Journal of Industrial and Management Optimization

https://www.aimspress.com/journal/jimo

JIMO, 22(6): 2647–2682.
DOI:10.3934/jimo.2026097
Received: 19 December 2025
Revised: 08 April 2026
Accepted: 13 April 2026
Published: 13 May 2026

Research article

mCube: A multinomial micro-level reserving model

Emmanuel Jordy Menvouta[1], **Robin Van Oirbeek**[2,*] **and Tim Verdonck**[1,3]

[1] KU Leuven, Department of Mathematics, Leuven, Belgium
[2] Allianz Benelux, Boulevard Roi Albert II 32, Brussels, Belgium
[3] University of Antwerp – imec, Department of Mathematics, Antwerp, Belgium

* **Correspondence:** Email: robin.vanoirbeek@gmail.com.

Abstract: This paper presents a multinomial multi-state micro-level reserving model, denoted mCube. We propose a unified framework for modelling the time and the payment process for incurred but not reported (IBNR) and reported but not settled (RBNS) claims and for modeling IBNR claim counts. mCube is a hybrid reserving framework: IBNR counts are estimated at the macro level, while post-reporting claim development is modelled at the micro level. We use multinomial distributions for the time process and spliced mixture models for the payment process. We illustrate the strong performance of the proposed model on a real data set of a major insurance company consisting of bodily injury claims. In our application, the proposed model produces the best estimate distribution. In the case study on a real insurance portfolio, this distribution is approximately centered around the true reserve. A simulation study on synthetic data confirms meaningful claim-level predictive accuracy across multiple scenarios.

Keywords: individual claims; IBNR; RBNS; micro-level reserving; multinomial model
Mathematics Subject Classification: 62P05

1. Introduction

A central part of an insurance company is the management of its future cash flows and solvency capital. To this end, insurers have to set aside reserves to cover outstanding claim liabilities. After an insured event has occurred, it always takes some time to settle the final payment of a claim. Taking into account two different sources of delay in general insurance, insurers typically set aside separate reserves for incurred but not reported claims (IBNR) and reported but not settled claims (RBNS). The number of RBNS claims is known, together with specific information relative to each claim. To determine the reserve for RBNS claims, the aim is to predict the amount that still needs to be paid. For IBNR claims, the insurer does not have information about each specific claim or the total number of these claims.

The goal when modelling IBNR reserves is first to estimate the number of these claims, and second, to estimate the cost for each claim.

New regulations such as the Solvency II and international financial reporting standards (IFRS) frameworks guide insurance companies towards best practices for the calculation of their reserves. Following these regulations, it is important that models for claims reserving not only accurately predict the ultimate reserve amount but also the distribution of future cash-flows conditional on currently available information. More information on the calculation of insurance reserves can be found in [1].

A first class of models developed for the task of claims reserving are collective or macro-level models that focus on aggregated data organized in a so-called run-off triangle (often on an annual or quarterly basis). Popular macro-level models include the chain-ladder method [2] and the Bornhuetter–Ferguson method [3]. These methods have been successfully applied for decades due to their ease of use and sound theoretical foundations (see, for example, [4, 5]). Furthermore, many extensions have been developed to produce more realistic results (e.g. [6–9]).

However, aggregating data may lead to several problems and yields a loss of information. Therefore, micro-level or individual claims reserving models focus on granular or claim-specific data of the individual claims. They aim to model two processes: a time process representing the individual states occupied by a claim, and a payment process which represents the amount paid for a claim in a particular state. The earliest articles on micro-level reserving include [10] and [11]. The modelling ideas from the early papers have been extended in a (semi-)parametric form [12, 13], as well as a nonparametric form [14, 15].

Beyond insurance reserving, individual-level modelling has proven fruitful across a range of financial risk applications. In credit risk, for instance, instance-based and Bayesian neural network approaches have been applied to investment decision-making in peer-to-peer lending [16, 17], demonstrating that granular, borrower-level models can outperform aggregate scoring methods in complex, high-dimensional settings. The mCube framework shares this individual-level philosophy, applied here to the claim development process in general insurance.

In this paper, we adopt the multi-state approach to loss reserving proposed by [18] and further considered by [19–21]. In particular, we

- choose a model for IBNR claim counts, which can be any model really, but in order to stay within the framework of the multinomial distribution, as detailed in Section 3, we go for an extension of the chain-ladder method [4];
- make a connection between the time process modelling and the multi-state competing risk framework;
- propose a new model for payments from RBNS claims that take into account the presence of large amounts of both signs. This is done by using a semi-parametric modelling of the payment distribution, through a mixture distribution with a generalised Pareto distribution (GPD) for the tails;
- include practical recommendations on how to apply the proposed methodology in order to develop IBNR and RBNS claims to their ultimate losses. These recommendations can be applied to any micro-level data set.
- include a feature engineering step which contains a binning step of the continuous variables, which is important to robustify the parametric assumptions which are typically made in most micro-level reserving models

- compare the predictive capabilities of our proposed method with other micro-reserving models.

We note that mCube is a hybrid framework by design: The IBNR count model operates at the macro level, leveraging the chain-ladder's well-established multinomial structure, while the claim development models operate at the micro level, exploiting individual claim information. This reflects the fundamental asymmetry that no individual covariate information is available for claims that have not yet been reported.

Thus, in our paper, we propose a novel approach by developing a unified approach for IBNR and RBNS claims utilizing the multinomial distribution as the central guiding principle. To the best of our knowledge, this is the first paper which attempts to accomplish this task. Our methodology relies on less parametric assumptions compared to other micro-level approaches, making it easier to fit. We also provide practical recommendations for actuaries who wish to implement the methodology on their own micro-level datasets. We recognize that the choice of tuning parameters depends on the specific dataset under consideration. To address this, we include a discussion on this issue, which is often missing in the literature. Additionally, we introduce a comparison of our methodology with two other published micro-level models.

2. The claims reserving problem

The development of a non-life insurance claim is presented in Figure 1.

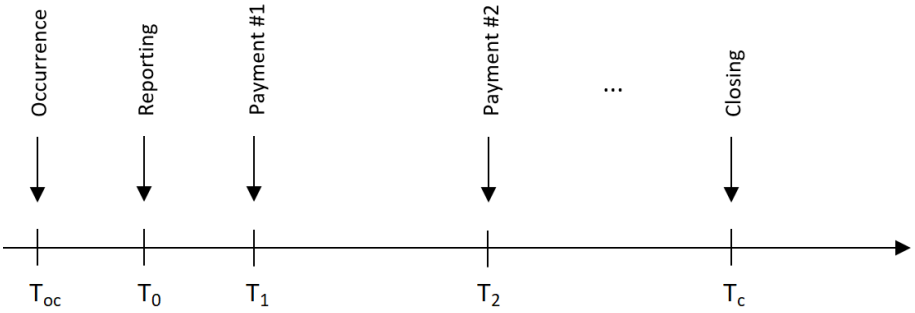


Figure 1. Claim development process.

The occurrence date, T_{oc}, is the date at which the claim event occurs, and the reporting date, T_0, refers to the date at which the claim is reported to the insurer. Once the insurer is aware of the claim and accepts the claim for reimbursement, some payments at different moments (here represented by T_1 and T_2) follow to compensate the insured for their loss. Once the insurance company reimburses the complete loss covered by the policy, the claim closes, which is represented by T_c. Note that we assume that once a claim is closed, it cannot be reopened.

At the moment of evaluation (commonly: end of a quarter, mid year, or end of book year), denoted τ, insurance companies have to set reserves aside to cover their future liabilities. These liabilities can come from four sources, which are enumerated below.

- Claims that have occurred before the evaluation period but which are not yet reported to the insurer, i.e. $T_{oc} \leq \tau < T_0$. These are called incurred but not reported (IBNR) claims.

- Claims that have occurred and have been reported before the evaluation period, but which are still open at evaluation date, i.e. $T_0 \leq \tau < T_c$. These are called reported but not settled (RBNS) claims.
- Reopened claims and unearned premiums. These liability sources exist in practice but are outside the scope of the present model. We assume that once a claim is closed, it cannot be reopened. If a dataset contains re-openings, a second micro-level RBNS model could be fitted to handle these separately. If needed, a binomial model could be used to estimate the number of reopened claims on a claims dataset.

Let us assume that we work on a sufficiently rich probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote $C_{k,t}$ as the random variable representing the cumulative amount paid for claim k at time t. Next, we want to predict the reserve at evaluation time τ, or in other words, the remaining amount to be paid for the claim until it is closed, denoted $R_{k,\tau}$. This can be estimated by $\hat{\mathbb{E}}[R_{k,\tau} \mid \mathcal{F}_\tau]$, denoted by $\hat{R}_{k,\tau}$, which is obtained by the difference of $\hat{C}_{k,T_c}$ and $C_{k,\tau}$. Here, F_τ represents the information available at time τ, and $\hat{C}_{k,T_c}$ equals the estimated total cost of the claim at closing time, which can be obtained by $\hat{\mathbb{E}}[C_{k,T_c} \mid \mathcal{F}_\tau]$.

Note that for IBNR claims, we have that $C_{k,\tau}$ equals zero and hence, $\hat{R}_{k,\tau} = \hat{C}_{k,T_c}$. Finally, the estimated reserve for the whole portfolio is calculated as follows:

$$\hat{R}_\tau = \sum_{k_1=1}^{\mathbf{n}^{RBNS}} \hat{R}_{k_1,\tau}^{RBNS} + \sum_{k_2=1}^{\hat{\mathbf{N}}^{IBNR}} \hat{R}_{k_2,\tau}^{IBNR}, \tag{1}$$

with $\mathbf{n}^{RBNS}$ representing the number of RBNS claims, $\hat{R}_\tau^{RBNS}$ the estimated RBNS reserve, $\hat{R}_\tau^{IBNR}$ the estimated IBNR reserve, and $\hat{\mathbf{N}}^{IBNR}$ the estimated number of IBNR claims. The goal of this paper is to determine $\hat{R}_\tau$, the estimated reserve of the whole portfolio, which is also called the best estimate cost of the portfolio. We use the term "best estimate" in the Solvency II sense: the probability-weighted average of future cash flows, as defined in Article 77 of the Solvency II Directive [1]. This does not imply optimality in a statistical sense, but rather denotes the central estimate of the reserve distribution.

2.1. The mCube framework

The mCube (multinomial multi-state micro-level) reserving framework is the composition of four components, each of which uses the multinomial distribution as its central building block:

1. **IBNR model** (Section 3): A Poisson regression model for the number of IBNR claims per accident year and reporting delay, rooted in the chain-ladder's implicit multinomial structure. The bootstrapped over-dispersed Poisson approach of [4] is used to obtain a distributional estimate.

2. **Multinomial time model** (Section 4.2): For each state S_j in the multi-state process, a multinomial logistic regression models the discrete-time transition probabilities $\lambda_{j,e}(t \mid \mathbf{x}_{k,t})$, conditional on the claim's covariate vector $\mathbf{x}_{k,t}$, where $e \in \{S_{j+1}, S_{tp}, S_{tn}\}$ denotes the possible destination states.

3. **Spliced payment model** (Section 4.3): Conditional on a payment transition, the payment amount is modelled by a multinomial logistic regression over L bins defined by a spliced distribution. The bins capture small payments (body), large payments (GPD tail), and negative payments, with bin probabilities depending on covariates.

4. **Simulation engine** (Section 4.4 and Algorithm 1 in Appendix C): The time and payment models are combined in a forward simulation that generates N_{sim} future trajectories for each open claim. The collection of simulated total costs yields the predictive distribution of the reserve $\hat{R}_{k,\tau}$ for each claim k, and the portfolio reserve $\hat{R}_\tau$ is obtained by aggregation via Equation (1).

The mCube framework is thus not a single statistical model but a modular reserving pipeline whose components are jointly calibrated and whose output is a full predictive distribution of reserves at the individual claim level. The modularity allows practitioners to replace individual components (e.g., substituting a different severity model) while retaining the overall structure. The four components are detailed in turn in Section 3 (IBNR model, Component 1), Section 4.2 (multinomial time model, Component 2), Section 4.3 (spliced payment model, Component 3), and Section 4.4 (simulation engine, Component 4), respectively.

3. IBNR model

In the context of micro-level reserving, it is of interest to know in which future years the remaining unreported IBNR claims will be reported. Before getting into the methodology that was used to this end in our study, we first want to add a reminder as to why the chain-ladder method is consistent with our central theme around the multinomial distribution. Let's start with some notation.

