymbol{d}$,
- the number of levels $s^{(a)}$ for each type of payment $a = 1, \ldots, A$,
- the static explanatory variables $\mathbf{X}_i$ and their parameters associated with each type of payment $\boldsymbol{\beta}^{(a)}$, $a = 1, \ldots, A$,
- the time dynamic explanatory variables $\mathbf{Z}_i$ and their parameters associated with each type of payment $\boldsymbol{\theta}^{(a)}$, $a = 1, \ldots, A$,
- the type of smoothing functions used for claim scores $f^{(a)}()$, and
- the underlying distributions of the payments.

For this dynamic risk score model, the *a priori* distribution parameters $\boldsymbol{\beta}^{(a)}$, $\boldsymbol{\theta}^{(a)}$, and $f^{(a)}()$ for each type of payment $a = 1, \ldots, A$ are obtained by maximizing the likelihood function given by

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

where $p()$ is the mass function of the number of claim payments at each k interval given their covariates, dynamic risk measure, and exposure. We suggest estimating the jump parameters $\psi^{(a)}$ by looking for the value that generates the best likelihood or the best predictions (based on an out-of-sample analysis).

Because the model of this paper distinguishes 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 covariate vectors that have micro-level information (i.e $\mathbf{X}_i$ and $\mathbf{Z}_i$) we only have access to covariate vector $\mathbf{Z}^*_{i,k}$. Thus, the likelihood function is given by

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

where $p^*()$ is the mass function. Also, a different estimation can be made for the jump parameter, $\psi^{*(a)}$, and it can be estimated, yet again, either by maximizing the likelihood function or by an doing an out-of-sample analysis.

4 Simulation procedure

As stated at the beginning of the paper, 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.

4.1 *IBNR* simulation procedure

The exact number of IBNR claims and their information is unknown at the evaluation date. Before we define the simulation procedure for the number of payments, we must perform a few steps. As indicated in section 2.4, for these claims all three delays must be simulated, that is $t_i^{(o)}$, the occurrence delay, $t_i^{(r)}$, the reporting delay, and $t_i^{(c)}$, the closure delay (see Table 1). To obtain the number of IBNR claims for each calendar year, Yanez and Pigeon [33] suggest using a Poisson distribution thinned by the registered exposure of each period and the distributions of the occurrence and report delays. Employing this methodology, we can define the simulation procedure of IBNR payments as follows:

- **Step 1:** Obtain $\widetilde{I}^*$, the simulated number of claims, (see Yanez and Pigeon [33]).
- **Step 2:** For $i = 1, \dots, \widetilde{I}^*$, follow the procedure from Yanez and Pigeon [33], to simulate the occurrence delay $\widetilde{T}_i^{(o)}$, the reporting delay $\widetilde{T}_i^{(r)}$ and the closure delay $\widetilde{T}_i^{(c)}$.
- **Step 3:** 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 $k = 0, \dots, K-1$ and $i = 1, \dots, \widetilde{I}^*$.
- **Step 4:** For $i = 1, \dots, \widetilde{I}^*$, go through each of the following sub-steps.
  - **Step 4a:** Set $k = 0$, the first time interval for which the exposure of claim i is positive and obtain its risk level by setting $\widetilde{L}_i^{(a)}(0) = \ell_0$ for $a = 1, \dots, A$.
  - **Step 4b:** Obtain $\widetilde{N}_{i,k}^{(a)}$, a simulated value of $\left(N_{i,k}^{(a)} | \mathbf{Z}_{i,k}^*, \widetilde{E}_{i,k}, \widetilde{\mathbf{L}}_i(k)\right)$, for $a = 1, \dots, A$.
  - **Step 4c:** Calculate the next risk level,

$$\widetilde{L}_i^{(a)}(k+1) = \min\left\{\max\left\{\widetilde{L}_i^{(a)}(k) - \widetilde{E}_{i,k}\mathbb{1}\left(\widetilde{N}_{i,k}^{(a)} = 0\right) + \psi^{(a)}\frac{\widetilde{N}_{i,k}^{(a)}}{\widetilde{E}_{i,k}}, 1\right\}, s^{(a)}\right\}$$

    for $a = 1, \dots, A$.
  - **Step 4d:**
    * 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 4b**.
    * If $\widetilde{E}_{i,k+1} = 0$ stop the simulation procedure of claim i.

4.1.1 *RBNS* simulation procedure

With RBNS claims, we have covariate 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 $\mathbf{L}_i(k)$ to simulate the unobserved potion of the development. The whole procedure is described as follows,

- **Step 1:** For each $i \in \mathcal{I}^{(O)}$, follow the procedure from Yanez and Pigeon [33], to obtain $\widetilde{T}_i^{(c)}$, the simulated value of $T_i^{(c)} > t_i^{(e)}$.
- **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 $k = 0, \dots, K-1$ 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 $k = \{k : d_{i,k} \le t_i^{(e)} < d_{i,k+1}\}$, the first time interval that takes place after the evaluation date and obtain its risk level by calculating

$$\widetilde{L}_i^{(a)}(k) = \begin{cases} \min\left\{\max\left\{L_i^{(a)}(k) - E_{i,k}\mathbb{1}\left(N_{i,k}^{(a)} = 0\right) + \psi^{(a)}\frac{N_{i,k}^{(a)}}{E_{i,k}}, 1\right\}, s^{(a)}\right\}, \text{if } d_{i,k} < t_i^{(e)} \\ L_i^{(a)}(k), \text{if } d_{i,k} = t_i^{(e)} \end{cases}$$

    for $a = 1, \ldots, A$. Notice that if a portion of the interval has been observed (that is, when $d_{i,k} < t_i^{(e)}$), then 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 (that is, when $d_{i,k} = t_i^{(e)}$), then the latest information available occurs at the previous time interval $(d_{i,k-1}, d_{i,k}]$, thus, the risk level is updated based on this information instead.
  - **Step 3b:** Obtain $\widetilde{N}_{i,k}^{(a)}$, a simulated value of $\left(N_{i,k}^{(a)}|\mathbf{X}_i, \mathbf{Z}_{i,k}, \widetilde{E}_{i,k}, \widetilde{\mathbf{L}}_i(k)\right)$, for $a = 1, \ldots, A$.
  - **Step 3c:** Calculate the next risk level,

$$\widetilde{L}_i^{(a)}(k+1) = \min\left\{\max\left\{\widetilde{L}_i^{(a)}(k) - \widetilde{E}_{i,k}\mathbb{1}\left(\widetilde{N}_{i,k}^{(a)} = 0\right) + \psi^{(a)}\frac{\widetilde{N}_{i,k}^{(a)}}{\widetilde{E}_{i,k}}, 1\right\}, s^{(a)}\right\}$$

    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.

5 Numerical results

5.1 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 insured, 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 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: Descriptive statistics

	Mean	SD	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 made some simplifying assumptions about the possible dependency that may exist in the data set. Firstly, 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. Secondly, 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.

