Individual Loss Reserving Using Activation Patterns

Marie Michaelides[1*], Mathieu Pigeon[1], Hélène Cossette[2]

[1]Department of Mathematics, Université du Québec à Montréal
[2]École d'actuariat, Université Laval

[*]Corresponding author: michaelides.marie@uqam.ca

Abstract

The occurrence of a claim often impacts not one but multiple insurance coverages provided in the contract. To account for this multivariate feature, we propose a new individual claims reserving model built around the activation of the different coverages to predict the reserve amounts. Using the framework of multinomial logistic regression, we model the activation of the different insurance coverages for each claim and their development in the following years, i.e. the activation of other coverages in the later years and all the possible payments that might result from them. As such, the model allows us to complete the individual development of the open claims in the portfolio. Using a recent automobile dataset from a major Canadian insurance company, we demonstrate that this approach generates accurate predictions of the total reserves as well as of the reserves per insurance coverage. This allows the insurer to get better insights in the dynamics of his claims reserves.

1 Introduction

Claims reserving is known to be one of the most crucial tasks performed by actuaries in insurance companies all over the world. Insurers must accurately predict the liabilities that will arise from open and future claims. This allows them to answer the reporting standards that they are subject to and to preserve a sufficient amount of capital with which they can set aside adequate reserves to fulfill their obligations to the policyholders and avoid financial ruin.

In the actuarial practice, claims data is commonly aggregated on an occurrence year and development year basis in run-off triangles. These then help actuaries to evaluate the reserves for the portfolio as a whole. One very common method that uses such triangles is the Chain Ladder model introduced by Mack [1993] and further discussed in Mack [1994], Mack [1999] or Mack and Venter [2000].

Over the years, several authors have challenged the robustness of this model. In particular, the recent increase in the quantity and availability of data has contributed to questioning the use of such aggregate methods. The early work of Buhlmann et al. [1980], Hachemeister [1980] and Norberg [1986] attempted to benefit from these larger quantities of data. A few more years were however necessary to obtain the necessary computing resources to move from the classical run-off triangles to the so-called micro-level claims reserving models. Antonio and Plat [2014] were the first to truly incorporate information related to the policyholder or even the claim itself into their model by building on the aforementioned work as well as on prior work performed by Norberg [1993], Norberg [1999] and Haastrup and Arjas [1996]. They demonstrate that using the detailed information available to the insurer at the claim level allows to obtain more accurate predictions for the reserves.

Micro-level models now abound in the actuarial literature thanks to the contributions of several authors. For example, Pigeon et al. [2013] propose a discrete time model for the payments and then extend their work in Pigeon et al. [2014] to include incurred losses as well. Other authors have also opened the way to non-parametric approaches to claims reserving. Wüthrich [2018] was the first to introduce the use of Breiman et al. [1984]'s Classification and Regression Tree (CART) algorithm in a micro-level reserving model. Building on this work, Lopez [2019] and Lopez et al. [2016] apply the CART algorithm with censored data using, respectively, survival analysis and copulas to account for the possible dependence between the development time of the claim and its ultimate amount.

In addition to the constant increase in the quantity of data, the growing diversification of the products offered by insurers contributes to further complexify the work of actuaries. Evolving in a very competitive world and taking advantage of the rise of new technologies, insurers must remain constantly aware of changes and evolutions in their clients' needs. To answer them, they often diversify their offer, resulting in the multiplication of coverages provided within a policy. Actuaries must thereby refine their models to keep up with this diverseness in the portfolios.

In this paper, we propose a model that not only makes use of the large quantity of data available to an insurer in a micro-level model but also takes into account the different coverages that a policy offers and their dynamics within the portfolio. We show that allowing dependence between coverages increases prediction accuracy of both the total portfolio reserves and the reserves per coverage. This in turn allows the insurer to gain further insights into the dynamics of his portfolio. To the best of our knowledge, such form of dependence between coverages has not yet been modelled in a granular reserving framework. Zhou and Zhao [2010] and Lopez [2019] both used copulas in such a context to model the dependence between the event times and delay in the development of a claim or the development time and the final amount of the claim. Pešta and Okhrin [2014] use time series and copulas to take into account the dependence between payment amounts made at different stages in the development of a claim. It is rather in the actuarial pricing literature that we find examples of dependency modelling between insurance coverages. Frees and Valdez [2008] and together with Frees et al. [2009] used copulas to model the dependence between different claim types. They begin by identifying the coverage(s) impacted by a claim, then those for which a payment is made before predicting the associated severity. More recently, Côté et al. [2022] extend their work by introducing a Bayesian model for multivariate and multilevel claim amounts, therefore facilitating the treatment of open claims which are of crucial importance in insurers' datasets.

We propose a micro-level reserving model that builds upon the work of Frees and Valdez [2008] and Frees et al. [2009] to include the dependence between multiple insurance coverages. Section 2 presents the statistical model that is based on activation patterns and illustrates how claims are handled at different stages of their development. In Section 3, we apply the model to a Canadian automobile insurance dataset with four different coverages. We begin by presenting the data, the process of model fitting and finally the results obtained. Section 4 draws the conclusion of our work.

2 An activation pattern model for claims reserving

This section introduces the model. We present the model components and the way they evolve as the claim develops. We use an example to illustrate the model and discuss the treatment of claims at various stages of their development.

Figure 1 shows the typical development process for a single claim. When a claim occurs, the policyholder reports it to the insurance company. This can take place either as soon as the claim occurs or with a certain reporting delay. Once reported, the insurer records the claim and opens a new file in his claims management system. Since insurance companies typically provide their policyholders with multiple coverages under a single insurance policy, the reporting of the claim activates at least one of these coverages. The insurer can then start making payments towards the policyholder. During its development, new information related to the claim can be brought to the insurer that can result in the activation of one or more additional coverages included in the insurance policy. Payments will then continue until settlement of the claim.

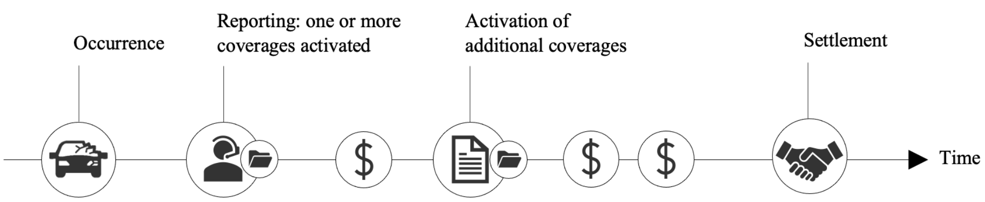

Figure 1: Development process of a claim

We seek to model the way in which a claim can activate multiple coverages either upon reporting or later, and the underlying dependence between them. Note that in the remainder of this paper, we work in discrete time and group the data per year. We could however have chosen any other time unit.

Although automobile claims often display shorter lifetimes than in other branches of insurance, it is still fairly common to observe claims that remain open for more than one year. We thereby structure our model based on the development years. In Section 2.1, we present the model for a given development year j before moving on to the year $j+1$ in Section 2.2 and to any additional development years in Section 2.3.

2.1 Development year j

Let $c = 1, ..., C$ denote a coverage provided by an insurer. Each policy in force can incur a claim that will impact one or more of the C insurance coverages that the policyholder benefits from. For each claim i with $i = 1, ..., n$ and a given development year j with $j = 1, ..., J$, we define the random vector $\boldsymbol{A}_{i,j}$ of dimension $1 \times C$. This vector indicates the pattern in which claim i activates the different coverages during year j.

With C insurance coverages and assuming that at least one must be activated when claim i is reported, there are $V = 2^C - 1$ different activation pattern vectors possible in year j. Let $\mathcal{V}$ be the ensemble of the V possible activation pattern vectors in year j. We then define the v^{th} possible pattern, a realisation of the random vector $\boldsymbol{A}_{i,j}$, as $\boldsymbol{a}_{i,j}^v \in \mathcal{V}$, with $v = 1, ..., V$ such that:

$$\boldsymbol{a}_{i,j}^v = \left[a_{i,j,1}^v \; a_{i,j,2}^v \; \dots \; a_{i,j,C}^v \right],$$

where

$$a_{i,j,c}^v = \begin{cases} 1, & \text{if coverage } c \text{ is activated in the } v^{\text{th}} \text{ pattern for claim } i \text{ in year } j \\ 0, & \text{otherwise.} \end{cases}$$

We use a multinomial logistic regression to model $\boldsymbol{A}_{i,j}$:

$$\mathrm{P}[\boldsymbol{A}_{i,j} = \boldsymbol{a}_{i,j}^v | \boldsymbol{x}_i, \boldsymbol{\beta}] = \frac{\exp\left(\boldsymbol{x}_i' \boldsymbol{\beta}_{j,v}\right)}{\sum_{k=1}^{V} \exp\left(\boldsymbol{x}_i' \boldsymbol{\beta}_{j,k}\right)}, \tag{1}$$

where $\boldsymbol{x}_i'$ is a $1 \times m$ vector of covariates for claim i that does not depend on the development year nor on the insurance coverage and $\boldsymbol{\beta}_{j,v}$ is the $1 \times m$ vector of parameters that can vary with the different coverages and thus depends on the v^{th} pattern.

Knowing which coverages claim i activates in year j thanks to $\boldsymbol{A}_{i,j}$, the insurer can move on to the next step of the claim development process depicted in Figure 1 and determine which of the active coverages will incur a payment within the year. We define $\boldsymbol{P}_{i,j}$, a $1 \times C$ vector in which each entry indicates whether a payment has been made for the corresponding active coverage or not:

$$P_{i,j,c} | (A_{i,j,c} = 1) = \begin{cases} 1, & \text{if a payment has been made for claim } i \\ & \quad \text{and coverage } c \text{ in year } j \text{ given } A_{i,j,c} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

For each insurance coverage c, we assume that

$$P_{i,j,c} | (A_{i,j,c} = 1) \sim \text{Bernoulli}\big(\pi_{j,c}(\boldsymbol{x}_i' \boldsymbol{\gamma}_{j,c})\big). \tag{2}$$

For development year j, the probability $\pi_{j,c}(\boldsymbol{x}_i' \boldsymbol{\gamma}_{j,c})$ is given by

$$\pi_{j,c}(\boldsymbol{x}_i' \boldsymbol{\gamma}_{j,c}) = \frac{\exp\left(\boldsymbol{x}_i' \boldsymbol{\gamma}_{j,c}\right)}{1 + \exp\left(\boldsymbol{x}_i' \boldsymbol{\gamma}_{j,c}\right)}.$$

The vector of covariates $\boldsymbol{x}_i'$ is the same as the one used for the activation patterns in Equation 1 and $\boldsymbol{\gamma}_{j,c}$ is the $1 \times m$ vector of coefficients varying with each coverage c and development year j.