Consider $N_{i,j}$ referring to the (incremental) number of reported claims for a given accident year $i = 1, \ldots, I$ and development year $j = 0, \ldots, J-1$. As such, the number of claims of accident year i that are observed at the moment of evaluation is written as N_i^{obs}, and since a claim can only be observed once reported, we have that $N_i^{obs} = \sum_{j=0}^{I-i} N_{i,j}$. N_i represents the total number of claims that occurred in accident year i, and $N_{i,j}$ represents the number of these claims that have been reported after j years. Since we assume that no claims will be reported beyond the last development year, i.e., there is no tail factor, we have $N_i = \sum_{j=0}^{J-1} N_{i,j}$. Finally, the number of IBNR claims of accident year i is given by $N_i^{IBNR} = N_i - N_i^{obs} = \sum_{j=I-i+1}^{J-1} N_{i,j}$, and the number of IBNR claims over all accident years N^{IBNR} equals $\sum_{i=1}^{I} N_i^{IBNR}$.

The chain-ladder method can be employed to obtain a point estimate for N_i^{IBNR} of accident year i and the future cells $N_{i,j}$ with $i + j > I$. $\forall\ 1 \leq i \leq I$ and $0 \leq j \leq J-1$, the main assumption of the chain-ladder method [2] is the following Markov property:

$$\mathbb{E}[N_{i,j} \mid N_{i,0}, \ldots, N_{i,j-1}] = \mathbb{E}[N_{i,j} \mid N_{i,j-1}] = f_{j-1} N_{i,j-1}, \tag{2}$$

meaning that once the development factors $(f_0, \ldots, f_{J-2}, f_{J-1})$ are known, the future cells of each accident year can be computed iteratively. For $0 \leq j \leq J-2$, the development factors typically are estimated using the column-sum estimator:

$$\hat{f}_{j-1} = \frac{\sum_{i=1}^{I-j} N_{i,j+1}}{\sum_{i=1}^{I-j} N_{i,j}}. \tag{3}$$

f_{J-1} corresponds to the tail factor and is not estimated with the column-sum estimator, but since we do not assume any developments beyond the observed reserving triangle, by definition $f_{J-1} = 1$

holds. Do notice from the formulae above that stationarity is assumed among all accident years by the chain-ladder method.

Let's define $\alpha_{I-i} = 1/\lambda_{I-i}$ with $\lambda_{I-i} = f_{I-i} \times \ldots \times f_{J-2}$ where α_{I-i} corresponds to the cumulative proportion of accident year i that is developed at the evaluation moment. Consider π_j as the probability that a claim will be reported in the jth development period. It can be shown that the difference of two consecutive α values is equal to this probability or $\hat{\pi}_j = \alpha_j - \alpha_{j-1}$, with $\hat{\pi}_0 = \alpha_0$ as the sole exception. Hence, in the context of the chain-ladder method and for each accident year i, it is assumed that $(N_{i,0}, \ldots, N_{i,J-1}) \mid N_i$ is distributed as a multinomial distribution with probabilities $\mathbf{\Pi}_i = (\pi_{i,0}, \ldots, \pi_{i,J-1})$. In other words, the application of the chain-ladder method implies the latter multinomial assumption, which fits the overall framework that we are going for in this paper.

However, the chain-ladder method alone is not enough to satisfy our purposes, since we want to derive reserve distribution for IBNR claims, and the chain-ladder method only provides a point estimate of this reserve. As a result, the bootstrapped approach of [4], using the over-dispersed Poisson process distribution, is adopted in this paper, where $N_{i,j}$ is estimated for the future cell of each bootstrap sample, resulting in our derived estimate of the IBNR reserve. Since [4] is rooted in the chain-ladder framework, it naturally carries a multinomial character that aligns with the central theme of our paper, but technically any other IBNR model could be used that results in estimates for the distribution of $N_{i,j}$ of all future cells. Further details on the approach of [4] can be found in the main article.

3.1. Relation to continuous-time likelihood-based approaches

The mCube framework operates in discrete time, modelling transition probabilities at each time period via multinomial regression. An alternative class of micro-level reserving models operates in continuous time, specifying transition *intensities* directly. We briefly position mCube relative to these approaches.

Norberg [11] proposed modelling claim development as a marked point process with nonhomogeneous Poisson intensities, enabling maximum likelihood estimation of transition rates and analytic reserve formulae via martingale theory. Antonio and Plat [13] extended this framework to a fully stochastic micro-level setting with covariate-dependent intensities. Badescu et al. [22] introduced latent heterogeneity through Cox processes, where a claim-specific frailty term captures unobserved risk characteristics. More recently, Maciak et al. [23] proposed marked Hawkes processes to capture self-exciting dynamics in claim arrivals, where past events increase the intensity of future events.

Compared to these continuous-time approaches, the discrete-time multinomial framework of mCube offers several practical advantages: (i) discrete-time models align naturally with how insurers record and process claims data (monthly or quarterly snapshots); (ii) multinomial regression allows flexible, nonparametric covariate effects without specifying a parametric intensity function; and (iii) the spliced mixture payment model accommodates the heavy tails and multimodality commonly observed in payment distributions.

The trade-off is that continuous-time models offer a more parsimonious parametric structure and direct access to theoretical properties (e.g., martingale representations), which can be advantageous for closed-form reserve formulae. The mCube framework instead relies on simulation for inference, which provides full predictive distributions at the cost of greater computation. A detailed empirical comparison on a common dataset is a promising direction for future research.

4. Micro-level model components (Components 2–4)

This section details Components 2–4 of the mCube framework introduced in Section 2.1. Component 1, the IBNR count model, was presented at the macro level in Section 3; the components developed here operate at the level of individual claims and exploit claim-specific covariate information unavailable before a claim is reported. We begin with the multi-state framework that defines the state space on which Components 2 and 3 are built (Section 4.1), followed by the multinomial time model (Component 2, Section 4.2), the spliced payment model (Component 3, Section 4.3), and the simulation engine (Component 4, Section 4.4).

4.1. Multi-state approach

This section uses the multi-state approach of [21] represented in Figure 2. In this approach, an RBNS claim occurred in state S_{oc} and is reported in state S_0. Once reported, either a first payment can occur, implying a transition from state S_0 to state S_1, or the claim can go to one of two absorbing states, S_{tn} or S_{tp}. Here, S_{tn} represents the fact that the claim went to a terminal state without payment, and S_{tp} represents that the claim went to a terminal state with payment. In this framework, the states S_j, $\{j \in \{0, n_{pmax} - 1\}\}$ are all strictly transient and moreover, for $j > 0$, state S_j implies j payments were made prior to the current time point. The integer n_{pmax} represents the maximum number of transitions a claim is allowed before being forced to an absorbing state. More specifically, we represent the multi-state model $(\mathscr{S}, \mathscr{T})$ with state space $\mathscr{S} = \{S_0, S_1, \ldots, S_{n_{pmax}-1}, S_{tn}, S_{tp}\}$ and set of direct transitions $\mathscr{T}$. An event corresponds to the transition from one state in $\mathscr{S}$ to another. The set of direct transitions $\mathscr{T}$ defines all possible transitions in the multi-state model, indicated by arrows in Figure 2. One advantage of individual claims reserving models is that they allow us to take into account covariate information, such as the history of incremental payments, line of business, reporting delay, as well as any other type of information which is available to describe individual claims and their development. We denote by $\mathcal{F}_{k,T}$ the filtration containing all the information concerning claim k at time T. In a multi-state framework, C_{k,T_c} can be computed by determining the next states the claim will occupy and summing the amount paid in these future states together with the amount already paid.

4.2. Multinomial model for the time process

We use the methodology from discrete time survival analysis, and model the time until an event or transition from one state to the other. We define an event as being the occurrence of a payment or as the transition to a terminal state without payment, as in [19] and [21]. Furthermore, we say that a claim is censored or open when it is not in one of the two absorbing states at the moment of evaluation. The choice of discretization for the multi-state model of the previous section is arbitrarily chosen to be monthly, with one month corresponding to 30 days. The proposed model can be adapted to work with any discretization, but the number of parameters increase with the granularity considered for the data.

Our goal is to model the time $T_{k,j}$ of the transition of claim k from a state S_j to a state S_{j+1}, $j \in \{0, 1, \ldots, n_{pmax} - 2\}$ or to a terminal state, using covariate information included in $\mathcal{F}_{k,t}$, through the covariate vector $\mathbf{x}_{k,t}$. If we denote by $\Delta_{k,j}$ the random and discrete censoring time of claim k in state S_j, we have the following assumption:

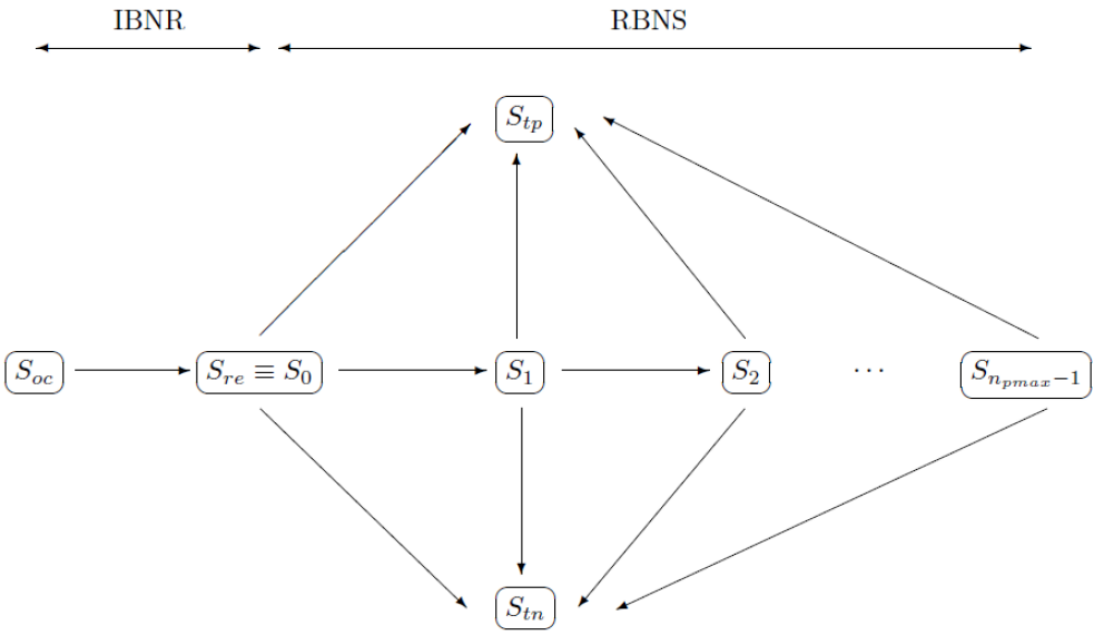


Figure 2. Discrete time multi-state model; source: [20].