5.2 Fitting the models

In this section, 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 3. 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.

Let's begin by describing the first steps to take to implement our model. First, we considered a time division vector with an even division between each period:

$$\boldsymbol{d} = \{0, 0.25, 0.5, \ldots, 4.75, 5\}\,.$$

Second, we decide to set the value of $s^{(a)}$ for each type of payment, $a = 1, \ldots, 3$, using the 99$^{\text{th}}$ quantile of the total observed frequency of claim payments from the training set:

$$Pr\left(\frac{\sum_k N_{i,k}^{(a)}}{\sum_k E_{i,k}} > s^{(a)}\right) = 0.01, \text{ where } i \in \mathcal{I}.$$

The values $s^{(a)}$ are listed in Table 3. For the covariates, we consider the ones listed on Table 1, as well as vector $\boldsymbol{d}$. It is worth noting that all the listed covariates are static ($\mathbf{X}_i$), but it is possible to consider time variable covariates ($\mathbf{Z}_{i,k}$), as shown throughout Section 2. Then, for the smoothing functions used for the risk scores, $\mathbf{L}_i(k)$, we consider cubic splines in the same light as suggested by Verschuren [29] (in a pricing context). Finally, we chose three distributions for our models, Poisson, Negative Binomial type I and type II.

The Negative Binomial type I can be described by its mean and variance

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

Table 3: Values of $s^{(a)}$

All	Medical	Disability	Expenses
47.70275	31.46552	7.237277	10.37915

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

$$\begin{aligned}\mathrm{E}\left[N_{i,k}^{(a)}|\mathbf{X}_{i,k},\mathbf{Z}_{i,k},E_{i,k},\mathbf{L}_i(k)\right]&=\mu_{i,k}^{(a)},\\ \mathrm{Var}\left[N_{i,k}^{(a)}|\mathbf{X}_{i,k},\mathbf{Z}_{i,k},E_{i,k},\mathbf{L}_i(k)\right]&=\mu_{i,k}^{(a)}+\sigma\left(\mu_{i,k}^{(a)}\right)^2.\end{aligned}$$

The Negative Binomial type II can be described in a similar manner,

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

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

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

Having specified these features we could then estimate parameters $\boldsymbol{\beta}^{(a)}$, $\boldsymbol{\theta}^{(a)}$, $\boldsymbol{\theta}^{*(a)}$, $\psi^{(a)}$ and $\psi^{*(a)}$ by maximum of likelihood. For each distribution and each type of payment, we also fitted a simpler model that does not use the dynamic risk score. We included these simpler models to compare them to the models that use the dynamic risk score. We list all estimated values in Appendix A.

We perform a goodness-of-fit analysis in the next subsection.

5.3 Goodness-of-fit analysis

Our main goal in this section is to assess the performance of the inclusion of the risk score $\mathbf{L}_i(k)$ into the count models, in terms of goodness-of-fit. First, we compare the likelihood, the Akaike information criterion and the Bayesian information criterion (or Schwarz information criterion) of two versions of our proposed models. The first version will include $\mathbf{L}_i(k)$ as a covariate and the second version will not. We present these results in Table 4 and Table 5. As shown in these tables, the inclusion of the risk scores provides better results in terms of BIC and AIC across all models and all types of payments.

Table 4: AIC and BIC of RBNS models with and without the risk score

		AIC		BIC	
Model	Payment type	with	without	with	without
NBII	Medical	291,287.49	301,585.04	291,691.25	301,960.63
	Disability	99,683.58	109,571.18	100,087.34	109,946.77
	Expenses	157,080.20	160,298.89	157,483.96	160,674.48
NBI	Medical	296,689.78	308,597.22	296,821.24	308,700.50
	Disability	101,039.15	112,260.51	101,170.61	112,363.79
	Expenses	164,760.37	170,638.74	164,891.83	170,742.03
POI	Medical	296,689.78	308,597.22	296,821.24	308,700.50
	Disability	101,039.15	112,260.51	101,170.61	112,363.79
	Expenses	164,760.37	170,638.74	164,891.83	170,742.03

With the same goal in mind (assessing the performance of the inclusion of the risk score covariate) we performed a likelihood ratio test between the models that use it and those that do not. Table 6 contains these results, where we also

Table 5: AIC and BIC of IBNR models with and without the risk score

		AIC		BIC	
Model	Payment	with	without	with	without
NBII	Medical	402,649.49	428,040.19	402,885.65	428,246.83
	Disability	131,999.99	159,272.96	132,236.16	159,479.61
	Expenses	212,020.49	223,972.00	212,256.66	224,178.64
NBI	Medical	404,536.01	425,922.65	404,772.18	426,129.29
	Disability	142,628.00	159,348.47	142,864.17	159,555.11
	Expenses	216,294.50	225,826.63	216,530.66	226,033.27
POI	Medical	473,852.46	563,348.57	474,078.78	563,545.37
	Disability	189,932.19	294,274.79	190,158.51	294,471.60
	Expenses	250,549.49	276,408.74	250,775.81	276,605.54

notice that models that include the risk score provide better results across all distributions and for both IBNR and RBNS claims.

Table 6: Likelihood Ratio (L. R.) test RBNS and IBNR models with and without the dynamic risk score.

Model	Payment	Restricted model covariates	Unrestricted model covariates	L.R. test statistic	p-value
NBII	Medical	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	21,104.47	< 0.01
	Disability	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	25,498.87	< 0.01
	Expenses	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	212,113.68	< 0.01
NBI	Medical	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	19,321.00	< 0.01
	Disability	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	16,654.54	< 0.01
	Expenses	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	211,382.58	< 0.01
POI	Medical	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	75,126.25	< 0.01
	Disability	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	94,570.46	< 0.01
	Expenses	$\mathbf{X}_i, \mathbf{Z}_{i,k}$	$\mathbf{X}_i, \mathbf{Z}_{i,k}, \mathbf{L}_i(k)$	302,934.58	< 0.01
NBII	Medical	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	25,396.70	< 0.01
	Disability	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	27,278.97	< 0.01
	Expenses	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	216,025.69	< 0.01
NBI	Medical	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	21,392.64	< 0.01
	Disability	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	16,726.46	< 0.01
	Expenses	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	209,634.15	< 0.01
POI	Medical	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	89,502.11	< 0.01
	Disability	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	104,348.61	< 0.01
	Expenses	$\mathbf{Z}^*_{i,k}$	$\mathbf{Z}^*_{i,k}, \mathbf{L}_i(k)$	312,805.08	< 0.01

Having assessed the increase in terms of goodness of fit, through AIC, BIC, and likelihood ratio test, we can also observe how changes in the dynamic risk score affect the mean of payment counts by plotting its relativity, that is,

$$\exp\left(f^{(a)}(\ell)\right), \text{ for } 1 \leq \ell \leq s^{(a)}$$

,

for the suggested distributions. Figures 3, 4 and 5 depict these results for RBNS payments. Using cubic splines, we notice that for all the distributions we have a similar trend, an increasing value of the relativity for the lower and higher values of the dynamic risk score and a plateau that indicates a more constant value in between these extremes. Also, the relativity tends to be similar between distributions except for the disability payments from the Negative Binomial type I model, which diverges from the other two distributions. Overall, we can say that high risk scores increase the number of subsequent payments.