Once we know which coverages are active in year j and which of these have incurred a payment, we can calculate the corresponding severity per coverage for claim i. The actuarial literature contains many examples of the use of Generalized Additive Models for the Location, Scale and Shape (GAMLSS) for this purpose, thanks to the level of flexibility that these models provide (see Stasinopoulos et al. [2017] for more details on the use of the GAMLSS). We use them to predict the expected severity incurred by claim i:

$$\mathrm{E}[Y_{i,j,c}|P_{i,j,c} = 1, A_{i,j,c} = 1] = g^{-1}(\boldsymbol{x}_i'\boldsymbol{\alpha}_{\boldsymbol{j,c}} + \alpha^*_{j,c}j). \tag{3}$$

Here, $Y_{i,j,c}$ represents the severity associated to claim i in year j for insurance coverage c, $g(.)$ is the link function, $\boldsymbol{\alpha}_{j,c}$ are the coefficients that depend on the insurance coverages and $\alpha^*_{j,c}$ measures the effect of development year j for coverage c.

With the activation patterns, payment vectors and corresponding severities, we know precisely which coverages claim i activated in year j and the severity associated to each of these coverages.

2.2 Development year $j + 1$

Knowing $\boldsymbol{A}_{i,j}$, we are aware of the coverages that claim i has activated at the start of year $j+1$. We assume that once an insurance coverage is active, it remains active until the settlement of the claim. For claim i and coverage c in development year j, we have

$$A_{i,j+k,c}|(A_{i,j,c} = 1) = 1, \quad \text{for all } k \geq 1. \tag{4}$$

As such, we now depict the set of possible patterns for $\boldsymbol{A}_{i,j+1}$ as $\mathcal{V}^* \subset \mathcal{V}$. More generally, we write

$$\mathrm{P}[\boldsymbol{A}_{i,j+1} = \boldsymbol{a}^{\nu*}_{i,j}|\boldsymbol{A}_{i,j}, \boldsymbol{x}_i, \boldsymbol{\beta}] = \begin{cases} \frac{\mathrm{P}[\boldsymbol{A}_{i,j+1}=\boldsymbol{a}^{\nu*}_{i,j}|\boldsymbol{x'}_i, \boldsymbol{\beta}_{j+1}]}{\sum_{\boldsymbol{a}^{\nu*}_{i,j} \in \mathcal{V}^*} \mathrm{P}[\boldsymbol{A}_{i,j+1}=\boldsymbol{a}^{\nu*}_{i,j}|\boldsymbol{x'}_i, \boldsymbol{\beta}_{j+1}]}, & \text{if } \boldsymbol{a}^{\nu*}_{i,j} \in \mathcal{V}^* \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where we further assume a Markovian property for the activation pattern vectors and build them using only the information available in the previous year. The vector of risk factors $\boldsymbol{x'}_i$ remains unchanged during the development of the claim and is thus the same as in year j. The vector of parameters $\boldsymbol{\beta}_{j+1}$, that we can write $\boldsymbol{\beta}_{j+1,\nu*}$ is still specific to the current development year, i.e. $j + 1$ and the new activation pattern observed.

Also assuming the Markovian property for the payment vectors, we then have:

$$P_{i,j+1,c}|(A_{i,j+1,c} = 1), \boldsymbol{A}_{i,j} \sim \mathrm{Bernoulli}\big(\pi_{j+1,c}(\boldsymbol{x}_i'\boldsymbol{\gamma}_{j+1,c})\big), \tag{6}$$

where

$$\pi_{j+1,c}(\boldsymbol{x}_i'\boldsymbol{\gamma}_{j+1,c}) = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\gamma}_{j+1,c})}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\gamma}_{j+1,c})},$$

with the parameter vectors $\boldsymbol{\gamma}_{j+1,c}$ depending once again on the specific insurance coverage c.

Knowing which coverages are active in year $j+1$ and which of them incurred a payment, we can finally determine the severity of these payments. As in development year j, we suggest the use of the GAMLSS to model the severity per coverage using the resulting vectors of activation and payment patterns:

$$\mathrm{E}[Y_{i,j+1,c}|P_{i,j+1,c}=1, A_{i,j+1,c}=1, \boldsymbol{A}_{i,j}] = g^{-1}(\boldsymbol{x}_i'\boldsymbol{\alpha}_{j+1,c} + \alpha^*_{j+1,c}(j+1)). \tag{7}$$

2.3 Development years $j+k$, for $k \geq 2$

If enough data is available and if the insurer deems it appropriate, he can repeat the model described in Sections 2.1 and 2.2 in years $j+k$, for $k \geq 2$ until development year j^*. In year j^*, we assume that even if the claim is still open, its development stabilizes, i.e. it does not activate any additional coverage. Additional payments can however still incur and although claims with such long settlement delays are rarer in automobile insurance, we should not disregard them. A longest lifetime often indicates a larger severity. If the claim is still open in development year j^*, we write its remaining severity starting from that year as $\boldsymbol{Y}_{i,j^*}|\boldsymbol{P}_{i,j+k}, \boldsymbol{A}_{i,j+k}, \boldsymbol{A}_{i,j+k-1}, \boldsymbol{Y}_{i,j^*} \geq \boldsymbol{Y}_{i,j+k}$ where we note the term $\boldsymbol{Y}_{i,j^*} \geq \boldsymbol{Y}_{i,j+k}$ in the condition. $\boldsymbol{Y}_{i,j^*}$ represents the severity of claim i for the development years starting from year j^* and encapsulates the possible remaining amounts that might be paid at later dates for claims with longer lifetimes.

We schematize the model described in Sections 2.1, 2.2 and 2.3 in Figure 2. We see how the different components interact and the dependence on the activation pattern vector for year j, meaning $\boldsymbol{A}_{i,j}$. As shown in Figure 2, we assume that a claim can activate a coverage at any time in a given development year j until the 30$^{\text{th}}$ of December. The insurer then records any payment made for the active coverages on the 31$^{\text{st}}$ of December.

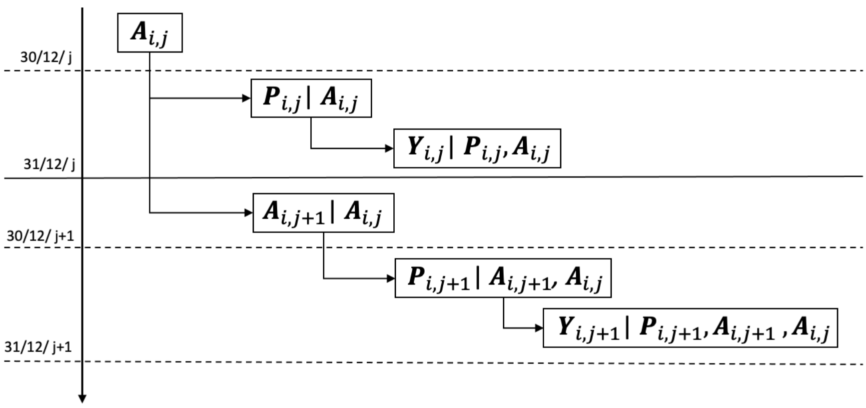

Figure 2: Summary of the activation patterns model

Example 1. To illustrate the model described in the previous sections, we consider in Figure 3 a simple example. Assume that an insurer offers $C = 2$ *different insurance coverages. The set of initial activation patterns of dimension* $V = 3$ *is given by* $\mathcal{V} = \{(0\,1)\ (1\,0)\ (1\,1)\}$. *We take a closer look at the development of claim* i *from development year* $j = 1$ *onwards.*

- *Year* $j = 1$
  Suppose that in development year $j = 1$, *we observe, for claim* i, *the activation pat-*

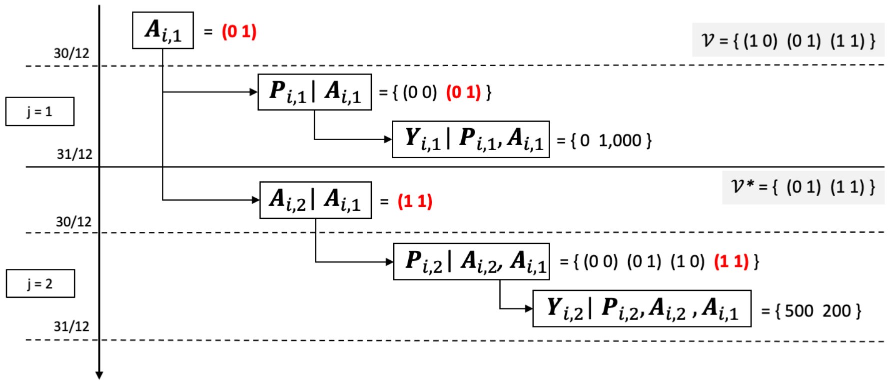

Figure 3: Illustration of the activation patterns model

tern $\boldsymbol{A}_{i,1} = \boldsymbol{v}_{i,1}^1 = (0\ 1)$*, i.e. claim* i *activates the second coverage but not the first one. Knowing* $\boldsymbol{A}_{i,1}$*, the possible payment pattern vectors are* $\boldsymbol{P}_{i,1}|\boldsymbol{A}_{i,1} = \{(0\ 0)\ \ (0\ 1)\}$*.*

If we assume that the resulting pattern for the payments is $(0\ 1)$*, then we have everything required to compute the severity for claim* i *in year 1. In this case, the severity for the first coverage, i.e.* $Y_{i,1,1}$*, will be equal to 0 since claim* i *did not activate it and hence, no payment incurred. The severity for the second coverage, i.e.* $Y_{i,1,2}$ *can be estimated using simulations drawn from an appropriate distribution. Let's say we obtain* $\boldsymbol{Y}_{i,j}$ *= (0 ; 1,000), then the total severity for claim* i *in year 1 is equal to 1,000.*

- *Year* $j+1=2$
  In the following year, the second coverage will remain active and the new set of possible activation patterns $\mathcal{V}^* \subset \mathcal{V}$ *is given by* $\mathcal{V}^* = \{(0\ 1)\ \ (1\ 1)\}$*.*

  Suppose that we obtain $\boldsymbol{A}_{i,2}|\boldsymbol{A}_{i,1} = (1\ 1)$*. The possible payment patterns are then* $\boldsymbol{P}_{i,2}|\boldsymbol{A}_{i,2}, \boldsymbol{A}_{i,1} \in \{(0\ 0)\ \ (0\ 1)\ \ (1\ 0)\ \ (1\ 1)\}$ *and we find, for example, that* $\boldsymbol{P}_{i,2}|\boldsymbol{A}_{i,2}, \boldsymbol{A}_{i,1} = (1\ 1)$*, i.e. the insurer recorded payments for both coverages.*

  We can finally compute the severity for both coverages for year 2 using the appropriate models. We find for example that $\boldsymbol{Y}_{i,j+1}|\boldsymbol{P}_{i,2}, \boldsymbol{A}_{i,2}, \boldsymbol{A}_{i,1}$ *= (500 ; 100), which gives a total severity for year 2 of* 600*.*

  The resulting total severity for claim i *is equal to* Y_i *= 1,600.*

2.4 IBNR, RBNS and RBNP claims

This section describes how the model handles claims at different stages of their development. We first provide an illustration of two claims with varying evaluation dates of the reserves and then describe the simulation routine that we use to predict the reserves.