Assumption 4.1. *We assume that $T_{k,j}$ and $\Delta_{k,j}$ are independent and that the censoring mechanism is noninformative, i.e. that censoring is independent from the transition type and transition time.*

Based on discrete-time competing risks literature [24], we represent the event type as a random variable $\epsilon_{k,j}$, taking values P, TP, and TN corresponding, respectively, to a transition due to a payment, a terminal payment, or a termination without payment. The discrete-time cause-specific hazard functions are then modelled by

$$\begin{aligned}
\lambda_{j,j+1}(t \mid \mathbf{x}_{k,t}) &= \mathbb{P}(T_{k,j} = t, \epsilon_{k,j} = P \mid T_{k,j} \geq t, \mathbf{x}_{k,t}) \\
&= \frac{\exp(\alpha_{j,j+1} + \beta_{j,j+1}^T \mathbf{x}_{k,t})}{1 + \sum_e \exp(\alpha_{j,e} + \beta_{j,e}^T \mathbf{x}_{k,t})}, \\
\lambda_{j,tp}(t \mid \mathbf{x}_{k,t}) &= \mathbb{P}(T_{k,j} = t, \epsilon_{k,j} = TP \mid T_{k,j} \geq t, \mathbf{x}_{k,t}) \\
&= \frac{\exp(\alpha_{j,TP} + \beta_{j,TP}^T \mathbf{x}_{k,t})}{1 + \sum_e \exp(\alpha_{j,e} + \beta_{j,e}^T \mathbf{x}_{k,t})}, \\
\lambda_{j,tn}(t \mid \mathbf{x}_{k,t}) &= \mathbb{P}(T_{k,j} = t, \epsilon_{k,j} = TN \mid T_{k,j} \geq t, \mathbf{x}_{k,t}) \\
&= \frac{\exp(\alpha_{j,TN} + \beta_{j,TN}^T \mathbf{x}_{k,t})}{1 + \sum_e \exp(\alpha_{j,e} + \beta_{j,e}^T \mathbf{x}_{k,t})}, \\
\lambda_{j,j}(t \mid \mathbf{x}_{k,t}) &= \mathbb{P}(T_{k,j} > t \mid T_{k,j} \geq t, \mathbf{x}_{k,t}) \\
&= 1 - \lambda_{j,j+1}(t \mid \mathbf{x}_{k,t}) - \lambda_{j,tp}(t \mid \mathbf{x}_{k,t}) - \lambda_{j,tn}(t \mid \mathbf{x}_{k,t}),
\end{aligned} \tag{4}$$

where e iterates over all event types. $\alpha_{j,e}$ and $\beta_{j,e}^T$ are the parameters in the multinomial regression model relating to event type e. Following Assumption 4.1, the parameters are estimated by their maximum

likelihood estimator*. For a claim that has occurred but has not yet been reported, only one event in the multi-state process is possible, namely reporting. Similarly to [21], we estimate the monthly probability of reporting using a binomial generalized linear model (GLM):

$$\begin{aligned}\lambda_{oc,0}(t \mid \mathbf{x}_{k,t}) &= \mathbb{P}(T_{k,oc} = t \mid T_{k,oc} \geq t, \mathbf{x}_{k,t}) \\ &= \frac{1}{1 + \exp(\alpha_{oc} + \beta_{oc}^T \mathbf{x}_{k,t})},\end{aligned} \tag{5}$$

with α_{oc} and β_{oc}^T representing the logistic regression parameters estimated by their maximum likelihood estimator. Note that $T_{k,j}$ denotes the time period at which the claim moves out of state S_j and is reset to 0 each time the claim enters a new non-absorbing state to signify that this is a new state and hence another discrete-time competing risk problem. We treat each discrete time unit as a separate observation in the data set. Hence, for a claim k in state S_j, there are as many lines as the number of time units the claim is in this state. We note that the number of IBNR claims are estimated on a yearly time scale, and we can estimate reporting probability for each month using Equation (5) with the RBNS claims. Once we know a claim will be reported in a given year, the month of reporting can be simulated by a random draw from a Bernoulli distribution using the estimated reporting probability. A representation of the different cause-specific hazards is shown in Figure 3.

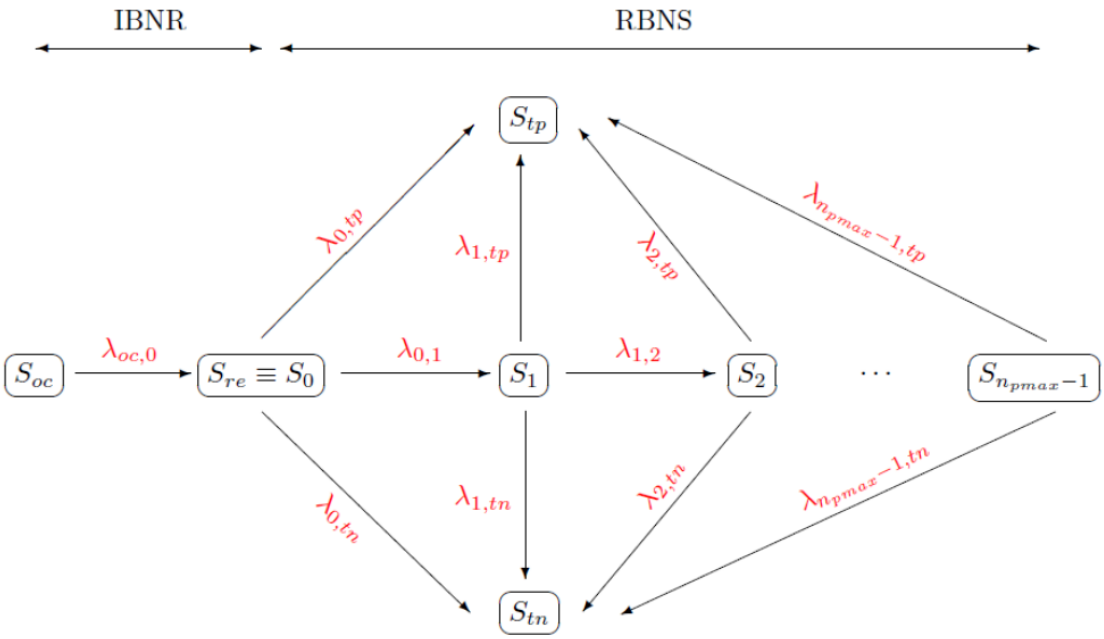


Figure 3. Representation of transition probabilities in the discrete time multi-state model; source: [20].

4.3. Modelling of the payment distributions

The difficulties in modelling the payment distribution arise from some stylised properties. First of all, negative payments can be present. For example, when the insurance company has to pay a

*using the function `multinom` in the package `nnet` [25] in `R`

third party and the insured has an insurance policy with a per-loss deductible of d, she has to pay d to the insurance company. Moreover, the payment distribution consists of a high number of small incremental payments, and a small number of very large payments in absolute value. Multiple models have been proposed to overcome these issues: Antonio and Plat [13] use a lognormal distribution to model payments, Bettonville et al. [21] use a mixture of lognormal distributions for the first payment and a mixture of lognormal or lognormal and pareto distributions for the link ratios, Frees and Valdez [26] use a generalised beta distribution of the second kind, and Pigeon et al. [12] use a multivariate extension of the univariate skew normal distribution. Reynkens et al. [27] propose to model censored losses using a mixed Erlang distribution for the body of the distribution and a generalised Pareto distribution (GPD) for the tail. The authors also propose to use the mean excess plot [28] to assess when to split the body and the tail of the distribution. This estimation procedure has the advantage of taking into account both random censoring and truncation. We propose to model the payment distribution using a data-driven modification of [29], which allows for the inclusion of covariate information and which can model the skewness of the distribution. We have decided on this approach because it allows us to model both positive and negative payments directly, can handle large payments of either sign, is easy to fit, and does not require the use of the expectation maximization algorithm, which can lead to computational instabilities. Let Y^j denote the random variable representing the payment size for a claim in state S_j. Moreover, we make the following assumption on the conditional distribution of Y^j given $\mathbf{x_t}$.

Assumption 4.2. *We assume that the density of Y^j conditional on $\mathbf{x_t}$ is an L-component mixture, i.e. $f^j(y \mid x) = \sum_{l=1}^{L} \xi_l^j(x) f_l^j(y)$, where $\xi_l^j(x)$ is the lth element of the covariate-dependent vector of multinomial logistic mixture weights and f_l^j are the densities of the mixture components. We further assume that L is known, f_1^j and f_L^j are densities of a GPD, and f_l for $l \in \{2, \dots, L-1\}$ are truncated normal distributions on the interval $[b_{l-1}^j, b_l^j]$. In this case, $b_1^j, \dots, b_{L-1}^j$ represent the splitting points separating the density into bins $\mathcal{B}_1^{(j)}, \dots, \mathcal{B}_L^{(j)}$.*

A representation of Assumption 4.2 is shown in Figure 4 with $L = 4$ bins. The choice of L involves a balance between interpretability, parsimony, and data requirements. A minimum of $L = 3$ bins is needed to separate the two GPD tails from a central body. The presence of negative payments in bodily injury portfolios — arising, for example, from deductible reimbursements — motivates a separate negative tail component, pushing the natural minimum to $L = 4$. Adding further interior bins ($L > 4$) provides finer granularity at the cost of additional parameters in the multinomial logistic weight model, which requires a sufficient number of observations per bin to estimate reliably. For the bodily injury case study, $L = 4$ is chosen for three reasons. First, the fixed splitting point $b_2^j = 0$ separates negative from positive payments by construction, yielding the four economically interpretable categories shown in Figure 4: large negative payments ($\mathcal{B}_1$, e.g. claw-backs or deductibles), small negative payments ($\mathcal{B}_2$), small positive payments ($\mathcal{B}_3$, e.g. routine medical costs), and large positive payments ($\mathcal{B}_4$, e.g. major disability settlements). Second, this four-category structure aligns with standard actuarial segmentation of bodily injury payment patterns, as further discussed in Appendix E. Third, the observation counts per bin across all states are sufficient to estimate the multinomial logistic weights with covariates, which would not be guaranteed for rarer transition states if L were increased further. In general, we recommend that practitioners select L based on the sign structure of their payment data and verify that each bin contains a sufficient number of observations to support the chosen covariate set in the weight model.

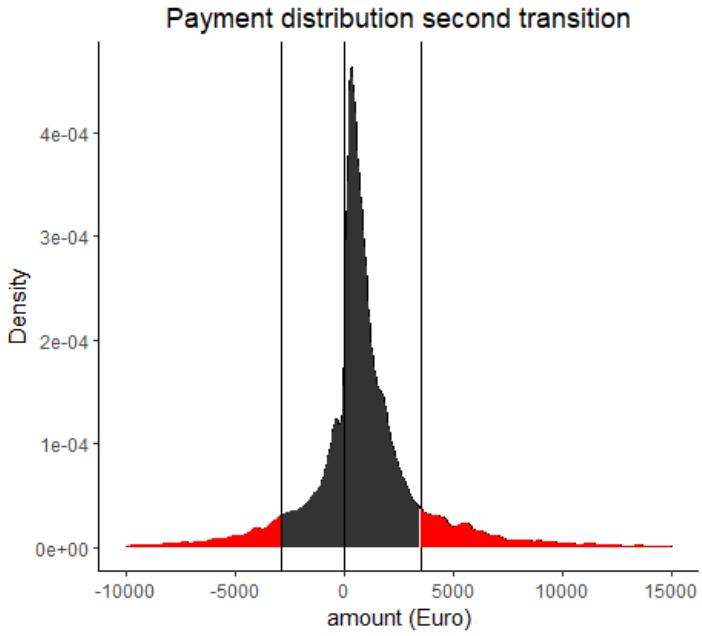


Figure 4. Spliced payment distribution with three splitting points.

The splitting points $b_2^j, \ldots, b_{L-2}^j$ can be chosen freely so that each bin has some interpretation. In practice, four splitting points can be chosen as shown in Figure 4 to represent small or large negative payments, as well as small or large positive payments. The leftmost and rightmost splitting points, respectively b_1^j and b_{L-1}^j, need to be well chosen in order to assure that the observations of these bins can be considered to be a sample from a GPD with b_1^j, ι_1^j, and φ_1^j (b_{L-1}^j, ι_L^j and φ_L^j) as the location, scale and shape parameter for $\mathcal{B}_1^j$ ($\mathcal{B}_L^j$). To this end, we can use tools from extreme value theory, such as the mean excess plot or the Gerstengarbe plot [30]. Following Assumption 4.2, the expected payment for a claim k in state S_j conditional on its covariate vector $\mathbf{x}_{k,t}$ is given by

$$\mathbb{E}[Y^j \mid \mathbf{x}_{k,t}] = \sum_{l=1}^{L} \xi_l^j(\mathbf{x}_{k,t})\mu_l \tag{6}$$

$$\xi_l^j(\mathbf{x}_{k,t}) = \frac{\exp(\gamma_{0,l}^{(j)} + \mathbf{x}_{k,t}^T \gamma_l^{(j)})}{1 + \sum_{m=1}^{L} \exp(\gamma_{0,m}^{(j)} + \mathbf{x}_{k,t}^T \gamma_m^{(j)})},$$

with μ_l being the mean of the *lth* component in the mixture. The parameter vectors $\gamma_0^{(j)}$ and $\gamma^{(j)}$ are estimated using maximum likelihood[†]. Also, the parameters $\mu_l, \ldots, \mu_L$ are obtained using maximum likelihood estimation. Hence, we can express the expected cumulative amount paid for claim k at closure as

$$C_{k,T_c} = \sum_{j:S_j \notin \{S_{tm}, S_{tp}\}} \mathbb{E}[Y^j \mid \mathbf{x}_{k,T_{k,j}}]. \tag{7}$$

[†]using the function `multinom` in the package `nnet` [25] in R

4.4. Total cost simulation for RBNS claims

Similarly to [31], we choose a one-period-ahead forecast to determine the estimated reserves $\hat{R}_{k,\tau}$ for each open claim k, as it allows us to intervene during the settlement process and helps us keep interpretability in the claims development process. In the modelling of the time process (4), the next state to visit is decided by a multinomial probability vector. Instead of assigning an open claim to the state with the highest multinomial probability, we take a sample of size one from the estimated multinomial distribution. Next, the expected payment for a claim in that state and with a given covariate vector is simulated using (6). All this has the advantage of adding variability to the predictive distribution. We note that using the expected value of the payments results in smoothed inputs in the simulation of RBNS total cost as mentioned in [31, 32]. This is not an issue in [31], as the authors do not use past payments as predictors, and hence, they do not need to simulate payments. However, when using our methodology, we use past payments as predictors hence, we need to validate the payment distribution through probability-probability (P-P) plots. Another strategy could be to take into account the full distribution of future payments by simulating an observation from a mixture component chosen based on the probabilities in (6).