Furthermore, we can compare these values with the relativity of the time intervals (defined by vector $\boldsymbol{d}$). As opposed to the values from the dynamic risk score, the relativity associated with the time intervals evolves differently and tends to decrease over time. For medical payments, there is an increase for $d_k \in (0.25, 0.5]$, a sharp decrease for values $d_k \in (0.25, 1]$, followed by a slow decrease from $d_k \in (1, 4]$. For disability and expense payments, there is a sharp decrease from $d_k \in (0.25, 1]$ and more stable values from then on (except for time intervals $d_k \in (4, 5]$ from expense payments, where values slightly decrease and increase at different times in the interval). Overall, we can say later time intervals decrease the number of subsequent payments compared to more recent ones.

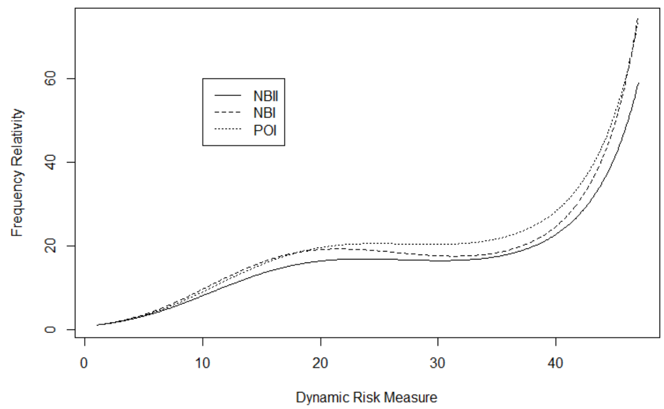

Figure 3: Relativity of the dynamic risk score to the mean of medical RBNS payments

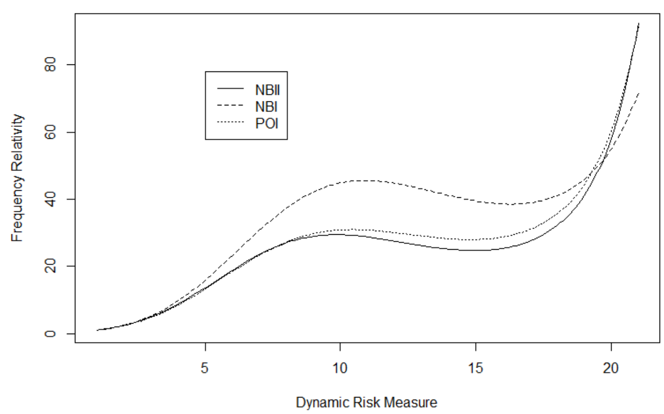

Figure 4: Relativity of the dynamic risk score to the mean of disability RBNS payments

5.4 Simulation analysis

We continue our numerical analysis by simulating the number of outstanding payments for each of the claims. By performing the algorithm described in Section 4 10,000 times, we can obtain the number of payments for the IBNR, RBNS, and total reserves of the various frequency models we fitted. Results are summarized in Tables 7 and 8, which contain, respectively, the values for models that use the dynamic risk score and those that do not.

We begin by analyzing the results regarding the exposure. We see that it is very well adjusted to the observed value for the RBNS claims; both the mean and the values-at-risk are close to it, but this is not the case for IBNR claims. This can be explained by the lack of individual information in IBNR claims. It will be reflected in the results obtained in terms of the frequency, as both the models with and without risk scores predict values that are lower than the observed values.

Next we focus on the frequency models, where the inclusion of the risk score improves the results of RBNS medical payments significantly, obtaining VaR and mean values that are higher than the observed values. Given that medical

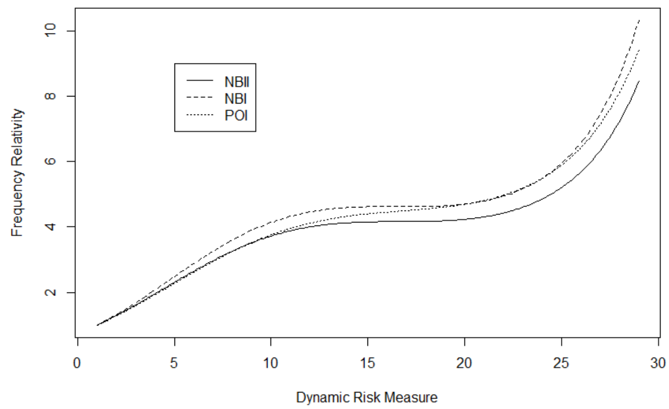

Figure 5: Relativity of the dynamic risk score to the mean of expense RBNS payments

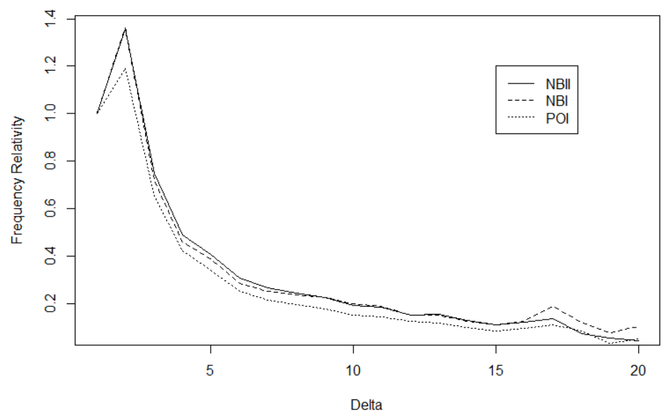

Figure 6: Relativity of the time intervals to the mean of medical RBNS payments

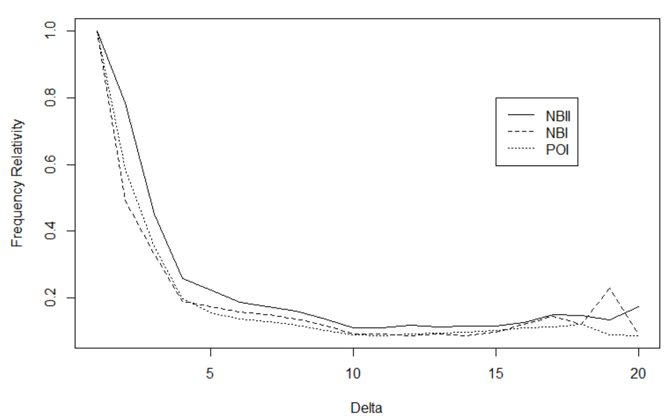

Figure 7: Relativity of the time intervals to the mean of disability RBNS payments