Example 2. We consider in Figures 4, 5 and 6 three different evaluation dates of the reserves. In Figure 4, we evaluate the reserves in development year $j=1$ *at time* t_1*, before the 31st of December. At this time, claim 1 occurred and is reported since we observe its activation pattern* $\boldsymbol{A}_{1,1}$*. However, the insurer has not recorded any payment yet, making it a reported but not paid claim (RNBP). We have to estimate its payment pattern* $\boldsymbol{P}_{1,1}$ *and corresponding severity for year* j *(which corresponds to the first year of development for this specific claim),* $\boldsymbol{Y}_{1,1}$*. Then we will also be able to simulate the necessary components for year* $j+1$*, i.e.* $\boldsymbol{A}_{1,2}, \boldsymbol{P}_{1,2}$ *and* $\boldsymbol{Y}_{1,2}$*. Claim 2 has not occurred yet and therefore, does not appear in the analysis.*

Figure 5 illustrates the case where we choose to evaluate the reserves at time t_2, a few months after t_1 during the year $j = 2$, for example on the 1st of January. This time, the insurer already recorded a payment for claim 1 but it is not settled yet. It is thus now a reported but not settled (RBNS) claim. Knowing $\boldsymbol{A}_{1,1}$, we are able to predict $\boldsymbol{A}_{1,2}, \boldsymbol{P}_{1,2}$ and $\boldsymbol{Y}_{1,2}$. Claim 2 has occurred but has not been reported yet and is thus an incurred but not reported (IBNR) claim. We will have to predict its first activation pattern $\boldsymbol{A}_{2,1}$ and all subsequent components required for the evaluation of the reserves.

Finally, we choose in Figure 6 t_3 as evaluation date. In this case, both claims 1 and 2 have been reported although they are not settled yet and a payment already took place for claim 1. Claim 1 is a RBNS claim while claim 2 is still a RBNP claim. For claim 1, we need to estimate the payment pattern and severity for development year 2, i.e. $\boldsymbol{P}_{1,2}$ and $\boldsymbol{Y}_{1,2}$. For claim 2 which is in its first year of developement, we have observed $\boldsymbol{A}_{2,1}$ but will still need to simulate $\boldsymbol{P}_{2,1}, \boldsymbol{Y}_{2,1}, \boldsymbol{A}_{2,2}, \boldsymbol{P}_{2,2}$ and $\boldsymbol{Y}_{2,2}$.

If the claims are still open after two development years, their final severity will be given by, respectively, $\boldsymbol{Y}_{1,2+k}|\boldsymbol{Y}_{1,2+k} \geq \boldsymbol{Y}_{1,2}$ and $\boldsymbol{Y}_{2,2+k}|\boldsymbol{Y}_{2,2+k} \geq \boldsymbol{Y}_{2,2}$, as discussed in Section 2.3.

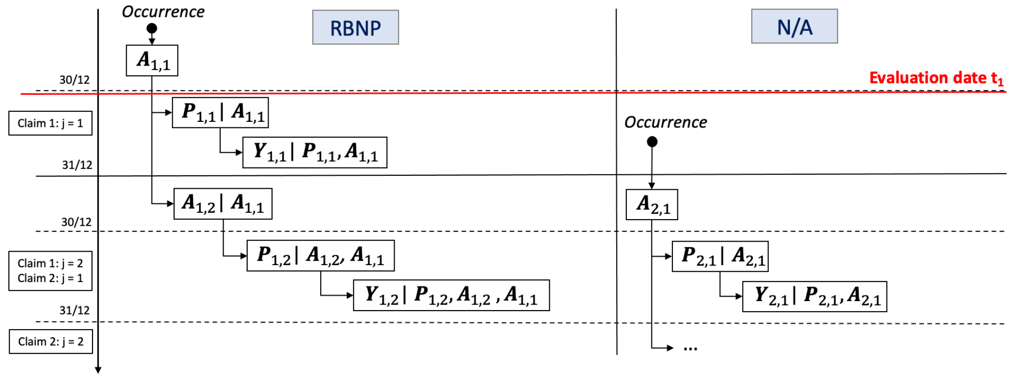

Figure 4: Illustration of evaluation date t_1

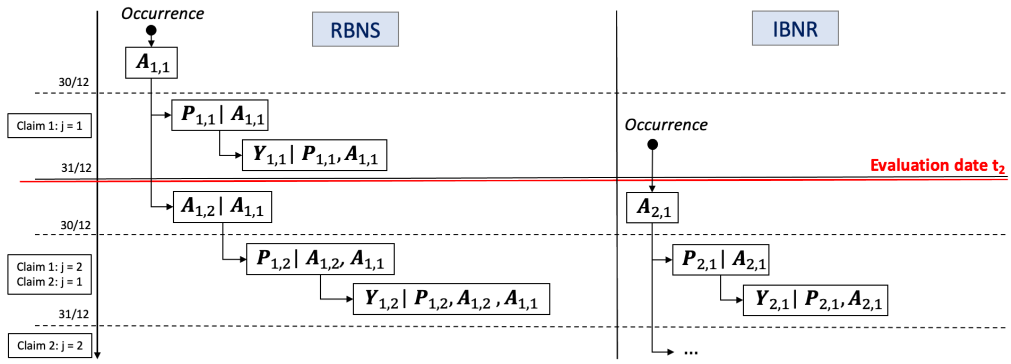

Figure 5: Illustration of evaluation date t_2

2.4.1 Simulation routine

To properly consider all open claims in a given dataset, we create a simulation routine that we can tailor to the specific development stage of different types of claims. We present the routine for the IBNR claims and the RBNS claims, including those

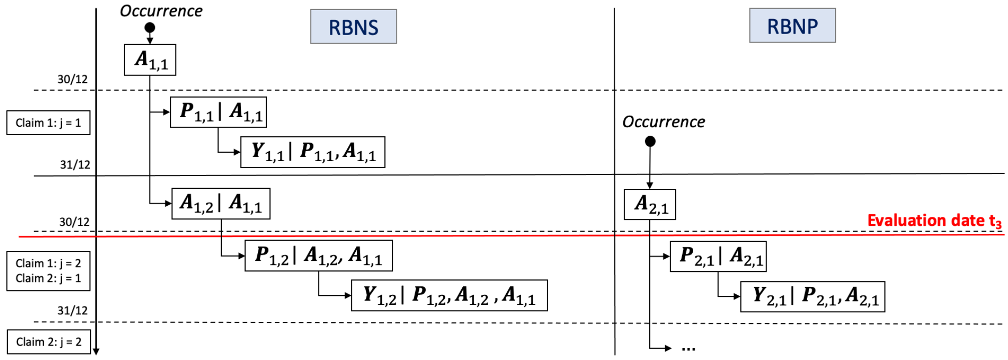

Figure 6: Illustration of evaluation date t_3

with longer development times for which we assume that they will not activate any additional coverages.

IBNR claims

1. For development year $j = 1$:

   (a) For each IBNR claim i_{IBNR}, with $i_{\text{IBNR}} = 1, \ldots, n_{\text{IBNR}}$, simulate the first activation pattern vectors $\boldsymbol{A}_{i_{\text{IBNR}},j}$ using the parameters estimated for the multinomial logit model from in Section 2.1.

   (b) For the inactive coverages, i.e. $A_{i_{\text{IBNR}},j,c} = 0$, then the corresponding $P_{i_{\text{IBNR}},j,c}$ is automatically generated as 0. For the active coverages only, simulate the payment patterns using the Bernoulli models with parameters $\hat{\pi}_{j,c}$.

   (c) If the payment indicator $P_{i_{\text{IBNR}},j,c} = 0$, either because the corresponding coverage is not active or because we simulated it as such in the previous step, then automatically set the corresponding severity for IBNR claim i_{IBNR} related to that coverage to 0, i.e. $Y_{i_{\text{IBNR}},j,c} = 0$. If $P_{i_{\text{IBNR}},j,c} = 1$ from the previous step, we simulate the severity from the distribution selected for the specific coverage.

2. For development years $j > 1$:

   (a) Knowing the activation pattern $\boldsymbol{A}_{i_{\text{IBNR}},j-1}$ for the previous development year, only keep the activation patterns that are now still possible. Renormalize the probabilities for the remaining activation patterns and simulate them using the multinomial logit model to obtain the vector $\boldsymbol{A}_{i_{\text{IBNR}},j}$.

   (b) If $A_{i_{\text{IBNR}},j,c} = 0$, $P_{i_{\text{IBNR}},j,c} = 0$. For the active coverages only, simulate the payment patterns using the Bernoulli models with parameters $\hat{\pi}_{j,c}$.

   (c) If the payment indicator $P_{i_{\text{IBNR}},j,c} = 0$, generate $Y_{i_{\text{IBNR}},j,c} = 0$. If $P_{i_{\text{IBNR}},j,c} = 1$ from the previous step, simulate the severity from the distribution selected for the specific coverage. The total severity simulated for the IBNR claims is given by $\boldsymbol{Y}^{IBNR} = \sum_{i_{\text{IBNR}}=1}^{n_{\text{IBNR}}} \sum_{j=1}^{J} \sum_{c=1}^{C} Y_{i_{\text{IBNR}},j,c}$.

RBNS claims At the time of evaluation, these claims are already in their j^{th} development year. We already know the activation patterns, payment patterns and severities for year $j-1$ and the simulation routine is:

1. For development year $j > 1$:
   (a) Knowing $\boldsymbol{A}_{i_{RBNS},j-1}$, with $i_{RBNS} = 1, ..., n_{RBNS}$, re-normalize the probabilities for the remaining possible activation patterns and simulate them using the multinomial logit model from Section 2.2 to obtain the vector $\boldsymbol{A}_{i_{RBNS},j}$.
   (b) If $A_{i_{RBNS},j,c} = 0$, $P_{i_{RBNS},j,c} = 0$. For the active coverages only, simulate the payment patterns using the Bernoulli models with parameters $\hat{\pi}_{j,c}$.
   (c) If $P_{i_{RBNS},j,c} = 0$, $Y_{i_{RBNS},j,c} = 0$. If $P_{i_{RBNS},j,c} = 1$ from the previous step, simulate the severity from the distribution selected for the specific coverage. The total severity simulated for these claims is given by $\boldsymbol{Y}^{\text{RBNS}} = \sum_{i_{RBNS}=1}^{n_{RBNS}} \sum_{j>1}^{J} \sum_{c=1}^{C} Y_{i_{RBNS},j,c}$.