Repeating these multinomial samplings N_{sim} times, we obtain a predictive distribution for C^r_{k,T_C}, the total payment of claim k, and its reserve $R^r_{k,\tau} = C^r_{k,T_C} - C_{k,\tau}$ with $r \in \{1, \ldots, N_{sim}\}$. Figure 5 summarises this simulation strategy schematically. The methodology illustrated here is slightly modified in Appendix C to make sure that the simulation for open claims completes in finite time. For example, a maximum number of transitions is set for each given claim as well as a maximum time that the claim is allowed to be open. Further details on the simulation process are also added in Appendix D.

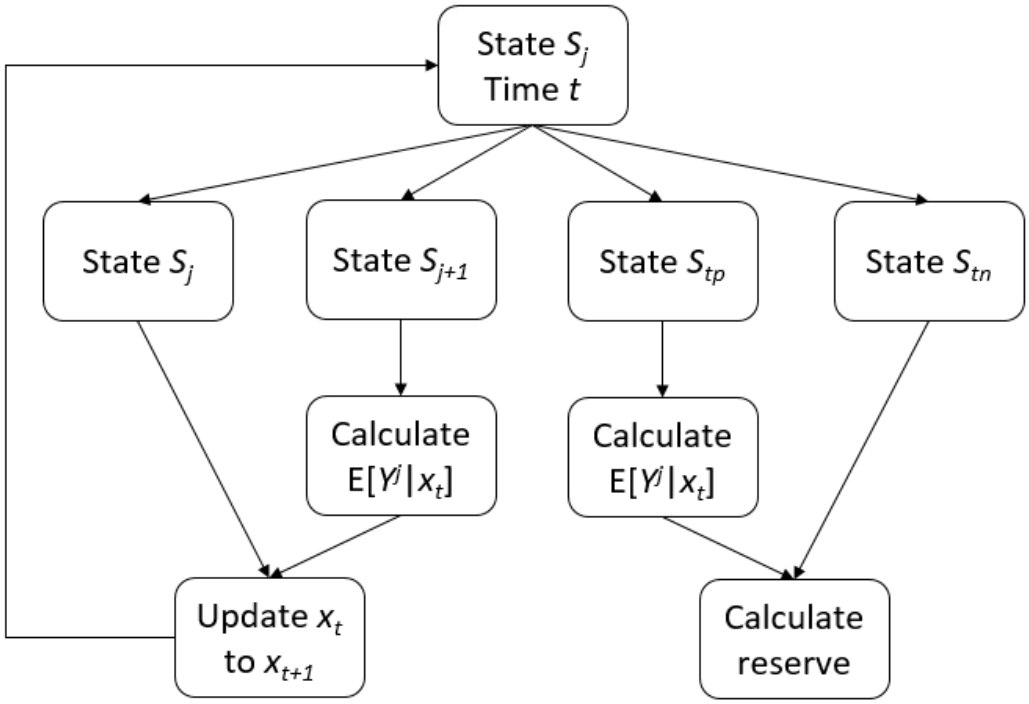


Figure 5. RBNS total cost simulation strategy.

4.5. Total cost simulation for IBNR claims

For IBNR claims, an extra step is added compared to RBNS claims, since we need to take into account the reporting delay. For each IBNR claim, this extra step consists in taking a sample of size one from a Bernoulli distribution using the estimated reporting probabilities (5). If this Bernoulli sample is 1, the RBNS total cost simulation strategy is applied. If the sample equals 0, then the covariate vector is updated and the claim remains in state S_{oc}. The Bernoulli sampling procedure is repeated until the claim leaves the state S_{oc}.

5. Hyper-parameter optimization

The flexibility of the proposed models lead to multiple hyper-parameters that need to be set prior to fitting the time and payment model. In this section, we explain the role of each hyper-parameter, as well as our tuning strategy. Furthermore, we want to alleviate the linearity assumption of the continuous variables that are used in the various regression models of mCube. To this end, each continuous variable will be binned into categories using the strategy explained in the next section. A binning strategy is applied such that the proposed noncomplex model can still capture complicated patterns.

5.1. Feature engineering

Individual claims data contain both static information, such as the line of business, claim type, and injured body part, as well as dynamic information, such as the cumulative payments and dates, that evolve throughout the claim's life. Table 1 shows a sample of the dynamic information available in an individual claim.

Table 1. Example of the dynamic claim information available from the database.

PolNumb	cumPay	bookDate	accDate	repDate	Status	closedDate
2640440	4,087.61	09-01-2012	01-01-2012	02-01-2012	O	28-08-2012
2640440	4,127.11	10-01-2012	01-01-2012	02-01-2012	O	28-08-2012
2640440	7.12	02-02-2012	01-01-2012	02-01-2012	O	28-08-2012
2640440	297.12	07-02-2012	01-01-2012	02-01-2012	O	28-08-2012
2640440	297.12	28-08-2012	01-01-2012	02-01-2012	C	28-08-2012

Since mCube requires that all time-varying variables are transformed into their respective values at the end of the fixed discrete time steps, the raw data set shown in Table 1 needs to be processed. More specifically, in our case, we have set 30 days as our fixed time step (*"perLen"*), and for example, for the cumulative payment, we record its value at the end of the subsequent time step. If this value differs in absolute value more than *"minPayVal"* from the previously recorded payment, we consider that a payment was performed and the claim will experience a transition from the current state. In the case where the claim resides in S_0, the previously recorded payment equals 0. As a result, transType will be set to P, TN, or TP, depending on the transition type. In the case where no payment has occurred, transType will be equal to N. See Table 2 for how the information in Table 1 will be transformed given the above described rules.

Next, we will work out how all commonly present time-varying variables are transformed to comply to mCube. These variables can be considered as the minimal set of variables that are present in

the data set used for any micro-level reserving model, and this set can be augmented by other time-varying variables that happen to be recorded by the insurance company at hand. As we work in a discrete time setting with a chosen period length (*"perLen"*), we denote the event times for claim k as $t_0^k \leq t_1^k < t_2^k \ldots \leq t_Q^k$ with t_0 representing the accident date, t_1 the reporting date, $t_2, \ldots, t_{Q-1}$ the payment dates, and t_Q representing the closing date. If $i \geq 1$ and t is such that $t_i^k \leq t < t_{i+1}^k$, we create the following variables for claim k:

- $x_{k,t}^1 = \max\left(1, \left\lceil \frac{t_1^k - t_0^k}{perLen} \right\rceil\right)$, to represent the reporting delay (deltRep).
- $x_{k,t}^2 = \mathbb{1}_{t_1^k = t_0^k}$, to a represent a fast reporting indicator (fastRep).
- $x_{k,t}^3 = \max\left(1, \left\lceil \frac{t - t_1^k}{perLen} \right\rceil\right)$, to represent the time since reporting (inProcTime).
- $x_{k,t}^4 = y_{i-1}^k$, the payment at time t_{i-1}^k, with $i \geq 3$ (delt1Pay).
- $x_{k,t}^5 = \left\lceil \frac{t - t_{i-1}^k}{perLen} \right\rceil$, with $i \geq 3$ for the time since the previous payment (delt1PayTime).
- $x_{k,t}^6 = \sum_{\{s: t_s^k < t\}} y_s^k$, for the cumulative payments up to time t (cumDel1tPay).
- $x_{k,t}^7 = x_{k,t}^3 \mathbb{1}_{\{i=1\}} + x_{k,t}^5 \mathbb{1}_{\{i>1\}}$ for the time spent in the current state (inStateTime).

Hence, at each time t, we have the claim's feature information vector $\mathbf{x}_{k,t} = (x_{k,t}^1, x_{k,t}^2, x_{k,t}^3, x_{k,t}^4, x_{k,t}^5, x_{k,t}^6, x_{k,t}^7, \mathbf{x}_{base})$ with $\mathbf{x}_{base}$ representing the remaining static information of the claim. We note that the "1" in the variable names represent the fact that we are taking the first lag for these variables. When we want to model the payment distribution, we also add an indicator of if it is a terminal payment or not as a covariate, as this information is always available before estimating a payment. Using the accident, reporting, booking, or settlement date, additional features, such as the day, month, quarter, or financial year during which any of these events have occurred, can be engineered. These help us capture calendar effects more easily. For the sake of simplicity, none of these seasonal variables were included in our analysis, even if there is no technical limitation not to do so. Table 2 contains the transformed database, which contains all of the proposed transformations of this section.

5.2. Binning continuous predictors

We use a modified version of the data-driven strategy to bin continuous payment-related variables of [33], where the authors propose to fit a generalized additive model (GAM), where the covariate effects of the continuous variables are fitted using cubic splines. More specifically, in our proposal, the spline estimates of the continuous predictors are binned using a regular regression tree. Therefore, we make the following adaptation to the algorithm of [33]

1. First, a sufficiently large bootstrap sample is taken from the data set. We recommend to sample between 50,000 to 100,000 observations for each bootstrap sample and to only take a limited number of bootstrap repetitions. In our case study, we have used a sample size per bootstrap of 100,000, and 10 bootstraps repetitions were taken.

2. In each bootstrap repetition, and for each continuous predictor, we split each continuous variable into 40 groups where the split points are the 0.025, 0.05,. . . , 0.95, 0.975 quantiles. For each group, the median value of the continuous variable of interest is chosen as the group representative or mediod.

Table 2. Example of the dynamic claim information available from the database.

polNumb	cumPay	bookDate	accDate	repDate	transType	closedDate
2640440	4,127.11	10-01-2012	01-01-2012	02-01-2012	P	28-08-2012
2640440	297.12	07-02-2012	01-01-2012	02-01-2012	P	28-08-2012
2640440	297.12	08-03-2012	01-01-2012	02-01-2012	N	28-08-2012
2640440	297.12	07-04-2012	01-01-2012	02-01-2012	N	28-08-2012
2640440	297.12	07-05-2012	01-01-2012	02-01-2012	N	28-08-2012
2640440	297.12	06-06-2012	01-01-2012	02-01-2012	N	28-08-2012
2640440	297.12	06-07-2012	01-01-2012	02-01-2012	N	28-08-2012
2640440	297.12	28-08-2012	01-01-2012	02-01-2012	TN	28-08-2012

deltRep	fastRep	inProcTime	delt1Pay	cumDelt1Pay	inStateTime	state
1	0	1	NA	NA	1	S_0
1	0	2	4,127.11	4,127.11	1	S_1
1	0	3	−3829.99	297.12	1	S_2
1	0	4	−3829.99	297.12	2	S_2
1	0	5	−3829.99	297.12	3	S_2
1	0	6	−3829.99	297.12	4	S_2
1	0	7	−3829.99	297.12	5	S_2
1	0	8	−3829.99	297.12	6	S_2

3. We then fit a multinomial regression similar to (4) in which the variable $x_{k,t}^7$ (inStateTime) and the medioids of interest obtained in the previous step are used as the predictors and the transition type is chosen as a response.

4. For each hazard function, the corresponding multinomial parameter estimates for each group representative are used as responses in a local regression (loess). The predictors of the local regression are the medioids. Note that the approach described from step (2) to (4) can also be replaced by using a spline estimate for the considered continuous variable in the multinomial model of step (3), instead of what is currently proposed. However, this requires a much longer fitting time than the current proposal, and the end result of both approaches was found to be relatively similar.

5. For each hazard function, a regression tree is then fitted using the predictions of the loess fit on the current bootstrap sample as a response, and the continuous variable of interest as a predictor. This gives a set of *"nGroups"* -1 splitting points for the continuous variable.