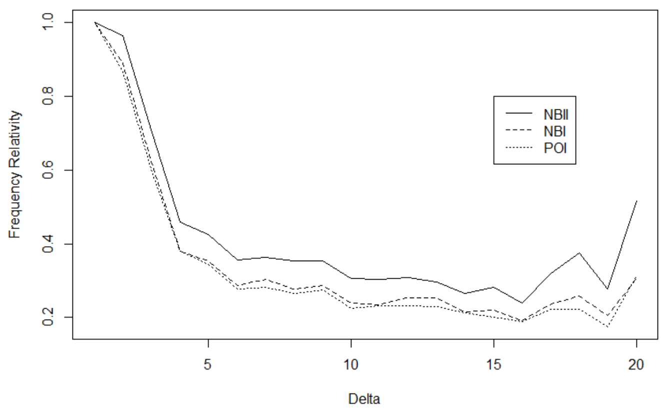

Figure 8: Relativity of the time intervals to the mean of expense RBNS payments

payments constitute the majority of payments this result is reflected into the total where we see values, again higher than the observed ones in terms of VaR and mean, which does not happen when the dynamic risk score is not included. In terms of RBNS disability payments, the inclusion of the risk score does not affect the mean significantly but it does increase the standard deviation, which gives more conservative results and is relevant to the NBII model because results for both models are lower than the observed values. As for the RBNS expense payments, including the risk score slightly decreases the model accuracy because models that do not use it already provide results over the observed values and its inclusion slightly increases the mean, standard deviation, and VaRs. Overall, however, all types of payments are not impacted in the same manner, including dynamic risk scores, which yield better results in terms of payment counts in this simulation comparison.

After analyzing the frequency models, we can now compare the best performing model (the one that uses the NBII distribution) to other models in the literature. However, because most models directly predict the total cost of payments rather than payment counts, we decided to compare total cost instead. Thus, we added a severity model to our dynamic score frequency model. We tested popular distributions, such as the Gamma, log-Normal and inverse normal distributions. We found that fitting distributions for each type of payment separately and including the risk score as a covariate were 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 [31]). We also considered the individual model by Yanez and Pigeon [33], which served as a comparative baseline for the inclusion of dynamic risk scores. Table 9 contains the results of $10,000$ simulations of each described model, and Figure 9 displays the results.

We discuss the results from Table 9 and Figure 9. 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 the individual models are lower than 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 similar, but the standard deviation and the 95 % and the 99 % VaRs are lower, further increasing the accuracy of the model while providing values over the observed reserve. Again, this shows an overall numerical preference for the model in this paper over the one suggested in Yanez and Pigeon [33].

6 Conclusion

In this paper, we introduced an innovative dynamic risk 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 utilize this updated information for the next interval. We applied this new method to the model by Yanez and Pigeon [33] 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 risk score introduced in this

Table 7: Simulation results for outstanding payment counts from models with risk scores

Reserve	Model	Payment	Mean	SD	95% VaR	99% VaR	Observed
RBNS	Weibull	Exposure	5892.83	52.71	5977.85	6016.12	5888.89
	NBII	Medical	57,555.00	1175.52	59,505.15	60,294.10	53,765
		Disability	17,784.90	509.59	18,635.10	18,914.08	18,401
		Expenses	21,245.61	369.21	21,848.40	22,114.06	16,653
		Total	96,585.51	1520.75	99,059.35	100,200.16	88,819
	NBI	Medical	64,539.33	1640.93	67,265.20	68,739.13	53,765
		Disability	18,780.31	767.09	20,070.20	20,592.35	18,401
		Expenses	21,196.75	436.98	21,920.05	22,279.12	16,653
		Total	104,516.39	2007.75	107,773.35	109,374.25	88,819
	POI	Medical	59,938.11	1049.61	61,588.00	62,267.09	53,765
		Disability	18,627.40	420.70	19,337.00	19,612.01	18,401
		Expenses	20,210.12	302.64	20,725.00	20,939.00	16,653
		Total	98,775.63	1405.06	101,034.00	101,934.06	88,819
IBNR	Weibull	Exposure	391.23	24.79	432.49	446.00	564.63
	NBII	Medical	2624.17	237.45	3022.30	3197.02	3424
		Disability	549.57	110.30	736.10	797.01	1016
		Expenses	767.64	81.56	906.05	989.11	1427
		Total	3941.38	335.03	4517.05	4689.02	5867
	NBI	Medical	2679.94	259.08	3110.35	3365.02	3424
		Disability	520.02	105.85	708.00	824.02	1016
		Expenses	749.43	82.10	891.00	956.00	1427
		Total	3949.38	330.45	4493.10	4779.02	5867
	POI	Medical	2381.70	165.00	2655.00	2775.04	3424
		Disability	765.65	107.71	953.00	1038.00	1016
		Expenses	770.37	66.55	885.05	933.02	1427
		Total	3917.72	277.93	4389.00	4553.07	5867
TOTAL	Weibull	Exposure	6284.06	58.10	6377.10	6421.57	6453.52
	NBII	Medical	58,248.00	1046.49	60,015.55	60,706.79	57,189
		Disability	17,054.64	465.71	17,862.30	18,096.99	19,417
		Expenses	21,991.42	371.91	22,576.25	22,895.19	18,080
		Total	97,294.07	1397.37	99,441.40	100,474.90	94,686
	NBI	Medical	67,219.27	1654.18	69,989.05	71,386.46	57,189
		Disability	19,300.33	770.92	20,601.50	21,180.04	19,417
		Expenses	21,946.17	445.50	22,701.15	23,018.12	18,080
		Total	108,465.77	2025.24	111,774.15	113,498.01	94,686
	POI	Medical	62,319.81	1059.18	63,984.35	64,640.03	57,189
		Disability	19,393.05	431.93	20,091.15	20,403.06	19,417
		Expenses	20,980.49	312.32	21,505.00	21,725.08	18,080
		Total	102,693.35	1428.71	105,039.30	105,875.52	94,686

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 risk 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 utilize 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, this work pioneers the introduction of risk scores in a loss reserve context. Thus, we left possible extensions of this idea for future projects. Examples are an implementation of a risk measure that is based on both the number of payments and their cost, or even a risk measure based on the previously observed total cost. Furthermore, correlation between different payment categories of the same claim was deemed complex enough to be considered in a separate future work.