RBNS claims with longer development times To remain thorough in our analysis, we must also consider the longer RBNS claims that we previously mentioned in Section 2.3. After a certain development year j^*, we assume that the activation pattern for claim i_{RBNS+}, with $i_{RBNS+} = 1, ..., n_{RBNS+}$, $\boldsymbol{A}_{i_{RBNS+},j^*}$ remains unchanged. We therefore simulate the additional severity that this claim can incur using the same severity distribution as the one we selected for year $j+k < j^*$, where $j+k$ denotes the last development year in which the activation patterns might have changed. Note that $Y_{i_{RBNS+},j*,c}$ must be at least equal to $Y_{i_{RBNS+},j+k,c}$. In what follows, we detail our method for coverage c. For claim i_{RBNS+}, we work with the distribution of the random variable $Z_{i_{RBNS+},c} = \max(Y_{i_{RBNS+},j+k,c}, Y_{i_{RBNS+},j^*,c})$. Its cumulative distribution function is given by

$$F_{Z_{i_{RBNS+},c}}(z) = \mathrm{P}[Z_{i_{RBNS+},c} \leq z] = \mathrm{P}[\max(Y_{i_{RBNS+},j+k,c}, Y_{i_{RBNS+},j^*,c}) \leq z].$$

We can rewrite this as

$$F_{Z_{i_{RBNS+},c}}(z) = \begin{cases} 0, & z \leq Y_{i_{RBNS+},j^*,c} \\ F_{\boldsymbol{Y}_{j+k,c}}(z), & z > Y_{i_{RBNS+},j^*,c}. \end{cases}$$

The inverse cumulative distribution function of Z is then given by

$$F^{-1}_{Z_{i_{RBNS+},c}}(u) = \begin{cases} Y_{i_{RBNS+},j^*,c}, & 0 \leq u \leq F_{\boldsymbol{Y}_{j+k,c}}^{-1}(Y_{i_{RBNS+},j^*,c}) \\ F^{-1}_{\boldsymbol{Y}_{j+k,c}}(u), & u > F^{-1}_{\boldsymbol{Y}_{j+k,c}}(Y_{i_{RBNS+},j^*,c}). \end{cases} \tag{8}$$

3 Numerical Application

We dedicate this section to the application of the model described in Section 2 to an automobile dataset from a Canadian insurance company. We first provide an overview and some exploratory analysis of the dataset before detailing the results obtained from the estimation of the different parameters of our model. We finally present the results obtained by simulations for the final reserves and compare them with those obtained using three other models that we describe in greater details below.

3.1 Data exploration

Our dataset contains 656,153 claims which have occurred between the 1st of January 2015 and the 30th of June 2021. Each of these claims comes with information related either to the insured, the car driven or the claim itself, and impacts at least one of four possible insurance coverages provided by the insurer.

3.1.1 Risk factors

Table 1 presents the risk factors that we will use in our analysis to build the different models. We provide further insights in Figure 7. Note that the risk factors `GENDER` and `YOB` contained a significant proportion of missing values. We used the `R` package `mice` to fill them in. This package uses Fully Conditional Specification (FCS) where binary data (`GENDER`) is imputed via logistic regression while unordered categorical data (`YOB`) is imputed with polytomous logistic regression. More details can be found in van Buuren and Groothuis-Oudshoorn [2011]. Note that this data filling procedure did not cause any significant changes in the results that we will discuss later on.

Table 1: Description of the risk factors provided for each claim in the dataset

Risk factors	
GENDER	Gender of the insured.
YOB	Year of birth: decade during which the insured was born.
VU	Use of the vehicle made by the insured.
AM	Annual distance driven by the insured (in km).
PROV	Place of occurrence of the claim: one of the Canadian provinces or the USA.
FR	Fault rating: evaluation of the insured's level of responsability in the accident.

Our dataset includes around 60% of male against 40% of female insureds, more than half of them being born between 1960 and 1989. Around half of the insureds use their vehicle for commuting purposes while an additional 37% use it for pleasure. The remaining 13% use their vehicle for commercial or business reasons, or did not disclose that information to the insurer. 76% of the insureds drive between 10,000 and 20,000 kilometers per year according to the annual mileage risk factor which is the only continuous one in our dataset. A bit more than 50% of the claims occur in Ontario, 20% in Alberta and the remaining 30% take place in the other Canadian provinces or in the United States of America. Finally, the insurance company assesses the insured's level of responsibility in an accident and issues a fault rating. In 37% of the claims, the insured was not considered to be at fault. The rating could not be applied for 32% of the claims and in 27% of the cases, the insured was deemed at fault. In very few cases could the company determine a partial level of responsibility or conclude that no fault was committed by neither of the parties involved in the claim.

3.1.2 Insurance coverages

As previously mentioned, each of the 656,153 claims in our dataset must impact at least one of four coverages offered to the policyholders to be activated in the systems of the insurer. We provide a brief definition of these four coverages in Table 2. To better understand the dynamics of our portfolio, Table 3 shows the importance of each coverage in terms of the proportion of claims and proportion of the total cost. Unsurprisingly, the Vehicle Damage coverage is the one that most often comes into play with 96.39% of the 656,153 claims activating it. A bit more than half of the claims impact the Loss of Use coverage. Only 5.70% of the claims in the portfolio

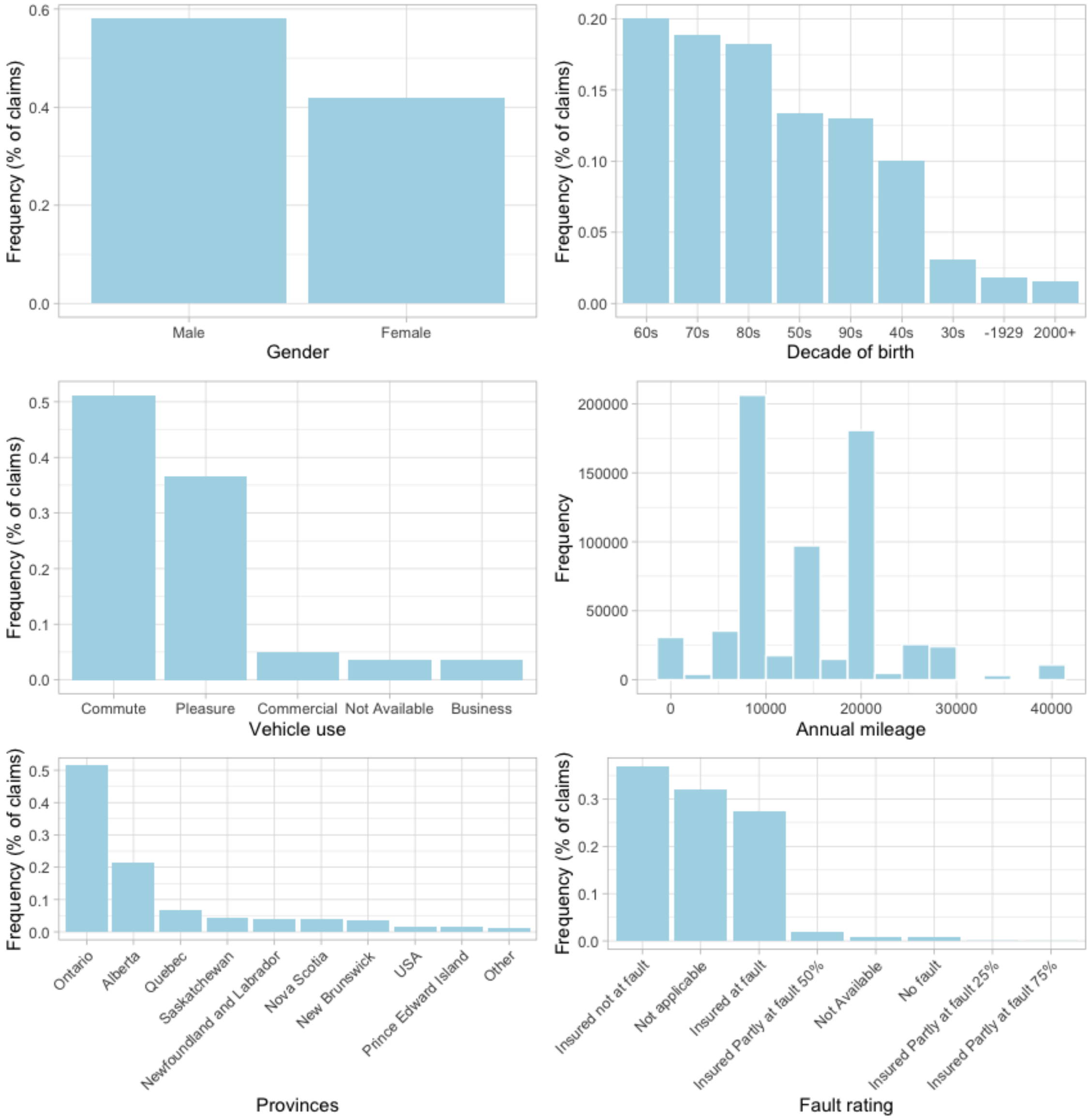

Figure 7: Risk factors.

activate the Bodily Injury coverage. Similarly and as expected, less than 10% of the total claims impact the Accident Benefits coverage.

The proportions are slightly different when we look at the repartition of the total portfolio cost over the four coverages. In particular, Loss of Use is the coverage that weights the less in terms of cost, representing less than 4% of the total cost of the portfolio. Despite being activated by less than 6% of all claims, the Bodily Injury coverage still represents around 13% of the total portfolio cost. Table 4 shows some descriptive statistics for each of the four coverages. Bodily Injury claims are those for which the average payments are the biggest, followed by the Accident Benefits coverage. The Loss of Use coverage presents by far the lowest payments as was already hinted by Table 3. We also notice that the payments for the Accident Benefits, Bodily Injury and Vehicle Damage coverages present large values in the higher quantiles, indicating that their corresponding distributions are probably heavy-tailed.

Table 2: Description of the four insurance coverages provided within the policy

Insurance coverages	
Accident Benefits	Compensation for loss of revenue, funeral expenses, medical expenses, death,...
Bodily Injury	Compensation for medical expenses or loss of revenue of a third party.
Vehicle Damage	Compensation for the damages incurred to the insured or another party's vehicle.
Loss of Use	Compensation of costs in case of the temporary replacement of a vehicle or any other alternative transportation means used during vehicle repairs.

Table 3: Weight of each coverage in the portfolio

Coverage	% of claims	% of the total cost
Accident Benefits	9.42	12.82
Bodily Injury	5.70	13.13
Vehicle Damage	96.39	70.44
Loss of Use	51.89	3.61

3.1.3 Activation and payment delays

In Table 5, we take a look at the average activation delays per coverage, i.e. the average delays between the reporting of a claim and the activation of the coverages in the insurer's systems.

The Accident Benefits, Vehicle Damage and Loss of Use coverages are typically activated without delay for the vast majority of the claims. For the Bodily Injury coverage however, around 10% of the claims only activate it with a delay of at least one year after the reporting date.