6. In the final step, the splitting points of the different hazard functions are merged, as to obtain a single set of splitting points that is used to bin the considered continuous variable in the same way for each considered transition function. This is done by ordering and merging the splitting points of all three hazard functions.

Note that we impose that each bin obtained by this method has at least *"nMinLev"* observations. The choice of the hyper-parameters *"nGroups"*, *"nGroupsFin"*, and *"nMinLev"* is discussed in Appendix D.

5.3. Remaining hyper-parameters

Due to the high flexibility of the time and payment models of mCube, multiple hyper-parameters need to be set prior to fitting any of the models. Due to the computational complexity of mCube and the large number of hyper-parameters and possible values for these hyper-parameters, a search in the hyper-parameter space is not feasible. We propose to choose values for the hyper-parameters based on patterns present in the data and business logic. We refer to Appendix C for more details on this matter.

6. Case study

In this section, the mCube is applied on a real data set of a major insurance company in order to explore its performance. The R code applied on simulated data is openly available at the GitHub repository `https://github.com/emmanueljordy/mCube`.

6.1. Data

The data used in this section is a random sample of the set of claims obtained from a European insurer and resulting in a total of 25,821 body injury claims occurring between 2006 and 2012, of which the latest evaluation moment was December 31, 2018 with some of the claims still being open at that time. To anonymize the payment data, we have multiplied all payments by a nondisclosed constant value. Furthermore, we adjust all payments for inflation. All the information that is present prior to and including December 31, 2012 is equal to the training set. Our aim is to predict the remaining cost of every open claim at the end of December 31, 2012 until December 31, 2018. The set of all such open claims corresponds to our entire test set. Since we have all information up to December 31, 2018, and since we have 6 years of information for each claim, the true cost evaluated at December 31, 2012 is known for all claims in our test set.

From Figures 6a, 6b, 6c and 6d, obtained from the entire data set, we observe that the number of accidents is stable over the different accident years and most claims are reported in the month in which they happened. The claim settlement distribution is right skewed with most claims being settled within 2 years. The distribution of the total number of transitions for claims in the data set is shown in Appendix I. We observe that only around 13% of claims have more than 6 transitions. Furthermore, we observe that most claims have 2 or 3 transitions in the multi-state process and about 64% of all selected claims have 3 transitions or less; hence, a substantial part of the data set consists of claims with a long and relatively complicated development pattern, which is rather typical for bodily injury claims. The distribution of the time spent in each state in Appendix J. We note that observations for states greater than 5 are lumped together in a state S_{5+} due to the small number of observations in each of the individual data sets and given that only a small percentage of the claims have more than 6 transitions (see also Appendix I). We observe that claims spend, on average, between 5 and 7 months (30-day periods) in each state. Furthermore, states S_0 and S_{5+} are the states with the lowest average time spent and also the most skewed. The reason for this is that these data sets have the most diverse types of claims.

The payment distribution in each state in Appendix J. From this table, we can see how complicated payment distributions are, due to the presence of negative payment amounts, small median payment amounts, right skewed distributions, and very large excess kurtosis. Furthermore, states S_0 and S_{5+} have by far the highest values for skewness and kurtosis, which underlines yet again the heterogeneity present in the claim that are in both states.

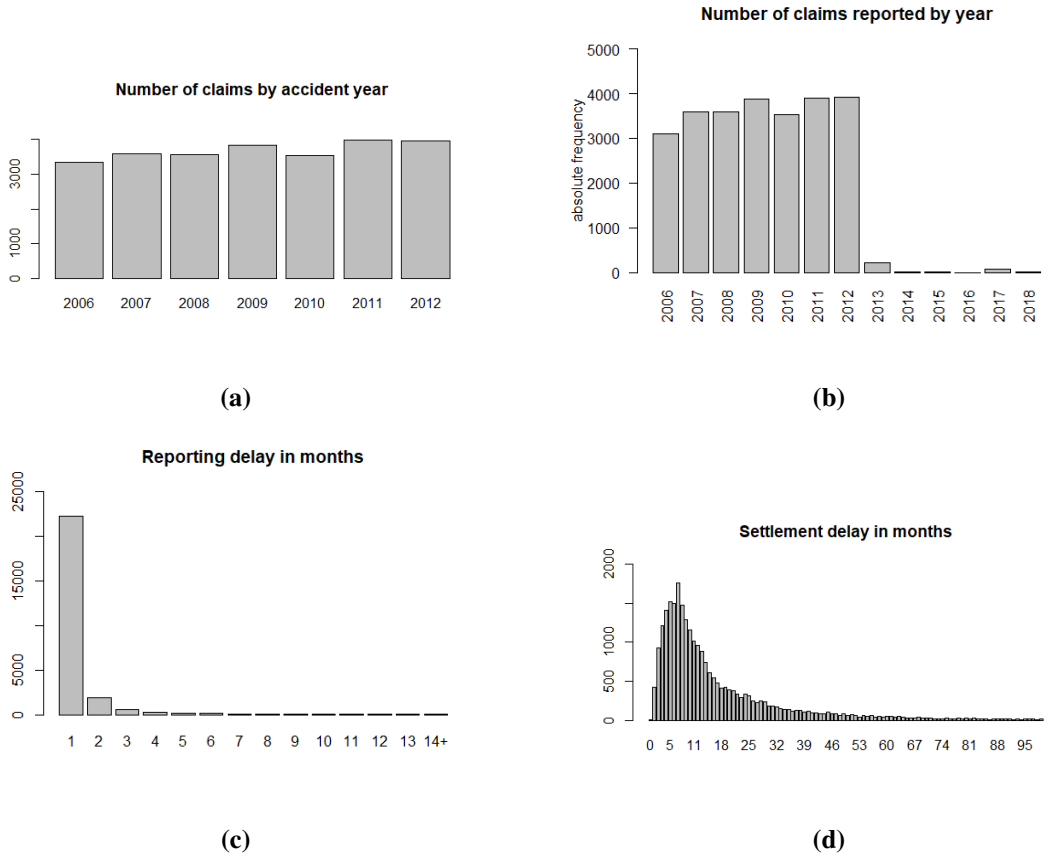


Figure 6. Distribution of the number of claims per accident and reporting year, together with the distribution of the reporting and settlement delay.

6.2. Time models evaluation

We will now evaluate the performance of the time models on the training data. Table 3 shows the number of observations (rows) for each data set, used to estimate the time models. Since we have added one row for each time period that a claim has spent in the respective state, we observe a high number of observations in Table 3. To evaluate the accuracy of the time models, we perform for each considered state a 5-fold cross-validation (CV) where for each claim in a hold-out set, we predict the time that it takes to exit as well as the state it will exit to. For a claim in a holdout set, we simulate 100 trajectories, and for each trajectory, we record the time it took to exit the state and the transition type. For each claim, we define the final transition type as the transition type that was simulated the most in the 100 trajectories. Furthermore, we note that once a claim has stayed more than 24 months in a state, we force it to exit the state, and once it has stayed more than 180 months in the process, we force the claim to an exit state. Table 4 shows the percentage of correctly predicted transitions and the mean bias time averaged for the 5 folds. We observe that for most states, we can predict around 80% of the correct transitions. State S_{5+} is more complicated as claims from different states are used to build that model, and hence we observe a drop in the performance. In state S_{5+}, claims with more than 5 payments are grouped, since there are not enough observations to estimate a separate model for each individual payment beyond the fifth payment. As a result, we are obliged, due to data limitations, to group claims together that are potentially very heterogeneous in nature. We also observe that for states $S_0, \ldots, S_4$, the mean bias (predicted - true) exit time is close to 0, and for state S_{5+}, it is negative; hence, the next states, as well as the time a claim stayed in a state, are correctly estimated.

Table 3. Number of observations in each state to model the time process.

	State					
	S_0	S_1	S_2	S_3	S_4	S_{5+}
abs.freq	64,253	76,977	58,342	27,476	15,462	53,520

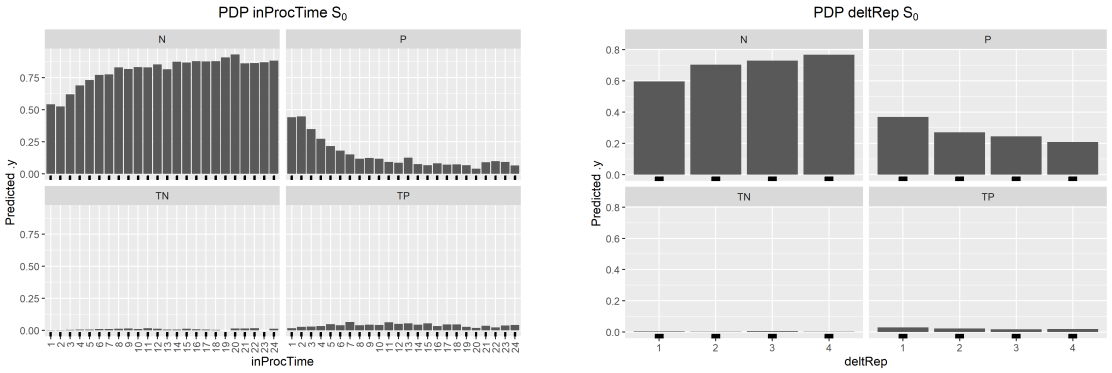
Table 4. Percentage of correct transitions and mean bias time (months) for the time models.

	State					
	S_0	S_1	S_2	S_3	S_4	S_{5+}
% correct transitions	92%	83%	82%	84%	86%	68%
mean bias time	0.14	0.09	−0.23	−0.08	0.04	−14.09

6.3. Interpreting marginal effects of covariates on the time models

In this section, we investigate the marginal effect of the covariates on the transition probabilities using partial dependence plots (PDP) introduced by [34] and implemented using the `iml` [35] package in R. These PDP illustrate the marginal effect of a covariate (predictor) on the predictions made by the models. This marginal effect is marginalized over the other covariates, meaning that to get the PDP for a categorical variable, we assign to each observation that same category for the variable of interest and average the predictions. Hence, as described by [36], the PDP of a category represents the average prediction made by the model if we force each observation to have that category for the given categorical variable.

From Figure 7, we observe that for claims in S_0, an increase in the time spent in the process, and hence in the time spent in the state, decreases the probability of a payment. Similarly, high values of the reporting delay decrease the probability of exiting the state.



(a) **(b)**

Figure 7. Partial dependence plots representing the marginal effect on transition probabilities of the time spent in the process and the reporting delay for claims in S_0.

For claims in S_1, we have two extra covariates, namely the binned version of the size of the first (hence previous) payment, where the bins are given in Appendix E, and the time spent in the current state. The transition probabilities of S_1 are shown in Figure 8. We observe that, as for the model for claims in S_0, high values for the time spent in the process, the time spent in the state, and the reporting delay decrease the probability of exiting the state. We also observe that a large cumulative previous payment increases the probability of having a payment.

Figure 9 shows the transition probabilities of S_2; we have one extra covariate, as compared to S_1, namely the binned size of the previous payment, where the bins are given in Appendix E, and the time spent in the current state. Just like for the S_1 time model, we observe that high values for the time spent in the process (inProcTime) and the time spent in the state (inStateTime) decrease the probability of a transition. We also observe that a large cumulative previous payment increases the probability of having a payment. However, the size of the previous payment only has a small impact on the transition probabilities. Transition probabilities for claims in S_3, S_4, and S_{5+} are shown in Appendix F, where we observe similar effects for the covariates as in S_2.