Table 8: Simulation results for outstanding payment counts from models without risk scores

Reserve	Model	Payment	Mean	SD VaR	95% VaR	99% VaR	Observed
RBNS	Weibull	Exposure	5892.83	52.71	5977.85	6016.12	5888.89
	NBII	Medical	41,785.18	549.99	42,696.00	43,085.42	53,765
		Disability	16,857.87	397.87	17,527.10	17,786.06	18,401
		Expenses	19,998.91	298.34	20,490.00	20,717.09	16,653
		Total	78,641.95	915.76	80,136.10	80,834.18	88,819
	NBI	Medical	47,642.49	641.25	48,717.00	49,207.06	53,765
		Disability	18,430.06	583.85	19,361.35	19,736.03	18,401
		Expenses	19,168.58	311.31	19,680	19,884.25	16,653
		Total	85,241.13	1092.88	87,003.45	87,895.25	88,819
	POI	Medical	47,145.71	510.23	47,997.35	48,193.39	53,765
		Disability	18,020.15	229.80	18,400.95	18,586.17	18,401
		Expenses	18,609.22	224.51	18,980.30	19,122.01	16,653
		Total	83,775.08	840.14	85,083.05	85,767.02	88,819
IBNR	Weibull	Exposure	391.23	24.79	432.49	446.00	564.63
	NBII	Medical	2846.31	200.30	3180.05	3309.01	3424
		Disability	886.12	105.34	1066.00	1137.08	1016
		Expenses	965.86	85.21	1110.00	1173.01	1427
		Total	4698.28	332.67	5258.05	5458.02	5867
	NBI	Medical	3008.36	218.95	3376.00	3534.01	3424
		Disability	858.06	109.15	1041.05	1140.01	1016
		Expenses	945.07	80.82	1082.00	1153.02	1427
		Total	4811.49	347.46	5389.05	5665.09	5867
	POI	Medical	1839.70	110.95	2024.20	2133.01	3424
		Disability	652.79	96.06	807.05	879.08	1016
		Expenses	933.54	89.19	1093.05	1152.05	1427
		Total	4741.26	382.15	5332.30	5646.07	5867
TOTAL	Weibull	Exposure	6284.06	58.10	6377.10	6421.57	6453.52
	NBII	Medical	44,631.49	584.63	45,581.15	46,007.09	57,189
		Disability	17,743.98	411.91	18,409.10	18,693.03	19,417
		Expenses	20,964.77	307.66	21,465.05	21,700.11	18,080
		Total	83,340.24	968.01	84,949.20	85,550.04	94,686
	NBI	Medical	50,650.85	680.48	51,771.20	52,287.08	57,189
		Disability	19,288.12	595.33	20,254.20	20,644.22	19,417
		Expenses	20,113.65	321.20	20,640.10	20,840.12	18,080
		Total	90,052.62	1149.97	91,939.70	92,730.19	94,686
	POI	Medical	50,016.84	562.36	50,958.05	51,255.08	57,189
		Disability	18,915.86	243.44	19,300.10	19,520.11	19,417
		Expenses	19,556.02	236.94	19,962.05	20,090.32	18,080
		Total	88,488.73	919.77	90,033.90	90,576.75	94,686

Table 9: Results of the total reserve predictions

	Mean	SD	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 IBNR	4,160,285	488,219	4,475,441	5,000,940	5,386,198
3-component total	149,620,225	3,678,054	152,066,762	155,830,382	158,291,786
Dynamic Risk RBNS	147,488,982	2,957,654	149,446,106	152,473,969	154,292,371
Dynamic Risk IBNR	3,565,489	352,761	3,798,531	4,154,814	4,483,246
Dynamic Risk total	151,054,472	2,980,171	153,058,465	156,103,266	157,962,819
Observed	147,703,974				

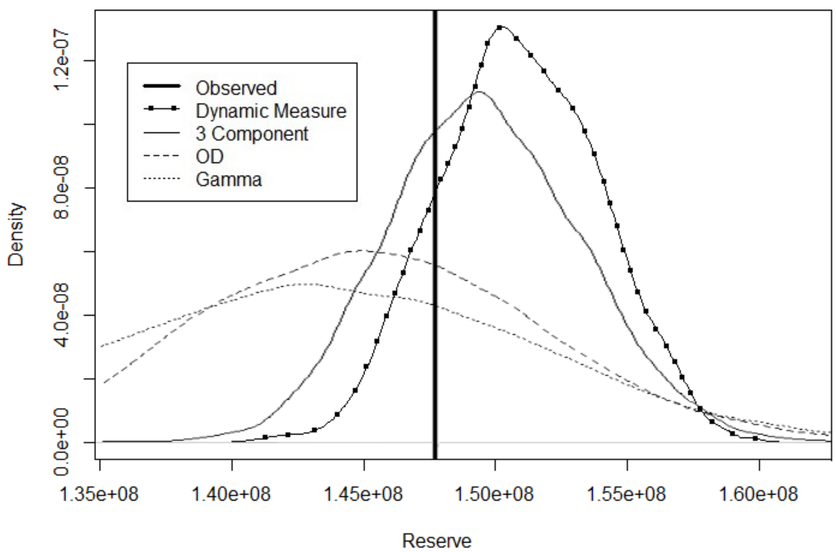

Figure 9: Total reserves for the chosen model

A Appendix

Table 10: Estimated values for the Negative Binomial I Model (RBNS)