Example 3. To illustrate this, consider the case where a policyholder is at fault in an accident with a third party. The policyholder reports the claim to the insurer who, considering the reparations needed to the vehicle, records it as a Vehicle Damage claim. The following year, the third party realises that he has been injured in the accident and demands a compensation. The insurer thereby needs to record the claim as a Bodily Injury as well as Vehicle Damage claim. This illustrates how in some cases, some coverages become active at a later time during the development of the claim rather than directly upon its reporting to the insurance company.

Appendix 4 further illustrates the dynamics between the different coverages by presenting the activation patterns observed in the dataset for the first development year of the claims.

Once a claim activates a coverage, we are interested in knowing when the first payment will take place. The payment delays refer to the time elapsed between the reporting date of a claim and the date at which the insurer records a first payment

Table 4: Descriptive statistics for the four insurance coverages

Coverage	Mean	Std. dev.	Quantiles				Max.
			0.5	0.75	0.95	0.99	
Accident Benefits	12,386	53,561	3,215	6,909	47,757	127,896	2,435,334
Bodily Injury	23,271	76,027	4,000	15,150	98,449	322,612	2,039,570
Vehicle Damage	5,040	8,121	2,605	5,830	17,984	40,611	149,399
Loss of Use	545	620	419	714	1,000	2,336	52,777

Table 5: Percentage of claims with different activation delays for the four coverages

	Activation delays		
Coverage	No delay	1 year	≥2 years
Accident Benefits	96.65	3.18	0.11
Bodily Injury	89.89	7.70	2.11
Vehicle Damage	98.20	1.73	0.07
Loss of Use	97.80	2.17	0.03

for the coverage(s) that the claim activated. Thereby, we do not assume that the payments directly follow the activation of the coverages. This is clearly shown in our data and illustrated in Figure 8. The Vehicle Damage coverage is the one for which payments happen the fastest with a payment taking place in the same year as the reporting year for 77% of the claims that activate that coverage. Payments come directly in the first development year for 66% of the claims that activate the Loss of Use coverage but only for 50% of the claims that activate the Accident Benefits coverage and 27% of the claims for the Bodily Injury coverage. For the rest of the claims, almost all remaining payments take place in the second year following the reporting date for the Vehicle Damage and Loss of Use coverages, such that when we reach the third development year, these coverages practically never incur additional payments. For the other two coverages however, the claimant might have to wait a bit longer to receive his payments. When we reach the fourth development year, only 63% of the claims have received their first payment for the Accident Benefits coverage against only 46% for the Bodily Injury coverage. It is only starting from the fifth development year that almost all claims have received their first payment regardless of the coverage impacted.

Example 3 (continued). Consider again our earlier example. Even though claim activated the Bodily Injury coverage in the second year after reporting, the third party that was injured might still wait a little to receive his payment while the insurance company consults with medical or legal experts to determine the final amount that the claimant is owed.

3.2 Estimation

We apply our model to the data introduced in Section 3.1. We present the results of the model fitting performed for the three components of the model, namely the activation patterns, payment patterns and severities. Since with work with automobile

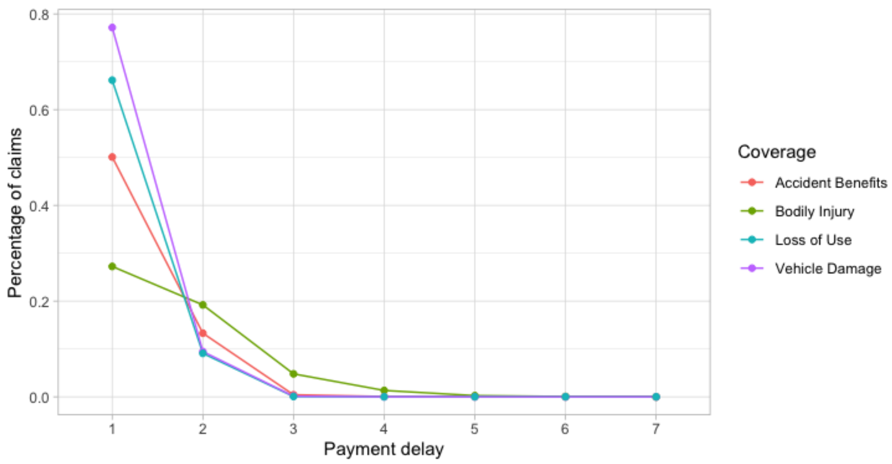

Figure 8: Payment delays per coverage

claims and given the limited number of development years available in our dataset, we work with the activation patterns and payment vectors for development years $j = 1$ and $j = 2$. For the claims that are already in at least their third development year at the evaluation of the reserves, we apply the methodology laid out in Section 2.3.

3.2.1 Activation patterns

Using the same notation as in Section 2, we have $C = 4$ insurance coverages, leading to $V = 15$ possible activation patterns. We fit the multinomial logit model for the activation patterns using maximum likelihood estimation with the likelihood function given by

$$\mathcal{L}(\boldsymbol{A}_{i,j}|\boldsymbol{x}_i\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_{j,v}\right)}{\sum_{k=1}^{V} \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}_{j,k}\right)}. \tag{9}$$

In Appendix 4, we provide the full list of parameters estimated by maximum likelihood for each of the $V = 15$ possible activation patterns.

We observe that the impact of some risk factors becomes very important for the 5$^{\text{th}}$ pattern, i.e. the one in which a claim simultaneously activates the Bodily Injury and Loss of Use coverages. In particular, the probability of observing that specific activation pattern increases quite substantially if the claim occurred in the USA. However, if the claim occurred in Quebec or in Saskatchewan, the probability of activating both the Bodily Injury and the Loss of Use coverages greatly decreases. This does not come as a surprise due to the legislation regarding automobile insurance in these two states. In Quebec, the Bodily Injury coverage does not exist: if an individual is injured in a car accident, he is covered by the public automobile insurance plan provided by the Société de l'assurance automobile du Quebec (SAAQ) rather than by a private insurance company. In Saskatchewan, the insured must choose between a "no fault" and a "tort" auto injury insurance coverage offered by the Saskatchewan driver's licensing and vehicle registration. Almost all residents choose the "tort" coverage under which they are insured regardless of whether they are at fault or not. As such, the Bodily Injury coverage very rarely appears for claims that occur in that province. Consequently, we see that the parameter estimates for Quebec and Saskatchewan are always quite negative for all the activation patterns that include the Bodily Injury coverage.

Moreover, when the fault rating is either not available or not applicable or when the degree of fault of the insured is evaluated at 25%, the probability of the claim activating both the Bodily Injury and Loss of Use coverages decreases. Once again, since the provincial legislation often impacts the way Bodily Injury claims are handled, we know that for many of those claims, insurers will not record a fault rating or deem it inapplicable to the specific situation. We can thus expect for example that when the fault rating is not applicable for a claim, this claim has occurred in Quebec where the Bodily Injury coverage does not apply and hence, the probability of activating it decreases.

Finally, the year of birth ia also a big driver for this particular pattern with the probability of observing it increasing for the younger insured individuals born in the 1990s and after 2000. It is commonly observed in automobile insurance data that younger drivers tend to cause more frequent and serious accidents. As a consequence, we can expect to observe more accidents with injuries or even casualties and loss of use of a vehicle in the younger population. Appendix 4 illustrates this well, with the insureds born after 2000 being the only cohort that always has an increasing impact on the probability of observing each activation pattern, even though they only represent 1.52% of all the insureds.

3.2.2 Payment patterns

As described in Section 2, the payment patterns help actuaries determine whether or not a payment took place for an active coverage c of claim i in development year j i.e. in the case where $A_{i,j,c} = 1$. We fit in total eight Bernoulli regressions: one for each of the four coverages in years $j = 1$ and $j = 2+$ and use again maximum likelihood optimization to obtain the parameter estimates for each of the models. Table 6 displays the results of these estimations. When comparing them with the payment delays observed in Figure 8, we see that the models provide overall a good fit. The Vehicle Damage coverage is the one for which payments come the fastest with around 80% of the claims receiving a payment in their reporting year, i.e. in development year $j = 1$. The Loss of Use coverage comes in second place with an estimated 70% of claims receiving a payment in the first year against around 65% of claims in the observed data. The estimated percentage $\pi_{j,\mathrm{AB}}$ for the Accident Benefits is very similar to the observed value with a payment happening in $j = 1$ for almost half of the claims. Finally, while we observe that a bit less than 30% of the claims receive a payment in the first year for the Bodily Injury coverage, the estimation obtained from our model is around 35%.

Table 6: Fitted average probabilities of observing a payment

Probability	$j = 1$	$j = 2+$
$\pi_{j,\mathrm{AB}}$	0.5155	0.3578
$\pi_{j,\mathrm{BI}}$	0.3484	0.2812
$\pi_{j,\mathrm{VD}}$	0.8203	0.1138
$\pi_{j,\mathrm{LoU}}$	0.7070	0.0720

3.2.3 Payment severities

From the observations in Table 4, we note the importance of considering long-tailed distributions to model the severity of payments, as is usually done in the actuarial literature. Among others, Frees et al. [2009] opt for the Generalized Beta of the second kind distribution to accommodate the long-tail nature of claims. Figure 9 shows the histogram of the payments made for the four coverages. Note that for the Accident Benefits and Bodily Injury coverages, we cut out most of the tail of the histograms in the right part of the graph to ease the readability. These graphs further illustrate the need on the one hand to consider long-tailed distributions for the coverages and on the other hand the benefits that we can earn from selecting different distributions for the coverages rather than a single distribution common to all. We are thereby able to tailor the model to the specific distribution profiles of the different types of claims.

For each coverage, we consider five distributions commonly used. We fit the following severity models for both years $j = 1$ and $j = 2+$: Log-Normal, Gamma, Pareto, Generalized Beta of the second kind and Weibull distributions. Appendix 4 presents the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for all the models for both time periods considered and all four insurance coverages. For the first development year, the Generalized Beta of the second kind seems to be the most appropriate distributional choice for Accident Benefits, Vehicle Damage and Loss of Use while the Pareto distribution is more suited to model the payments incurred in the Bodily Injury coverage. For the other development years, the Generalized Beta of the second kind is again the distribution that presents the lowest AIC for Accident Benefits, Vehicle Damage and Loss of Use while we prefer the Weibull distribution to model Bodily Injury claims. This illustrates the advantages of separating the severity modelling of the different coverages. From the model fitting, we see that the distribution profile of the Bodily Injury coverage is different than that of the others and varies with the development periods.