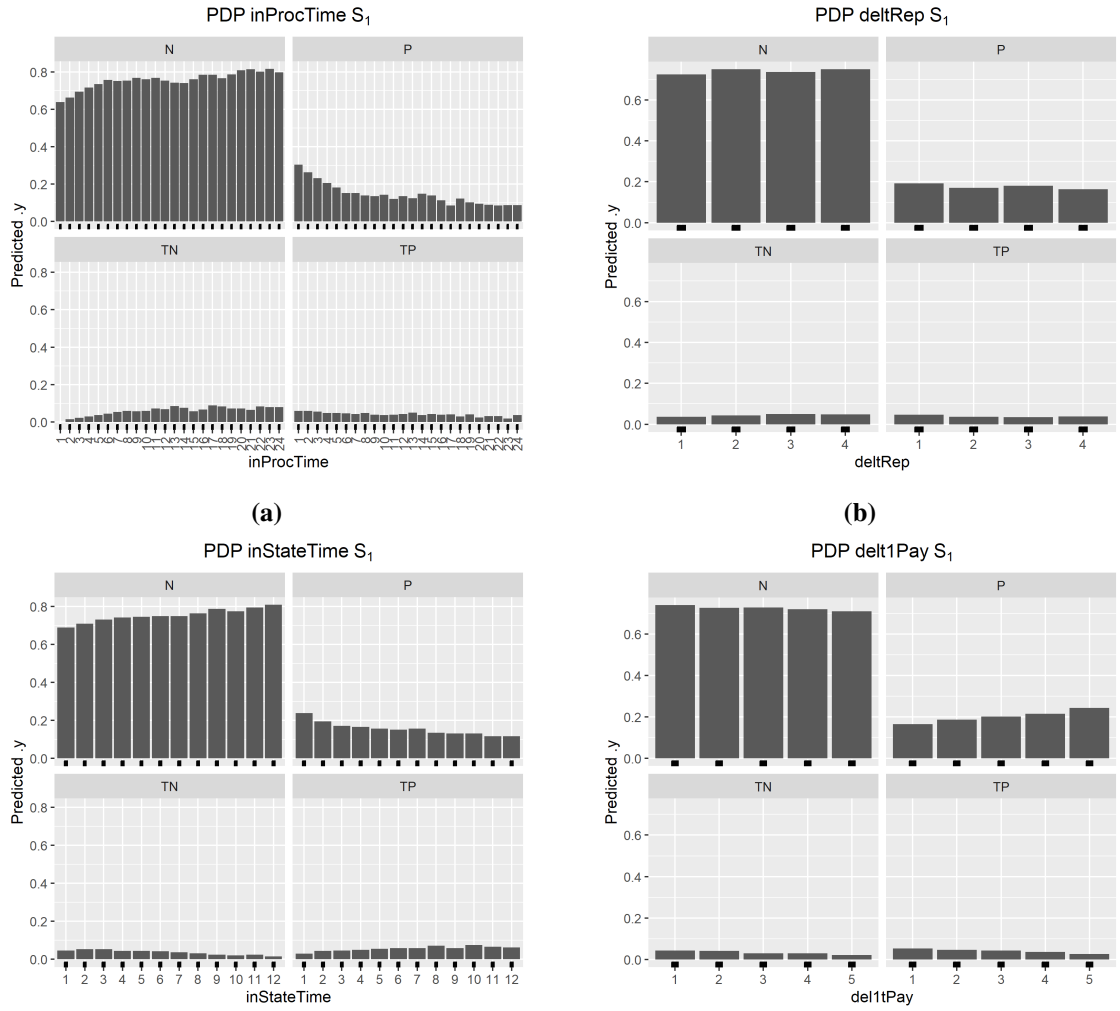


Figure 8. Partial dependence plots representing the marginal effect on transition probabilities of the time spent in the process (a), the reporting delay (b), the time spent in the state (c), and the previous payment size (d) for claims in S_1.

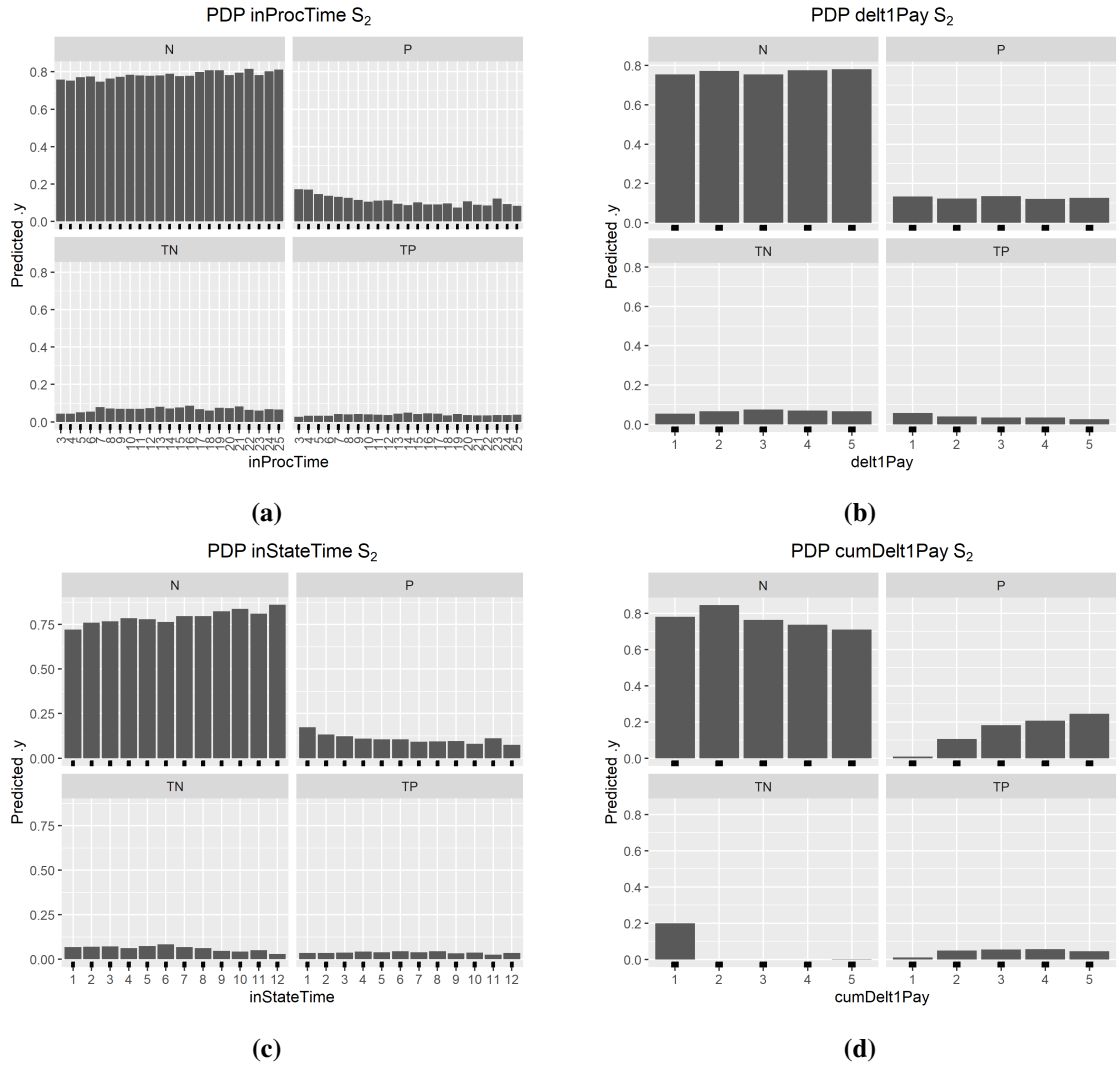


Figure 9. Partial dependence plots representing the marginal effect on transition probabilities of the time spent in the process (a), the previous payment size (b), the time spent in the state (c), and the cumulative previous payment size (d) for claims in S_2.

6.4. Payment models evaluation

In this section we evaluate the accuracy of the payment models used through 5-fold cross-validation. Table 5 shows the number of observations in each state, which will be used to estimate the payment distributions. We observe that the number of observations used to fit the payment model of a given transition is lower than the number of observations of the time fit of the same transition (see Table 5 versus Table 3); this is because we model a payment distribution conditionally on the occurrence of a payment. Hence, to model the payment distribution, we only use the lines in the data sets where the transition type contain a payment, i.e lines where transType is equal to 'P' or 'TP' in Table 2. During the 5-fold cross-validation, the payment models are trained on the training portion and evaluated on each holdout set, and we compute the root mean square error (RMSE) and median absolute error (MAE) between the true and predicted payments. For each fold, we set $b_2 = 0$, and other splitting points for the

payment distribution are estimated using the Gerstengarbe plot, as explained in Section 4.3. Specifically, the splitting points b_1^j and b_{L-1}^j that determine the GPD tail regions are re-estimated on the training portion of each fold only; the held-out fold is not used at any stage of the threshold selection procedure. The fixed point $b_2^j = 0$ is set by construction and carries no information from the data. In this way, the four bins can be interpreted respectively as small and large negative or positive payments. From Table 6, we observe that the center of the payment distribution is well captured as the MAE is between 911 and 2,323 euro, which is to be compared to the summary statistics of the payment distribution shown in Appendix I. We observe furthermore that the payment distributions in data sets for states S_0 and S_{5+} are the most difficult to model, given the nonhomogeneity of claims in those data sets. We also add probability-probability (PP) plots for the fitted distribution in Appendix J, where we observe an acceptable fit of the payment distributions, especially in the tails, which are of particular interest in a reserving context.

Table 5. Number of observations in each data set to model the payment process.

	State					
	S_0	S_1	S_2	S_3	S_4	S_{5+}
abs.freq	24,480	18,465	10,062	5,849	3,941	19,084

Table 6. 5-fold CV RMSE and MAE of the payment distributions in each data set.

	State					
	S_0	S_1	S_2	S_3	S_4	S_{5+}
RMSE	9,561	4,946	4,494	5,238	4,391	8,008
MAE	1,668	1,060	911	1,193	1,478	2,323

6.5. Interpreting marginal effects of covariates on the payment models

We now look at the marginal effects of the covariates on the prediction of either a large or small negative payment, represented by the bins B_1 and B_2, respectively. Similarly, we also look at the probabilities of obtaining a small or large positive payment, represented by the bins B_3 and B_4. Table 7 shows the splitting points for the payment distribution in each data set, and in Table 8, the predicted mean payment of each bin is shown, estimated as explained in Section 4.3.

Table 7. Splitting points (b_l^j) for the payment distribution in each data set.

	State					
	S_0	S_1	S_2	S_3	S_4	S_{5+}
b_1	$-1,230$	$-3,000$	$-2,310$	$-1,780$	$-2,140$	$-2,690$
b_2	0	0	0	0	0	0
b_3	3,500	3,200	2,970	3,107	2,500	2,530

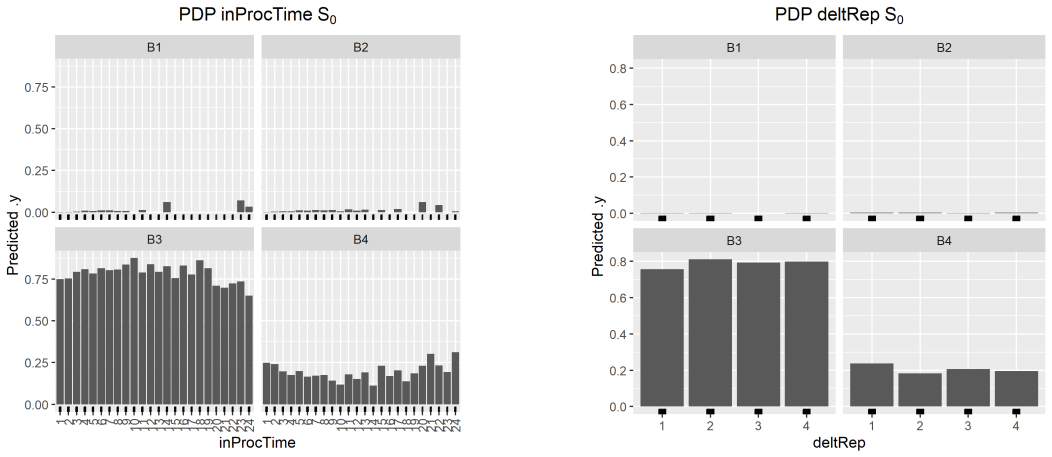
From Figure 10, we observe that the effect of the time spent in the process together with the reporting delay has little impact on the probability of having a small or large payment. Note in Table 9 that almost all payments in S_0 are positive, hence pertaining to B_3 or B_4.

Table 8. Mean payment (μ_l) in bins for each data set.

	State					
	S_0	S_1	S_2	S_3	S_4	S_{5+}
B_1	−6,043	−11,725	−10,644	−8,024	−10,047	−13,725
B_2	−475	−1,248	−800	−670	−788	−993
B_3	1,404	1,152	1,070	1,094	948	909
B_4	7,230	7,510	6,915	7,400	7,004	9,914

Table 9. Percentage of claims in each bin per state in the training data.