Variable	Category	With the risk score			Without the risk score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Type of loss	Single vehicle	0.12	0.48	0.50	0.32	0.72	0.54
	Multi vehicle	0.19	0.17	0.35	0.22	0.22	0.33
	Hit pedestrian	0.37	0.80	0.75	0.76	1.12	0.91
	Other	0.39	0.63	0.46	0.50	0.85	0.50
Injured gender	Male	-0.14	0.08	0.10	-0.17	0.08	0.12
	Unknown	0.02	1.35	0.83	0.11	2.26	0.98
Region	Ontario	-0.08	0.65	1.84	-0.21	0.81	2.29
	West	0.52	0.18	0.46	0.45	0.48	0.72
Injured age	$(18, 25]$	0.07	0.43	0.25	0.05	0.41	0.31
	$(25, 30]$	0.19	0.57	0.36	0.16	0.58	0.42
	$[30, 50]$	0.22	0.59	0.37	0.20	0.59	0.42
	$(50, 70]$	0.27	0.63	0.46	0.26	0.62	0.53
	$(70, \infty)$	0.32	0.61	0.62	0.39	0.56	0.72
	Unknown	-0.21	-1.12	-0.33	-0.31	-1.88	-0.38
Vehicle age	$(3, 6]$	0.02	0.09	-0.00	0.04	0.13	-0.01
	$(6, 10]$	0.01	0.14	0.05	0.06	0.24	0.07
	$(10, 20]$	0.05	0.30	0.15	0.08	0.42	0.19
	$(20, \infty)$	0.03	0.49	0.12	0.12	0.74	0.11
	Unknown	-0.00	0.05	0.05	-0.07	-0.09	0.01
$t_\ell^{(r)}$	$(1, 7]$	-0.03	0.12	0.11	-0.04	0.10	0.08
	$(7, 30]$	-0.12	-0.07	0.08	-0.20	-0.18	0.02
	$(30, 90]$	-0.23	-0.34	0.03	-0.39	-0.65	-0.08
	$(90, 180]$	-0.48	-0.48	0.01	-0.79	-0.96	-0.21
	$(180, 365]$	-0.50	-0.59	-0.10	-0.79	-1.09	-0.33
	$(365, \infty)$	-0.38	-0.27	0.25	-0.46	-0.59	0.03
Initial reserve	$(1000, 5000]$	-0.12	-0.25	-0.34	-0.18	-0.30	-0.38
	$(5000, 10000]$	-0.05	0.24	-0.09	-0.08	0.22	-0.11
	$(10000, 20000]$	-0.02	0.35	0.04	-0.05	0.33	0.07
	$(20000, \infty)$	0.11	0.65	0.23	0.30	0.87	0.30
Time intervals	$(0.25, 0.5]$	0.30	-0.80	-0.12	0.71	0.59	0.28
	$(0.5, 0.75]$	-0.35	-1.28	-0.48	0.46	0.75	0.27
	$(0.75, 1]$	-0.80	-1.90	-0.97	0.36	0.81	0.05
	$(1, 1.25]$	-0.97	-2.03	-1.04	0.45	0.96	0.14
	$(1.25, 1.5]$	-1.28	-2.12	-1.25	0.39	1.00	0.04
	$(1.5, 1.75]$	-1.40	-2.16	-1.20	0.41	1.07	0.18
	$(1.75, 2]$	-1.46	-2.23	-1.29	0.47	1.08	0.19
	$(2, 2.25]$	-1.51	-2.36	-1.25	0.48	0.91	0.27
	$(2.25, 2.5]$	-1.63	-2.61	-1.43	0.43	0.77	0.14
	$(2.5, 2.75]$	-1.69	-2.59	-1.45	0.49	0.75	0.21
	$(2.75, 3]$	-1.91	-2.66	-1.38	0.45	0.78	0.30
	$(3, 3.25]$	-1.93	-2.59	-1.38	0.43	0.78	0.36
	$(3.25, 3.5]$	-2.09	-2.61	-1.54	0.31	0.87	0.26
	$(3.5, 3.75]$	-2.23	-2.43	-1.51	0.23	1.01	0.32
	$(3.75, 4]$	-2.10	-2.20	-1.66	0.40	1.27	0.18
	$(4, 4.25]$	-1.68	-2.01	-1.44	0.66	1.43	0.41
	$(4.25, 4.5]$	-2.13	-2.14	-1.36	0.48	1.70	0.47
	$(4.5, 4.75]$	-2.52	-1.49	-1.58	-0.12	1.85	0.29
	$(4.75, 5]$	-2.22	-2.53	-1.18	0.01	1.44	0.53
Cubic spline	1^{st} polynomial	5.40	6.38	2.77			
	2^{d} polynomial	0.98	0.54	0.53			
	3^{d} polynomial	4.17	4.45	2.33			
Intercept		1.37	-1.29	-1.72	1.46	-1.55	-2.11
$\sigma^{(a)}$		0.08	1.48	0.57	0.50	2.46	0.80
$\psi^{(a)}$		0.11	0.32	0.49			

Table 11: Estimated values for the Negative Binomial II Model (RBNS)

Variable	Category	With the risk score			Without the risk score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Type of loss	Single vehicle	0.06	0.37	0.38	0.13	0.67	0.48
	Multi vehicle	0.17	0.16	0.33	0.22	0.18	0.34
	Hit pedestrian	0.22	0.51	0.59	0.46	0.89	0.81
	Other	0.27	0.41	0.42	0.39	0.69	0.52
Injured gender	Male	-0.12	0.07	0.08	-0.17	0.07	0.11
	Unknown	0.03	0.87	0.75	0.01	1.10	0.88
Region	Ontario	-0.11	0.07	1.69	-0.28	0.26	2.08
	West	0.46	0.22	0.52	0.40	0.58	0.77
Injured age	$(18, 25]$	0.03	0.15	0.26	0.06	0.45	0.32
	$(25, 30]$	0.13	0.26	0.32	0.17	0.57	0.37
	$[30, 50]$	0.16	0.27	0.35	0.23	0.56	0.39
	$(50, 70]$	0.20	0.31	0.42	0.30	0.56	0.49
	$(70, \infty)$	0.22	0.47	0.48	0.38	0.52	0.64
	Unknown	-0.15	-0.82	-0.30	-0.11	-0.73	-0.35
Vehicle age	$(3, 6]$	0.02	0.08	-0.01	0.00	0.11	-0.01
	$(6, 10]$	0.02	0.08	0.04	0.05	0.18	0.06
	$(10, 20]$	0.04	0.14	0.11	0.05	0.34	0.18
	$(20, \infty)$	-0.00	0.31	0.03	0.01	0.66	0.10
	Unknown	0.03	0.07	0.07	-0.05	-0.09	0.01
$t_\ell^{(r)}$	$(1, 7]$	-0.03	0.08	0.07	-0.06	0.12	0.08
	$(7, 30]$	-0.11	0.02	0.06	-0.19	-0.13	0.02
	$(30, 90]$	-0.22	-0.04	0.02	-0.41	-0.55	-0.10
	$(90, 180]$	-0.56	-0.22	-0.03	-0.85	-0.96	-0.24
	$(180, 365]$	-0.60	-0.32	-0.07	-0.92	-0.96	-0.33
	$(365, \infty)$	-0.84	-0.25	0.01	-1.25	-1.24	-0.18
Initial reserve	$(1000, 5000]$	-0.10	-0.22	-0.29	-0.15	-0.27	-0.34
	$(5000, 10000]$	-0.06	0.02	-0.08	-0.08	0.17	-0.08
	$(10000, 20000]$	-0.06	0.13	0.01	-0.11	0.26	0.05
	$(20000, \infty)$	0.07	0.37	0.15	0.06	0.85	0.20
Time intervals	$(0.25, 0.5]$	0.30	-0.24	-0.04	0.59	0.53	0.29
	$(0.5, 0.75]$	-0.29	-0.80	-0.35	0.36	0.57	0.29
	$(0.75, 1]$	-0.71	-1.35	-0.78	0.20	0.55	0.09
	$(1, 1.25]$	-0.90	-1.49	-0.85	0.25	0.76	0.17
	$(1.25, 1.5]$	-1.18	-1.67	-1.03	0.15	0.75	0.11
	$(1.5, 1.75]$	-1.32	-1.75	-1.01	0.15	0.80	0.21
	$(1.75, 2]$	-1.40	-1.83	-1.04	0.20	0.80	0.25
	$(2, 2.25]$	-1.49	-1.99	-1.04	0.21	0.72	0.30
	$(2.25, 2.5]$	-1.65	-2.20	-1.18	0.14	0.55	0.24
	$(2.5, 2.75]$	-1.69	-2.19	-1.19	0.18	0.61	0.26
	$(2.75, 3]$	-1.88	-2.13	-1.18	0.06	0.67	0.33
	$(3, 3.25]$	-1.85	-2.18	-1.21	0.18	0.56	0.32
	$(3.25, 3.5]$	-2.03	-2.15	-1.33	0.09	0.64	0.24
	$(3.5, 3.75]$	-2.19	-2.15	-1.27	-0.02	0.68	0.35
	$(3.75, 4]$	-2.09	-2.07	-1.43	0.09	0.84	0.19
	$(4, 4.25]$	-1.99	-1.90	-1.14	0.23	1.18	0.48
	$(4.25, 4.5]$	-2.59	-1.90	-0.98	-0.33	1.24	0.67
	$(4.5, 4.75]$	-2.92	-2.01	-1.28	-0.21	1.35	0.40
	$(4.75, 5]$	-3.09	-1.74	-0.66	-0.81	1.38	0.85
Cubic spline	1^{st} polynomial	5.16	6.46	2.54			
	2^{d} polynomial	1.01	0.98	0.54			
	3^{d} polynomial	4.08	4.53	2.13			
Intercept		1.50	-0.62	-1.57	1.63	-0.85	-1.97
$\sigma^{(a)}$		0.55	1.18	0.25	0.93	1.90	0.40
$\psi^{(a)}$		0.11	0.32	0.49			