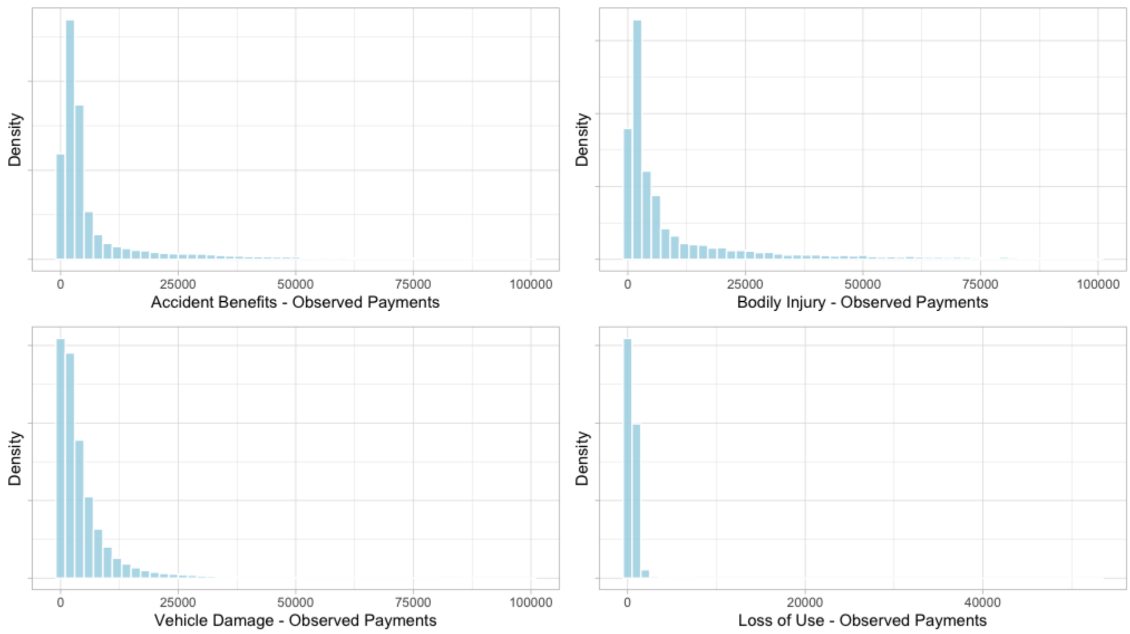

Figure 9: Histograms of the observed payments per coverage

3.3 Predictive distributions and model comparisons

With the data presented in Section 3.1 and the models fitted in Section 3.2, we can now focus in this section on presenting the results obtained after performing the evaluation of the reserves for our portfolio of automobile insurance claims, choosing the 1^{st} of January 2019 as our evaluation date. We present the predictive distributions for the reserve of the whole portfolio and the four coverages separately. We compare these results to the true reserve and to that obtained from fitting some more classical and commonly used reserving models.

For all the models that we will consider in this section, we perform 5,000 simulations using the simulation routine described in Section 2.4.1. As shown in Appendix 4, this is a sufficiently large number to bring stability to our results. We first present the estimations from the activation patterns model before comparing them to other reserving models.

Figure 10 displays the predictive distribution for the total reserves in our portfolio, i.e. for both the IBNR and RBNS claims on the left and the total reserves only for the RBNS claims on the right. In both cases and in all graphs that will follow, the red line marks the true reserve, the blue dotted line shows the average value of the predictions and the continuous blue line depicts the 95^{th} quantile of the distribution. For all claims, the true reserve amounts to 542.55M CAD, whereas our model produces an average predicted amount of 561.62M CAD. For the RBNS claims only, the true reserve is equal to 524.91M CAD and the simulations give 536.86M CAD. Note that the true reserve amounts that we provide here are minimum amounts since at the evaluation date, a bit more than 3% of the claims are still open in the portfolio. To handle them, we chose to consider the date of the last observed payment as the settlement date. Thereby, we must be aware that the true severity of these claims will in all likelihood be greater than the one observed.

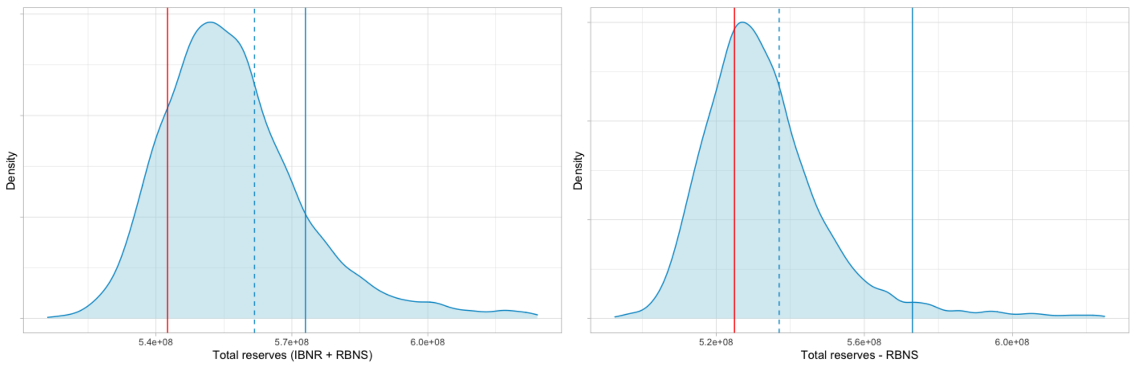

Figure 10: Simulated reserves as of the 1st of January 2019. The red line, dashed blue line and continuous blue lines depict, respectively, the true reserve amount, the average and the 95^{th} quantile of the simulations

With the IBNR claims representing only 3.25% of the total final cost in our evaluation dataset, we choose to focus on the results of the RBNS claims only. Figure 11 presents the predictive distributions obtained for the reserves of the four insurance coverages. As in Figure 10, the red lines show the true reserve amounts while the dotted and continuous blue lines represent, respectively, the mean and the 95^{th} quantile of the distributions. In Table 7, we provide further details on these results. Since financial legislation around the world typically require insurers to set aside an amount based on a high quantile of the predictive distribution of their reserves, we provide the 95% Value-at-Risk (VaR) in addition to the mean for each of the four coverages and for the portfolio as a whole.

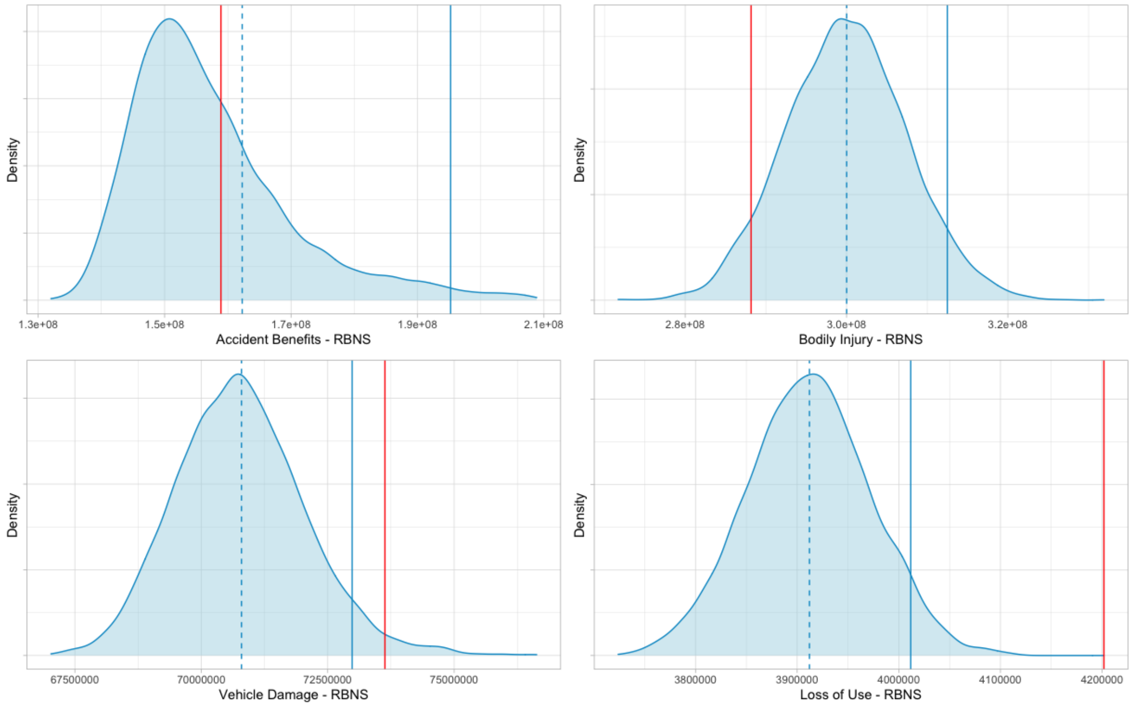

Figure 11: Simulated reserves per insurance coverage (RBNS). The red lines, dashed blue lines and continuous blue lines depict, respectively, the true reserve amounts, the average and the 95^{th} quantiles of the simulations

Table 7: Simulated RBNS reserves (in M CAD).

Coverage	True reserves	Activation Patterns	
		Mean	$\text{VaR}_{0.95}$
Accident Benefits	>158.90	162.15	196.64
Bodily Injury	>288.17	300.02	312.58
Vehicle Damage	>73.63	70.80	72.95
Loss of Use	>4.20	3.91	4.01
Total	>524.91	536.86	574.58

For the full portfolio, the activation patterns model overestimates the true reserves, providing rather conservative estimates. This is also the case for the Accident Benefits and Bodily Injury coverages. For the Vehicle Damage and Loss of Use coverages that represent together around 15% of the total RBNS reserve, the predictions slightly underestimate the true amounts.

In Sections 3.3.1, 3.3.2 and 3.3.3, we pursue the analysis of our results by comparing them to those obtained using, respectively, an aggregate model, a replica of the activation patterns model but without taking dependence into account and the individual reserving model proposed by Antonio and Plat [2014].

3.3.1 Activation patterns model vs Aggregate models

Table 8 compares the reserve estimates obtained with the proposed model based on the activation patterns to those of the classical Overdispersed Poisson (ODP) Chain Ladder and to the true reserves, both for the portfolio as a whole and for the four insurance coverages taken separately.

On average, the aggregate model underestimates the true reserves in each case. We need to take the 95% VaR to obtain an estimate of the whole RBNS reserve that is above the minimum observed value of 524.91M CAD. Considering that we expect this amount to be higher due to the open claims, our model gives a more conservative estimate of the total RBNS reserve with an average simulated amount of 536,86M CAD. Most importantly, the aggregate model does not perform well at the coverage level. In each case and particularly for the Accident Benefits and Bodily Injury coverages, the predictions underestimate the true reserve amounts. Using such a model, the insurer is thus blind to the dynamics of his portfolio and can only get an idea of what his total reserve amounts to.

In addition, since our data spans from 2015 to mid-2021, we only have a limited number of development years and thereby rather small claim triangles to perform the analysis with aggregate models. To obtain more appropriate estimates in this specific case, the insurer should consider making use of tail factors.