	State					
TransType	S_0	S_1	S_2	S_3	S_4	S_{5+}
B_1	0.3 %	6.9 %	3.5%	3.5%	1.8%	1%
B_2	0.4 %	17.5 %	9.3%	6.1%	4.8%	2.5%
B_3	76.2 %	60.9 %	71.9%	72.7%	70.2%	66.3%
B_4	23.1 %	14.7 %	15.3%	17.7%	23.2%	30.2%



(a) **(b)**

Figure 10. Partial dependence plots representing the marginal effect on probabilities to belong to a payment bin of the time spent in the process and the reporting delay for claims in S_0.

In S_1 we see that about 25% of all payments are negative. From Figure 11, we can see that the different covariates do have a small, yet much bigger effect than for S_0. The most likely payment is a small positive payment. Furthermore, we observe a probability of around 40% for a large positive previous payment to lead to a large payment of either sign.

From Figure 12, we observe that claims that stay longer in the process tend to have a higher probability of a positive payment. This relates to the fact that these are longer tailed claims that tend to cost more. We also observe that claims with a high cumulative previous payment tend to produce higher payments, hence claims that are costly early in development tend to stay costly as development progresses. The marginal effect of covariates for states S_3, S_4, and S_{5+} are similar to those of state S_2 and are shown in Appendix G.

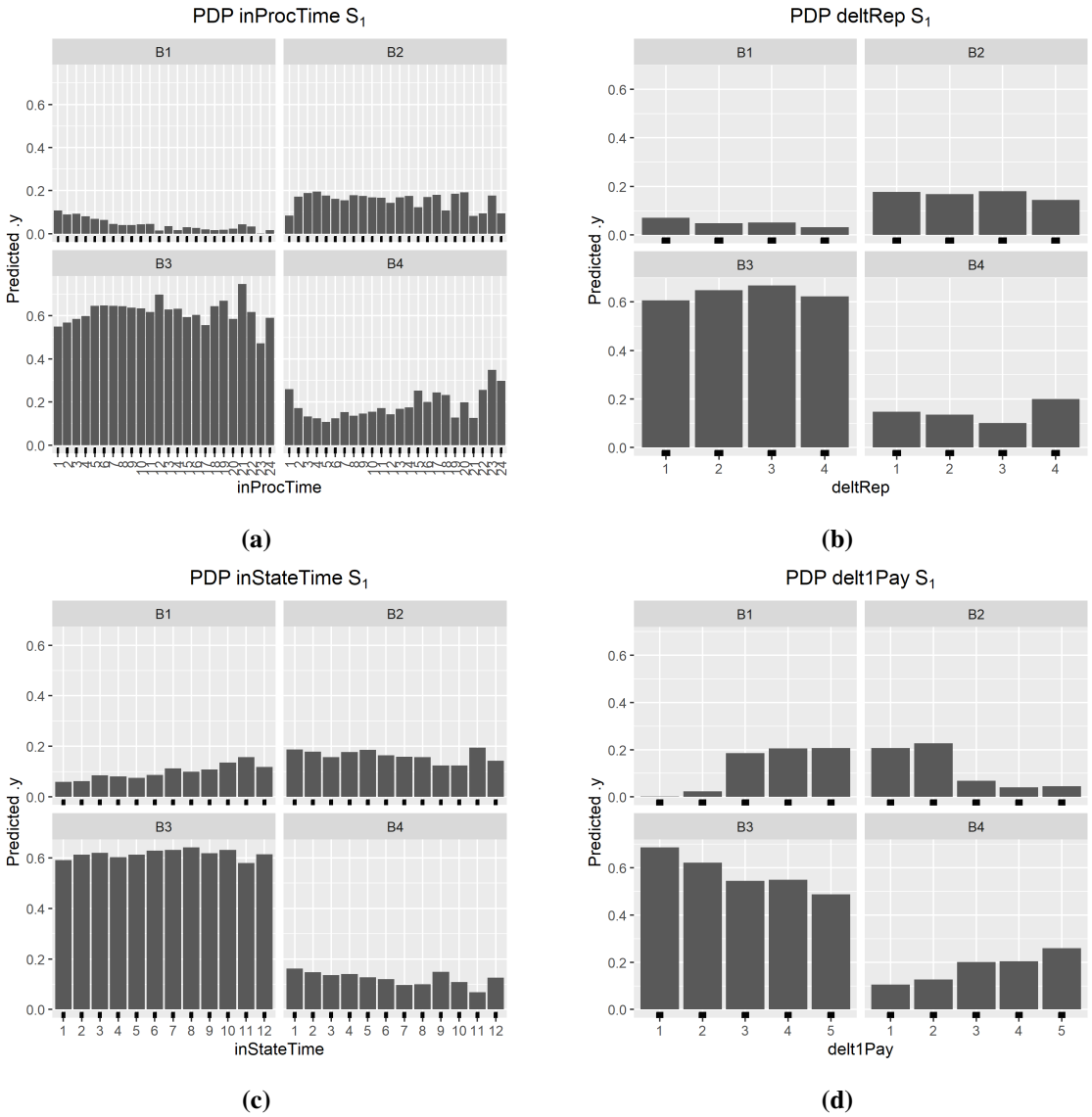


Figure 11. Partial dependence plots representing the marginal effect on probabilities to belong to a payment bin of the time spent in the process (a), the reporting delay (b), the time spent in the state (c), and the previous payment size (d) for claims in S_1.

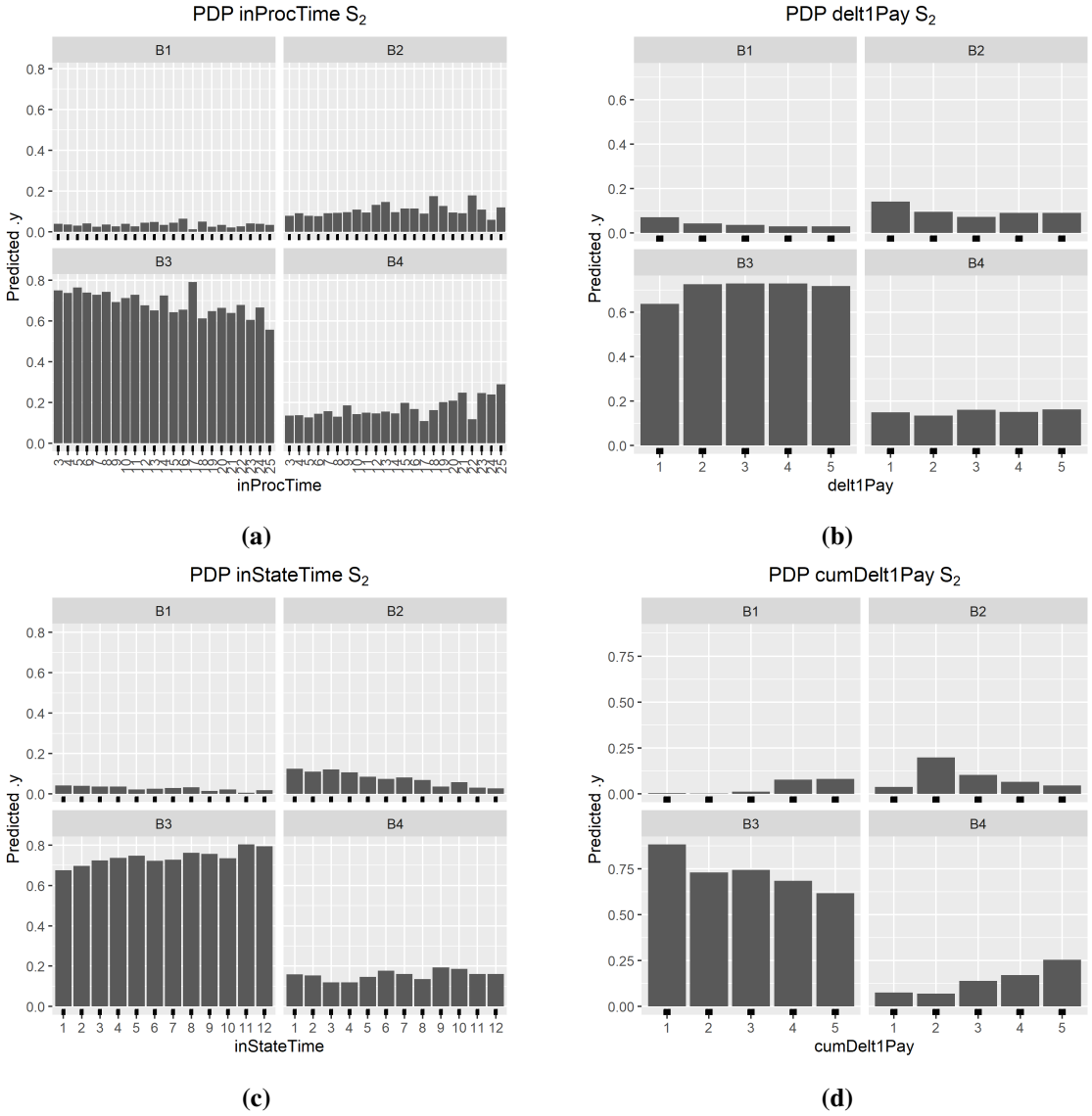


Figure 12. Partial dependence plots representing the marginal effect on probabilities to belong to a payment bin of the time spent in the process (a), the previous payment size (b), the time spent in the state (c), and the cumulative previous payment size (d) for claims in S_2.

6.6. IBNR count comparison with the chain-ladder

We start by computing the yearly IBNR claim counts based on the methodology presented in Section 3, i.e. the approach of [4]. To this end, a bootstrap sample of a sufficient size is taken (1000) to obtain the mean and 95% quantiles of these IBNR predictions, which are shown in Table 10. For each accident year, the mean of these IBNR counts are used to build the IBNR data set as explained in Appendix B.

Clearly, most of the IBNR claims come from the last observed accident year, as these are bodily injury claims, which are typically reported rather fast. As mentioned in Section 3, the average number of IBNR claims predicted by the approach of [4] corresponds to the number of IBNR claims predicted by the chain-ladder method. However, these IBNR predictions underestimate the number of IBNR claims

in later accident years. This is due to a change in the pattern of the occurrence and reporting of claims for the last two accident years as shown in Appendix H. This change in pattern creates a deviation from the assumptions of the model for the later accident years. When recomputing for the number of IBNR claims after having removed the accidents from years 2011 and 2012, the estimated number of IBNR claims was observed to be much closer to the ground truth.

A formal test for pattern changes in IBNR counts can be conducted via a chi-square goodness-of-fit test on the Poisson regression residuals, or by testing for a structural break in the accident year coefficients. However, with the available data, such tests have limited power to detect gradual trend changes in the most recent accident years, which are precisely the years most relevant for IBNR estimation. We recommend that practitioners complement formal tests with visual inspection of the residual patterns, as illustrated in Appendix H.

Table 10. Predicted yearly IBNR claim counts.

	2006	2007	2008	2009	2010	2011	2012	Total
Database	0	0	0	0	2	29	283	314
CL mean	0	0	1	1	3	15	173	193
CL 95%	0	0	4	4	8	26	213	233

6.7. Comparison with other micro-reserving models

In this section, we compare the performance of the proposed methodology on an individual level to the multi-state model of [21] and the hierarchical GLM of [37]. The goal is to see how well the predictive distributions of the reserves capture the true reserves for the RBNS claims. Let $\hat{q}_k^\alpha$ denote the α-quantile of the predictive distribution of $R_{k,\tau}$, consisting of N_{sim} possible reserve values $\hat{R}_{k,\tau}^1, \ldots, \hat{R}_{k,\tau}^{N_{sim}}$. Then, we can define the following measures:

- The interval score, IS := $mean(\hat{q}_k^\alpha - \hat{q}_k^{1-\alpha})$, measures the width of the prediction intervals.
- The prediction interval coverage probability, PICP = $mean(\mathbb{1}_{R_{k,\tau} \in [\hat{q}_k^{1-\alpha}, \hat{q}_k^\alpha]})$, represents which fraction of the true reserves falls in the prediction intervals of the different reserving methods.
- The continuous ranked probability score (CRPS) of [38] is a strictly proper scoring rule, which assesses the quality of probabilistic forecasts and rewards the forecaster for honest estimation of the predictive distribution. We use the formulation from [39], given by

$$CRPS(R_{k,\tau}) = \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} \mid \hat{R}_{k,\tau}^i - R_{k,\tau} \mid - \frac{1}{2N_{sim}^2} \sum_{i=1}^{N_{sim}} \sum_{r=1}^{N_{sim}} \mid \hat{R}_{k,\tau}^i - \hat{R}_{k,\tau}^r \mid,$$

  where a lower score for CRPS represents a more accurate probabilistic forecast.