Table 12: Estimated values for the Poisson Model (RBNS)

Variable	Category	With the risk score			Without the risk score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Type of loss	Single vehicle	0.14	0.40	0.41	0.33	0.76	0.49
	Multi vehicle	0.18	0.19	0.29	0.23	0.26	0.28
	Hit pedestrian	0.33	0.53	0.63	0.74	1.01	0.86
	Other	0.30	0.41	0.42	0.45	0.74	0.51
Injured gender	Male	-0.10	0.06	0.07	-0.15	0.05	0.10
	Unknown	0.10	1.03	0.70	0.11	1.42	0.82
Region	Ontario	-0.08	0.43	1.92	-0.23	0.79	2.34
	West	0.36	0.20	0.49	0.26	0.64	0.74
Injured age	$(18, 25]$	0.05	0.07	0.24	0.06	0.27	0.32
	$(25, 30]$	0.13	0.17	0.32	0.15	0.37	0.37
	$[30, 50]$	0.15	0.19	0.33	0.18	0.39	0.39
	$(50, 70]$	0.20	0.24	0.41	0.24	0.41	0.48
	$(70, \infty)$	0.22	0.46	0.50	0.34	0.54	0.67
	Unknown	-0.21	-0.97	-0.32	-0.23	-1.02	-0.36
Vehicle age	$(3, 6]$	0.04	0.08	-0.01	0.04	0.16	-0.00
	$(6, 10]$	0.02	0.09	0.04	0.06	0.24	0.07
	$(10, 20]$	0.05	0.16	0.13	0.08	0.38	0.21
	$(20, \infty)$	0.04	0.30	0.09	0.13	0.75	0.16
	Unknown	0.00	0.04	0.02	-0.09	-0.14	-0.04
$t_\ell^{(r)}$	$(1, 7]$	-0.03	0.05	0.04	-0.05	0.09	0.03
	$(7, 30]$	-0.11	-0.02	-0.01	-0.21	-0.19	-0.08
	$(30, 90]$	-0.21	-0.12	-0.08	-0.43	-0.67	-0.22
	$(90, 180]$	-0.50	-0.22	-0.15	-0.89	-1.02	-0.41
	$(180, 365]$	-0.48	-0.36	-0.21	-0.87	-1.11	-0.50
	$(365, \infty)$	-0.43	-0.22	-0.05	-0.75	-1.11	-0.27
Initial reserve	$(1000, 5000]$	-0.12	-0.19	-0.26	-0.20	-0.28	-0.31
	$(5000, 10000]$	-0.05	0.02	-0.06	-0.09	0.10	-0.06
	$(10000, 20000]$	-0.01	0.11	0.06	-0.06	0.20	0.10
	$(20000, \infty)$	0.15	0.36	0.22	0.31	0.80	0.31
Time intervals	$(0.25, 0.5]$	0.17	-0.61	-0.13	0.52	0.62	0.26
	$(0.5, 0.75]$	-0.43	-1.11	-0.50	0.36	0.77	0.22
	$(0.75, 1]$	-0.86	-1.68	-0.95	0.27	0.82	0.01
	$(1, 1.25]$	-1.08	-1.90	-1.05	0.34	0.88	0.08
	$(1.25, 1.5]$	-1.38	-2.03	-1.27	0.28	0.92	-0.03
	$(1.5, 1.75]$	-1.53	-2.07	-1.25	0.31	0.98	0.09
	$(1.75, 2]$	-1.63	-2.15	-1.31	0.35	0.94	0.09
	$(2, 2.25]$	-1.73	-2.29	-1.28	0.37	0.83	0.19
	$(2.25, 2.5]$	-1.88	-2.42	-1.49	0.34	0.73	0.05
	$(2.5, 2.75]$	-1.94	-2.43	-1.46	0.38	0.75	0.13
	$(2.75, 3]$	-2.06	-2.39	-1.46	0.36	0.80	0.18
	$(3, 3.25]$	-2.12	-2.36	-1.48	0.36	0.80	0.20
	$(3.25, 3.5]$	-2.30	-2.31	-1.56	0.25	0.90	0.15
	$(3.5, 3.75]$	-2.45	-2.27	-1.62	0.14	0.98	0.14
	$(3.75, 4]$	-2.33	-2.19	-1.69	0.30	1.14	0.09
	$(4, 4.25]$	-2.18	-2.15	-1.51	0.52	1.26	0.25
	$(4.25, 4.5]$	-2.48	-2.10	-1.51	0.32	1.37	0.26
	$(4.5, 4.75]$	-3.35	-2.38	-1.73	-0.51	1.12	0.07
	$(4.75, 5]$	-2.96	-2.40	-1.09	-0.53	0.63	0.44
Cubic spline	1^{st} polynomial	5.39	6.48	2.63			
	2^{d} polynomial	1.22	1.02	0.67			
	3^{d} polynomial	4.30	4.51	2.33			
Intercept		1.54	-0.63	-1.56	1.61	-1.34	-2.00
$\psi^{(a)}$		0.11	0.32	0.49			

Table 13: Estimated values for the Negative Binomial II Model (IBNR)