Table 8: Comparison between the activation patterns model and the Overdispersed Poisson Chain Ladder (in M CAD)

Coverage	True reserve	Activation patterns		ODP	
		Mean	$\text{VaR}_{0.95}$	Mean	$\text{VaR}_{0.95}$
Accident Benefits	>158.90	162.15	196.64	71.01	81.66
Bodily Injury	>288.17	300.02	312.58	177.64	204.63
Vehicle Damage	>73.63	70.80	72.95	40.87	48.43
Loss of Use	>4.20	3.91	4.01	3.02	3.88
Total	>524.91	536.86	574.58	493.14	537.16

3.3.2 Activation patterns model vs independence model

To strengthen our argument in favour of a model that takes into account the possible dependence between the insurance coverages, we perform a series of independence tests on our data. First, we test for independence between each possible pair of insurance coverages using the chi-squared goodness of fit test. Each of these two-way interaction tests clearly rejects the hypothesis of independence. We also perform likelihood ratio tests where we compare models with no interaction, two, three and four-way interactions. Once again, the tests clearly show the presence of dependence between the coverages and reject the simpler models in favour of the more complex ones in each scenario.

To assess the impact of modelling this dependence on the reserve estimates, we build an independence model. We reproduce the activation patterns model but rather than modelling the activation patterns $\boldsymbol{A}_{i,j}$ and $\boldsymbol{A}_{i,j+1}$ with the multinomial logit model described in Section 2, we model the activation of the coverages using four separate and independent Bernoulli regressions for development periods $j = 1$ and $j = 2+$. We present the results obtained with this model in Table 9 and compare them to the true reserves and the estimates from our activation patterns model.

With an average estimated amount and average 95% VaR of, respectively, 495.32M CAD and 526.93M CAD, the independence model clearly underestimates the true reserves. The predictions are also less accurate for each of the four coverages and the activation patterns model outperforms the independence model in each case. Leaving the dependence between the coverages out of the modelling process can thereby lead to underestimated reserves which can put the insurer at risk of financial sanctions.

Table 9: Comparison between the activation patterns model and the independence model (in M CAD)

Coverage	True reserve	Activation patterns		Independence	
		Mean	$\text{VaR}_{0.95}$	Mean	$\text{VaR}_{0.95}$
Accident Benefits	>158.90	162.15	196.64	143.41	173.50
Bodily Injury	>288.17	300.02	312.58	278.18	290.31
Vehicle Damage	>73.63	70.80	72.95	69.94	72.07
Loss of Use	>4.20	3.41	4.01	3.78	3.89
Total	>524.91	536.86	574.58	495.32	526.93

3.3.3 Activation patterns model vs another individual model

Next to the aggregate and independence models, we compare the model based on the activation patterns to the individual reserving model first introduced by K. Antonio and R. Plat in 2014. Working in continuous time, the authors use a hierarchical process to estimate the reserves for the RBNS claims. The interested reader can find more details on their model in Antonio and Plat [2014].

We focus here solely on the Accident Benefits coverage. Figure 12 shows the predictive distributions for the RBNS reserves of this coverage obtained with the activation patterns model (in blue) and Antonio and Plat [2014]'s individual model (in grey). As before, the red line shows the true amount of the reserves equal to 158,90M CAD while the dotted and continuous lines represent, respectively, the average reserve estimates and 95% VaR obtained with each model. For the activation patterns model and Antonio and Plat [2014]'s model, the average predictions are, respectively, 162,15M CAD and 143,46M CAD while the 95% VaR amount to 196,64M CAD and 168,30M CAD. The predictive distribution of the activation patterns model is thereby closer on average to the true reserve. However, Antonio and Plat [2014]'s model gets a closer estimate in terms of the VaR whereas in this case, our model overestimates the reserve.

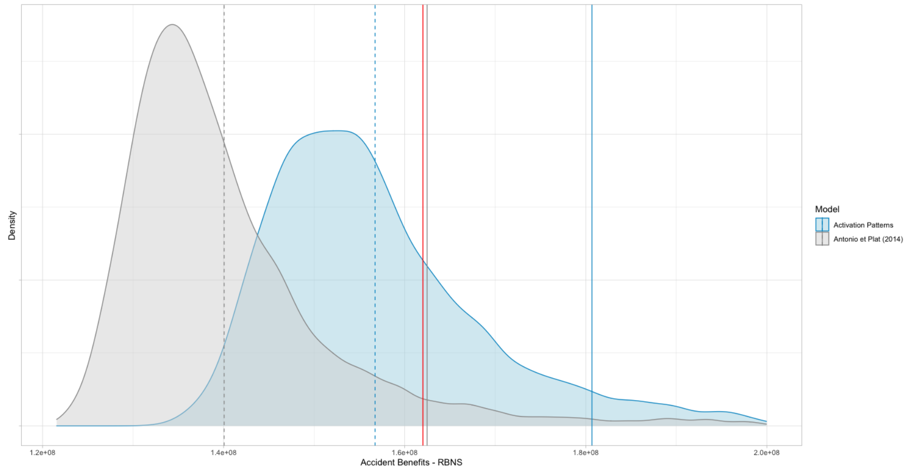

Figure 12: Predictive distributions of the RBNS reserves obtained with the activation patterns model and the model of Antonio and Plat Antonio and Plat [2014]. The red line shows the true reserve amount and the blue and grey dotted and continuous lines depict the average predictions and 95% VaR obtained with, respectively, Antonio and Plat Antonio and Plat [2014]'s model and the model based on the activation patterns

3.3.4 Results summary

Table 10 summarizes the results discussed in Sections 3.3.1 and 3.3.2. Of the four different models, the independence model is the one that underestimates the total reserve the most in terms of the Value-at-Risk.

The aggregate model presents the lowest average estimates but its predictive distribution is more heavy-tailed than that of the independence model and the 95% VaR of 537,16M CAD seems to be a more appropriate estimate. However, this model performs very poorly when we take the coverages separately and consistently underestimates the true reserves. It is thereby not fitted to gain a better understanding of the dynamics of the portfolio. In addition and as mentioned in Section 3.3.1, the limited amount of development years available forces the insurer to use tail factors, making the model less reliable or apt for predictions.

The model based on the activation patterns provides a greater level of accuracy in the estimation of the total reserves both on average and in the higher quantiles of the predictive distributions. Even though it is the model that overestimates the true reserve the most, especially in terms of the VaRwhere we could prefer the total estimate of the aggregate model, it provides more accurate predictive distributions at the level of the different insurance coverages.

4 Conclusion

In this paper, we introduce a model based on activation patterns for the insurance coverages to estimate the claims reserves on an individual basis while taking into account the dependence between the different coverages provided by an insurer. More specifically, we analyze the way in which a single claim can simultaneously activate multiple coverages. Based on the predictions of the so-called activation patterns of the coverages in year j, we then predict whether an active coverage

Table 10: Model comparison - summary (in M CAD)

Coverage	Observed	Simul.	Act. pat.	ODP	Ind.	[1]
Accident Benefits	>158.90	Mean	162.15	71.01	143.41	143.46
		$\text{VaR}_{0.95\%}$	196.64	81.66	173.50	169.30
Bodily Injury	>288.17	Mean	300.02	177.64	278.18	-
		$\text{VaR}_{0.95\%}$	312.58	204.63	290.31	-
Vehicle Damage	>73.63	Mean	70.80	40.87	69.94	-
		$\text{VaR}_{0.95\%}$	72.95	48.43	72.07	-
Loss of Use	>4.20	Mean	3.91	3.02	3.78	-
		$\text{VaR}_{0.95\%}$	4.01	3.88	3.89	-
Total	>524.91	Mean	536.86	493.14	495.32	-
		$\text{VaR}_{0.95\%}$	574.58	537.16	526.93	-

will incur a payment and its corresponding amount. Using a Canadian automobile dataset, we apply the model and compare the results obtained to those of three additional models: the classical Overdispersed Poisson Chain Ladder, a replica of the activation patterns models that does not take the dependence into account and Antonio and Plat [2014]'s individual reserving model. We observe that when taking the dependence between the coverages into account, we reach on average more accurate estimates of the reserves than the other models. The predictive distribution that we obtain is however more heavy-tailed than the others, leading to larger Value-at-Risk. We can however consider these as more conservative compared to other models for which the estimates are sometimes too close to the observed amount. The main contribution of our model resides in its capacity to predict the specific reserve estimates for each insurance coverages more accurately, thereby allowing the insurer to better understand the dynamics of his portfolio. This is particularly interesting considering some regulatory frameworks such as Solvency II that requires a detailed analysis of the reserves on a line of business basis and that consistently advocates for the use of more prudent and conservative modelling techniques.

Appendix A: Observed activation patterns

Table 11 presents the activation patterns $\boldsymbol{A}_{i,1}$ observed in the dataset for the first development year of the claims. The value 1 stands for the activation of the coverage. 85.67% of all claims activate the Vehicle Damage coverage either alone or simultaneously with the Loss of Use coverage. The Accident Benefits coverage is most often activated together with the Vehicle Damage coverage or on its own. We also observe that around 1% of all claims simultaneously activate all four coverages upon their reporting.

Table 11: Frequency of the activation patterns observed in the first development year

Accident Benefits	Bodily Injury	Vehicle Damage	Loss of Use	% of claims
0	0	1	0	42.95
0	0	1	1	42.72
1	0	1	0	4.98
0	1	1	1	1.99
1	0	0	0	1.44
0	1	1	0	1.24
1	0	1	0	1.07
0	0	0	1	1.04
1	1	1	1	1.01
0	1	0	0	0.61
1	1	1	0	0.43
1	1	0	0	0.38
1	0	0	1	0.10
0	1	0	1	0.04
1	1	0	1	0.01

Appendix B: Multinomial logit model

Table 12: Parameter estimates for the multinomial logit model.

	Risk factors	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Intercept		3.45	3.44	-0.26	-12.78	1.31	1.93	0.30	-1.04	1.14	1.67	0.28	-1.68	0.56	1.69
PROV	New Brunswick	0.21	-0.16	1.11	0.43	0.05	-0.49	-0.93	0.29	-0.18	0.35	0.48	1.07	0.40	0.35
	Newfoundland and Labrador	0.55	0.02	0.58	0.63	0.09	0.27	-0.06	-0.16	-0.09	0.26	0.79	0.83	0.61	0.65
	Nova Scotia	0.36	-0.05	1.09	0.18	-0.08	-0.72	-1.08	0.23	-0.73	0.30	0.48	0.01	-0.01	0.29
	Ontario	-0.14	-0.12	0.93	0.25	-0.35	-0.32	-0.50	-0.15	-1.26	-0.34	0.15	0.41	-0.65	-0.31
	Other	0.57	-0.14	0.61	-0.62	-0.06	-0.08	0.25	0.10	0.34	-0.28	0.76	1.87	0.71	0.29
	Prince Edward Island	0.93	-0.13	0.93	0.98	0.17	-0.75	-0.57	0.31	0.01	0.31	0.58	-3.21	1.11	0.61
	Quebec	1.64	-0.11	-2.10	-19.74	-2.77	-3.31	-2.34	-2.10	-1.97	-2.01	-1.66	-0.22	-2.46	-2.34
	Saskatchewan	1.53	-0.40	-1.21	-23.02	-1.04	-2.84	-1.05	-0.78	-0.41	-1.27	-2.73	-390.59	-0.60	-2.37
	USA	-5.52	-1.50	-3.83	11.02	-5.16	-2.24	-7.22	0.32	-5.19	-0.99	-3.53	0.10	-3.45	-0.86
YOB	2000+	0.71	0.70	3.71	14.79	3.87	3.80	3.90	2.92	3.61	3.48	3.94	3.49	3.84	3.62
	1930s	-0.34	-0.21	-1.13	9.95	-1.09	-0.84	-0.96	-1.10	-0.72	-1.0	-1.47	-1.95	-1.07	-1.04
	1940s	-0.41	-0.26	-0.72	10.46	-0.49	-0.56	-1.39	-1.82	-1.16	-1.08	-1.90	-2.24	-1.68	-1.64
	1950s	-0.55	-0.27	-0.82	10.36	-0.56	-0.73	-0.84	-1.57	-0.69	-0.82	-1.54	-1.81	-1.36	-1.29
	1960s	-0.55	-0.29	-0.58	10.44	-0.33	-0.44	-0.81	-1.31	-0.82	-0.88	-1.79	-1.36	-1.57	-1.60
	1970s	-0.60	-0.26	-0.54	10.37	-0.30	-0.37	-0.75	-1.32	-0.74	-0.81	-1.88	-1.43	-1.52	-1.51
	1980s	-0.73	-0.33	-0.48	10.18	-0.29	-0.40	-0.69	-1.14	-0.78	-0.71	-1.62	-1.18	-1.27	-1.26
	1990s	-0.44	-0.30	-0.30	11.04	-0.08	-0.12	0.15	-1.20	-0.07	-0.48	-0.50	-0.50	-0.39	-0.63
GENDER	Male	0.24	0.08	-0.22	-0.44	-0.23	-0.28	-0.26	-0.89	-0.35	-0.89	-0.28	-0.69	-0.45	-0.92
AM		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VU	Commercial	0.68	-0.11	0.83	0.25	0.71	-0.19	0.28	-0.299	0.89	-0.19	0.71	-0.55	0.84	-0.27
	Commute	-0.13	-0.00	-0.08	-0.20	-0.05	-0.14	0.03	-0.11	0.15	0.09	0.07	-0.34	0.17	-0.02
	Not Available	2.37	0.13	2.73	-0.07	2.59	0.28	2.25	-1.88	1.44	-1.13	0.11	-3.26	1.28	-1.50
	Pleasure	0.15	-0.03	-0.00	-0.15	0.05	-0.21	0.25	-0.22	0.57	0.04	0.34	-0.46	0.57	-0.06
FR	Insured not at fault	-0.94	-0.06	-2.53	-2.32	-3.38	-2.70	0.55	0.84	-0.23	0.72	-1.28	-1.39	-1.82	-1.35
	Insured Partly at fault 25%	1.10	1.37	2.96	-12.13	1.62	2.13	0.89	-9.41	2.23	2.10	3.51	-2.41	3.43	2.98
	Insured Partly at fault 50%	0.34	0.56	-0.01	-0.42	-0.12	-0.01	-0.92	-0.38	0.18	0.71	-0.07	0.16	0.17	0.50
	Insured Partly at fault 75%	-0.55	-0.27	0.61	-8.55	0.88	0.34	0.93	-6.36	0.61	0.67	2.64	-3.71	0.80	1.49
	No fault	4.30	3.23	-1.19	-5.06	-1.25	0.05	6.37	1.61	2.30	4.04	1.29	-2.13	-0.83	0.87
	Not applicable	1.46	-0.07	-1.88	-28.36	-3.83	-4.53	1.05	-1.28	-1.12	-1.36	-0.77	-2.98	-2.35	-3.16
	Not Available	5.81	1.34	3.74	-12.24	4.27	1.26	7.88	0.74	4.41	0.99	4.16	-0.43	3.32	1.06

Appendix C: Choice of the severity models

Table 13: Choice of the distributions for the severity of the payments

Coverage	Model	$j = 1$ AIC	$j = 1$ BIC	$j = 2+$ AIC	$j = 2+$ BIC
Accident Benefits	Log-Normal	266,408	266,651	181,577	181,827
	Gamma	264,411	264,654	183,670	183,921
	Pareto	262,903	263,146	181,281	181,532
	Generalized Beta II	**261,110**	**261,368**	**181,222**	**181,487**
	Weibull	264,101	264,344	182,545	182,796
Bodily Injury	Log-Normal	100,523	100,735	106,985	107,214
	Gamma	99,675	99,887	107,096	107,325
	Pareto	**98,971**	**99,183**	106,691	106,919
	Generalized Beta II	99,966	100,185	106,884	107,119
	Weibull	99,187	99,400	**106,673**	**106,902**
Vehicle Damage	Log-Normal	5,375,907	5,376,246	620,136	620,431
	Gamma	5,385,180	5,385,519	619,456	619,751
	Pareto	5,342,022	5,342,361	617,247	**617,542**
	Generalized Beta II	**5,330,989**	**5,331,349**	**617,245**	617,557
	Weibull	5,371,787	5,372,126	618,531	618,827
Loss of Use	Log-Normal	1,874,185	1,874,498	193,705	193,967
	Gamma	1,868,867	1,869,180	192,929	193,191
	Pareto	1,894,927	1,895,241	194,376	194,638
	Generalized Beta II	**1,861,585**	**1,861,918**	**192,212**	**192,489**
	Weibull	1,877,508	1,877,821	193,459	193,721

Appendix D: Results stability

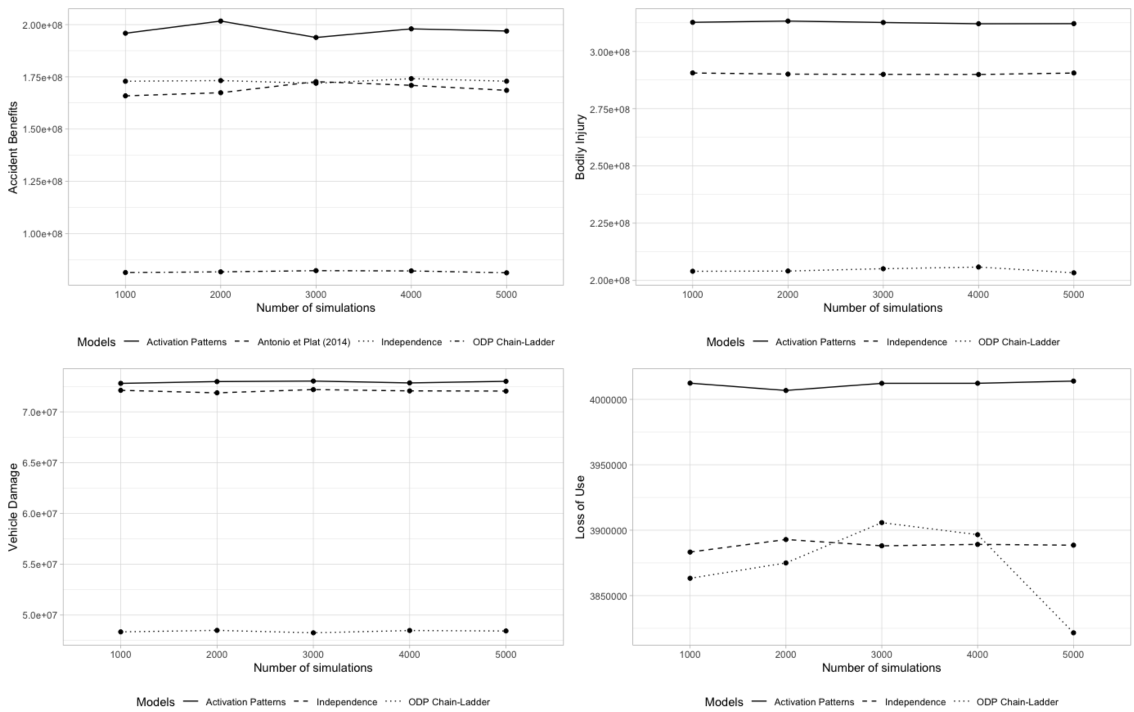

Figure 13: Results of the RBNS reserves (95% VaR) based on the number of simulations performed

References

K. Antonio and R. Plat. Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014(7):649–669, 2014.

L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth Statistics/Probability Series, New York, 1984.

H. Buhlmann, R. Schnieper, and E. Straub. Claims reserves in casualty insurance based on a probabilistic model. *Bulletin of the Association of Swiss Actuaries*, 80:21–45, 1980.

M.-P. Côté, C. Genest, and D. Stephens. A bayesian approach to modeling multivariate multilevel insurance claims in the presence of unsettled claims. *Bayesian Anal.*, 17(1):67–93, 2022.

E. Frees and E. Valdez. Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457–1469, 2008.

E. Frees, P. Shi, and E. Valdez. Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin*, 39(1):165–197, 2009.

S. Haastrup and E. Arjas. Claims reserving in continuous time: a non-parametric bayesian approach. *ASTIN Bulletin*, 26(2):139–164, 1996.

C. Hachemeister. A stochastic model for loss reserving. *Transactions of the 21st International Congress of Actuaries*, 1:185–194, 1980.

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

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

T. Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, 23(2):213–225, 1993.

T. Mack. Which stochastic model is underlying the chain ladder model? *Insurance: Mathematics and Economics*, 15:133–138, 1994.

T. Mack. The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor. *ASTIN Bulletin*, 29(3):361–366, 1999.

T. Mack and G. Venter. A comparison of stochastic models that reproduce chain ladder reserve estimates. *Insurance: Mathematics and Economics*, 26(1):101–107, 2000.

R. Norberg. A contribution to modeling of ibnr claims. *Scandinavian Actuarial Journal*, 1986:155–203, 1986.

R. Norberg. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin*, 23(1):95–115, 1993.

R. Norberg. Prediction of outstanding liabilities ii. model variations and extensions. *ASTIN Bulletin*, 29(1):5–25, 1999.

M. Pešta and O. Okhrin. Conditional least squares and copulae in claims reserving for a single line of business. *Insurance: Mathematics and Economics*, 56:28–37, 2014.

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

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

M. Stasinopoulos, R. Ridgby, G. Heller, V. Voudouris, and D. De Bastinani. *Flexible regression and smoothing: using gamlss in R*. Chapman and Hall/CRC, New York, 2017.

S. van Buuren and K. Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011.

M. V. Wüthrich. Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, 6:465–480, 2018.

X. Zhou and X. Zhao. Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, 46(2):10, 2010.