Variable	Category	With the risk score			Without the risk score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Time intervals	$(0.25, 0.5]$	0.12	-0.58	0.08	0.45	0.47	0.50
	$(0.5, 0.75]$	-0.60	-1.19	-0.28	0.16	0.51	0.57
	$(0.75, 1]$	-1.07	-1.46	-0.73	-0.02	0.51	0.39
	$(1, 1.25]$	-1.27	-1.43	-0.82	0.01	0.74	0.49
	$(1.25, 1.5]$	-1.56	-1.49	-1.01	-0.09	0.75	0.44
	$(1.5, 1.75]$	-1.71	-1.47	-1.00	-0.10	0.81	0.54
	$(1.75, 2]$	-1.82	-1.50	-1.04	-0.07	0.81	0.60
	$(2, 2.25]$	-1.93	-1.61	-1.04	-0.09	0.73	0.66
	$(2.25, 2.5]$	-2.10	-1.79	-1.18	-0.17	0.56	0.60
	$(2.5, 2.75]$	-2.15	-1.75	-1.21	-0.13	0.61	0.61
	$(2.75, 3]$	-2.33	-1.70	-1.18	-0.26	0.67	0.69
	$(3, 3.25]$	-2.31	-1.76	-1.22	-0.15	0.56	0.69
	$(3.25, 3.5]$	-2.46	-1.71	-1.34	-0.24	0.64	0.62
	$(3.5, 3.75]$	-2.67	-1.67	-1.26	-0.36	0.69	0.75
	$(3.75, 4]$	-2.60	-1.55	-1.45	-0.24	0.91	0.59
	$(4, 4.25]$	-2.45	-1.38	-1.16	-0.06	1.27	0.89
	$(4.25, 4.5]$	-2.99	-1.36	-1.03	-0.58	1.37	1.05
	$(4.5, 4.75]$	-3.13	-1.34	-1.33	-0.38	1.60	0.81
	$(4.75, 5]$	-3.00	-1.10	-0.74	-0.96	1.71	1.34
Cubic spline	1^{st} polynomial	4.05	5.39	3.32			
	2^{d} polynomial	1.87	-0.42	0.68			
	3^{d} polynomial	3.61	3.99	2.54			
Intercept		1.78	0.03	0.45	1.88	0.37	0.51
$\sigma^{(a)}$		0.60	1.30	0.34	1.02	1.96	0.57
$\psi^{(a)}$		0.20	1.15	0.57	0.20	1.15	0.57

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

Variable	Category	With the risk score			Without the risk score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Time intervals	$(0.25, 0.5]$	0.11	-0.78	-0.10	0.45	0.47	0.50
	$(0.5, 0.75]$	-0.67	-1.23	-0.51	0.16	0.51	0.57
	$(0.75, 1]$	-1.17	-1.78	-1.05	-0.02	0.51	0.39
	$(1, 1.25]$	-1.37	-1.82	-1.14	0.01	0.74	0.49
	$(1.25, 1.5]$	-1.69	-1.88	-1.36	-0.09	0.75	0.44
	$(1.5, 1.75]$	-1.83	-1.89	-1.34	-0.10	0.81	0.54
	$(1.75, 2]$	-1.89	-1.93	-1.43	-0.07	0.81	0.60
	$(2, 2.25]$	-1.99	-2.02	-1.39	-0.09	0.73	0.66
	$(2.25, 2.5]$	-2.11	-2.25	-1.57	-0.17	0.56	0.60
	$(2.5, 2.75]$	-2.18	-2.21	-1.58	-0.13	0.61	0.61
	$(2.75, 3]$	-2.39	-2.26	-1.51	-0.26	0.67	0.69
	$(3, 3.25]$	-2.38	-2.23	-1.53	-0.15	0.56	0.69
	$(3.25, 3.5]$	-2.56	-2.25	-1.69	-0.24	0.64	0.62
	$(3.5, 3.75]$	-2.74	-2.09	-1.67	-0.36	0.69	0.75
	$(3.75, 4]$	-2.66	-1.92	-1.79	-0.24	0.91	0.59
	$(4, 4.25]$	-2.21	-1.72	-1.65	-0.06	1.27	0.89
	$(4.25, 4.5]$	-2.70	-2.00	-1.58	-0.58	1.37	1.05
	$(4.5, 4.75]$	-3.03	-1.19	-1.85	-0.38	1.60	0.81
	$(4.75, 5]$	-2.60	-2.25	-1.29	-0.96	1.71	1.34
Cubic spline	1^{st} polynomial	4.52	5.16	3.45			
	2^{d} polynomial	1.71	0.29	0.74			
	3^{d} polynomial	3.89	4.27	2.77			
Intercept		1.78	0.07	0.55	1.88	0.37	0.51
$\sigma^{(a)}$		0.18	1.62	0.82	1.02	1.96	0.57
$\psi^{(a)}$		0.19	1.08	0.61			

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

Variable	Category	With the risk score			Without the risk score		
		Medical	Disability	Expenses	Medical	Disability	Expenses
Time intervals	$(0.25, 0.5]$	0.07	-0.97	-0.03	0.41	0.67	0.48
	$(0.5, 0.75]$	-0.59	-1.50	-0.46	0.22	0.84	0.51
	$(0.75, 1]$	-1.05	-1.70	-0.94	0.13	0.93	0.32
	$(1, 1.25]$	-1.28	-1.77	-1.06	0.19	1.01	0.41
	$(1.25, 1.5]$	-1.59	-1.79	-1.29	0.13	1.06	0.31
	$(1.5, 1.75]$	-1.75	-1.77	-1.27	0.16	1.13	0.44
	$(1.75, 2]$	-1.87	-1.81	-1.35	0.18	1.10	0.46
	$(2, 2.25]$	-1.99	-1.91	-1.33	0.18	0.98	0.57
	$(2.25, 2.5]$	-2.14	-2.02	-1.53	0.14	0.89	0.43
	$(2.5, 2.75]$	-2.21	-2.02	-1.52	0.18	0.88	0.50
	$(2.75, 3]$	-2.34	-1.98	-1.53	0.14	0.93	0.55
	$(3, 3.25]$	-2.40	-1.95	-1.53	0.14	0.94	0.58
	$(3.25, 3.5]$	-2.58	-1.90	-1.62	0.02	1.04	0.55
	$(3.5, 3.75]$	-2.72	-1.84	-1.68	-0.10	1.12	0.54
	$(3.75, 4]$	-2.60	-1.75	-1.75	0.06	1.31	0.50
	$(4, 4.25]$	-2.44	-1.77	-1.59	0.29	1.42	0.66
	$(4.25, 4.5]$	-2.73	-1.70	-1.60	0.13	1.57	0.66
	$(4.5, 4.75]$	-3.51	-1.93	-1.82	-0.61	1.43	0.48
	$(4.75, 5]$	-3.06	-2.02	-1.22	-0.61	1.07	0.94
Cubic spline	1^{st} polynomial	5.79	5.82	3.42			
	2^{d} polynomial	1.12	-0.19	0.79			
	3^{d} polynomial	4.44	4.19	2.78			
Intercept		1.83	0.14	0.56	1.83	0.14	0.56
$\psi^{(a)}$		0.12	1.16	0.52	0.12	1.16	0.52

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

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

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

[6] 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.

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

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

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

[15] 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.

[16] 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.

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

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

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

[25] 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.

[26] 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.

[27] 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.

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

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

[30] 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.

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

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

[33] 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.

[34] 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.

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