From Table 11, we observe that the predictive distribution produced by mCube provides a better representation of the true observed reserves than the two other micro-reserving models. mCube obtains the lowest mean CRPS, representing that the predictive distribution for the RBNS claims are the most accurate. However, the method from [21] obtains the lowest median CRPS, representing the fact that it is more suited for the claims with a lower reserve amount. Moreover, the prediction interval coverage probabilities are the highest for mCube, although none of the methods have the required coverage of 95% or 99%.

Table 11. CRPS, IS, and PICP for the RBNS claims of the 3 competing micro-reserving methods on the subset of Allianz bodily injury claims.

	mCube	Bettonville	Crevecoeur
mean_CRPS	**11**, **148**.**82**	14,166.51	15,405.99
median_CRPS	4,077.85	**2**, **377**.**34**	4,251.77
PICP_99	**0**.**60**	0.37	0.12
IS_99	101,642.41	**385**, **729**.**23**	71,808.31
PICP_95	**0**.**57**	0.29	0.11
IS_95	66,168.42	**66**, **284**.**85**	61,955.80

We also compare the observed reserves with the mean of the predictive distributions of the methods. To this end, we use the following pointwise accuracy measures, where $R_{k,\tau}$ represents the true observed reserve and $\hat{R}_{k,\tau}$ the mean of the predictive distribution obtained from the methods:

- Bias := $\sum_k (R_{k,\tau} - \hat{R}_{k,\tau})$.
- Mean absolute error (MAE) := $mean(\mid \hat{R}_{k,\tau} - R_{k,\tau} \mid)$.
- Root mean square error (RMSE) := $\sqrt{mean(\mid \hat{R}_{k,\tau} - R_{k,\tau} \mid^2)}$.
- Symmetric mean absolute percentage error (sMAPE):= $mean(200\times \mid R_{k,\tau} - \hat{R}_{k,\tau} \mid /(R_{k,\tau} + \hat{R}_{k,\tau}))$.

Table 12. Pointwise accuracy measures for the individual RBNS reserves on the subset of Allianz bodily injury claims.

	mCube	Bettonville	Crevecoeur
bias	-582.34	2,145.68	**−130**.**53**
MAE	**16**, **089**.**18**	24,746.39	21,442.60
RMSE	**41**, **119**.**29**	100,108.94	48,911.23
sMAPE	**1**.**30**	1.63	1.59

From Table 12, we observe that mCube produces the best pointwise forecast for most of the metrics under consideration. We remark however that the method of [37] has the lowest absolute bias. These findings highlight the superiority of mCube with respect to the other methods under consideration for the data set of Allianz bodily injury claims.

6.8. Best estimate comparison with the chain-ladder and other micro-reserving models

In this section, we compare the performance of mCube to the chain-ladder and other micro-reserving models for predicting best estimate reserves on the same Allianz data set. To obtain the reserve predicted by mCube, we follow Sections 4.4 and 4.5. Given that most claims are reported within one month of their occurrence, as shown in Figure 6c, we use $\hat{\lambda}_{oc,0} = 1$ in Equation (5). This assumption is reasonable, since in Figure 6 (c), around 90% of the claims have a reporting delay less than 1 month and 99% of the claims have a reporting delay of less than 3 months. Furthermore, when performing the logistic regression of Equation 5, we observe that the probability of being reported in the first month is at least 85%. From Figure 13, we observe the good performance of mCube on the prediction of the reserves, as the true reserve is near the center of the best estimate distribution. In particular, from

Table 13, we observe that we perform very well on the RBNS reserves, but less well on the IBNR reserves, given that we do not model the number of IBNR claims sufficiently well, as explained in the previous section. The chain-ladder however underperforms on this data set, showing the added value of our proposed methodology. The predicted reserve by mCube is 73,916,247, giving a percentage error (PE) of 0.37%, whereas the chain-ladder predicts 62,433,801, giving a percentage error of -15%. This represents the reserve for both IBNR and RBNS claims. As mentioned in Appendix H, the reason why the Chain-Ladder under-reserves is due to the fact that its assumptions are not satisfied by the data under consideration. In Appendix H, we recompute the reserve of the chain-ladder after removing years 2011–2012 and using the year 2010 as reference. In this reduced data set, the chain-ladder does not under-reserve and produces closer reserves to the true observed reserve. Given that the method from [37] does not produce IBNR claims, we choose to simulate reserves of IBNR claims for [21] and [37] using the methodology explained in Section 3. The methods produce reserves of 58,104,262 and 78,436,112, given a percentage error of, respectively, 21% and 7%.

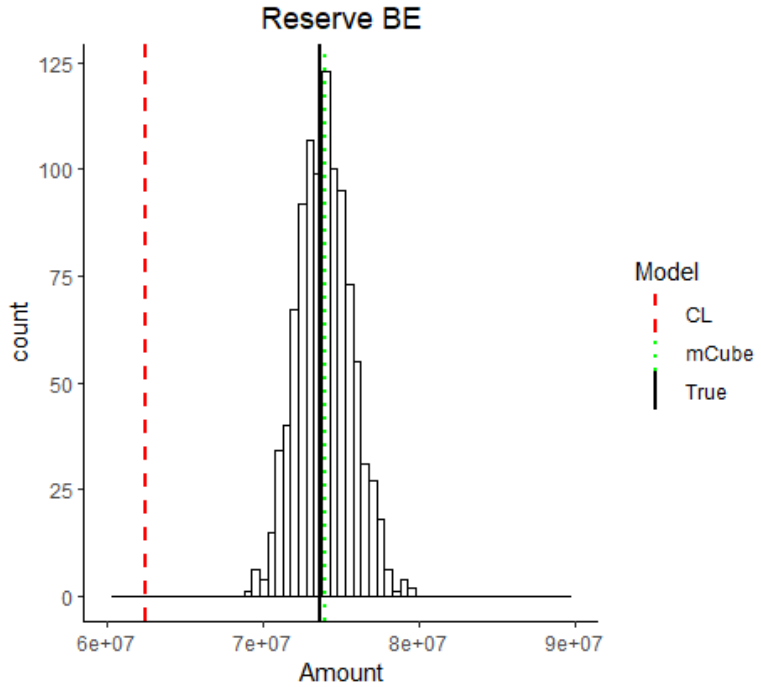


Figure 13. Best estimate distribution for all claims (RBNS and IBNR).

Table 13. Observed reserves and mean of the predicted reserves for the subset of bodily injury claims.

	Database	mCube	chain-ladder	Bettonville	Crevecoeur
RBNS reserve	69,837,157	72,689,539	-	57,628,135	68,725,695
IBNR reserve	3,808,605	1,226,708	-	476,127	9,710,417
Total reserve	73,645,764	73,916,247	62,433,801	58,104,262	78,436,112
PE	0	0.37%	−15%	−21%	7%

7. Simulation study on synthetic data

To assess the generalisability of mCube beyond the case study of Section 6.1 and to provide a fully reproducible benchmark, we conduct a simulation study using the `SynthETIC` individual claims simulator [40].

7.1. Design

`SynthETIC` generates complete individual claim histories — occurrence, notification, partial payments, and settlement — according to a parametric data-generating process for general insurance claims. Since the full claim development is known, we can evaluate mCube's predictive accuracy at the individual claim level by comparing simulated future costs to the true future costs. We consider five scenarios that vary the characteristics of the claims portfolio:

1. **Base case:** Default `SynthETIC` parameters with a reference claim size of 200,000 and quarterly time units, generating approximately 3,300 claims over a 10-year observation window.
2. **Heavy tails:** The reference claim size is doubled to 400,000, shifting the entire claim size distribution upward and producing heavier-tailed payments.
3. **Long development:** Settlement delays are multiplied by a factor of 1.5, resulting in slower claim closure and a larger proportion of open claims at the evaluation date.
4. **High frequency:** The claim frequency is doubled from 0.03 to 0.06 per unit of exposure, generating approximately 6,600 claims. This tests scalability and the effect of a larger training set.
5. **Non-stationary:** The claim frequency increases linearly from 0.02 to 0.04 over the observation window, violating the stationarity assumption implicit in the chain-ladder method.

Each scenario is replicated 50 times with different random seeds. For each replication, both mCube and the Mack chain-ladder method [2] are applied.

7.2. Validation approach

We employ the *opened-closed-claims* validation strategy, which is the natural evaluation framework for micro-level reserving models. For each replication:

1. The mCube pipeline (data preparation, multinomial time model, spliced payment model) is fitted on the upper triangle, i.e., all claim information available up to the evaluation date.
2. A random sample of 400 *closed* claims is selected.
3. Each sampled claim is "reopened" at a randomly chosen point in its development history. A random transition and a random time within that transition are selected, and the claim is treated as if it were still open at that point.
4. From the reopened state, $N_{\text{sim}} = 50$ future trajectories are simulated using Algorithm 1.
5. The mean simulated future cost $\hat{R}_k$ is compared to the true observed future cost R_k for each reopened claim k.

This approach avoids the selection bias inherent in evaluating only genuinely open claims at the evaluation date, which are disproportionately the large, slow-developing claims. By reopening closed claims at random points, the evaluation set is representative of the full claims population.

For the chain-ladder comparison, we construct an annual paid triangle from each replication and apply the Mack method [2] to estimate the total outstanding reserve. The chain-ladder percentage error is computed against the true total reserve (RBNS + IBNR).

7.3. Adaptations for synthetic data

The `SynthETIC` data-generating process produces claims with a simpler covariate structure than the real insurance portfolio of Section 6.1. In particular, the payment covariates (previous payment size, cumulative payment) are highly collinear with the time spent in the process, leading to overfitting in the multinomial time model when the full covariate set is used. We therefore restrict the time model to use only the time spent in the current state (`inStateTimeTrans`) as a predictor, while the payment model retains the full covariate set. This adaptation is data-driven: A practitioner applying mCube to a new portfolio should assess covariate relevance through the standard model diagnostics (partial dependence plots, cross-validated prediction accuracy) described in Sections 6.2–6.4.

7.4. Results

Table 14 summarises the results across all five scenarios. The chain-ladder method overestimates the total reserve by 37–46% on average across all scenarios, with substantial variability (standard deviation 15–22%). This systematic overestimation is expected: The chain-ladder method estimates the total reserve (RBNS + IBNR) from aggregate triangles, and does not distinguish between the two components. Notably, the non-stationary scenario produces a lower chain-ladder PE (37%), consistent with the fact that increasing claim frequency in later years inflates the lower-right portion of the triangle, partially offsetting the typical upward bias.

For the claim-level validation of mCube, the concordance probability [41] — the proportion of claim pairs for which mCube correctly identifies the more expensive claim — ranges from 0.630 to 0.676 across scenarios, consistently and significantly above the uninformative baseline of 0.5. The Spearman rank correlation between predicted and observed future costs is 0.39–0.50, confirming a moderate to strong positive association at the individual claim level. These metrics demonstrate that mCube captures meaningful heterogeneity in claim severity, which is a capability that aggregate methods such as chain-ladder do not possess.

The highest claim-level discrimination is observed in the high-frequency scenario (concordance 0.676, Spearman 0.502), where the doubled portfolio size provides a richer training set. This suggests that mCube's predictive accuracy scales favourably with portfolio size, which is an encouraging property for practical applications involving large insurance portfolios.

The mean bias of mCube's claim-level predictions is −12% to −36% depending on the scenario, indicating a systematic underestimation. This underestimation is a known consequence of right-censoring in the training data: The payment model is fitted on completed transitions, which underrepresent the large payments from claims that are still developing at the evaluation date. The censoring effect is strongest in the high-frequency scenario (bias −36%, aggregate PE −34%), where the larger number of open claims at the evaluation date amplifies the bias. Conversely, the

long-development and non-stationary scenarios exhibit the smallest aggregate percentage errors (−4% and −5%, respectively), because the slower settlement or increasing frequency produce a richer set of completed transitions for model fitting.

The aggregate percentage error of mCube on the reopened claims ranges from −5% to −34%, which in four out of five scenarios is substantially smaller in magnitude than the chain-ladder overestimation. Moreover, the concordance and Spearman metrics are stable across scenarios (standard deviations of 0.01–0.02 and 0.02–0.03, respectively), indicating that the claim-level discrimination is robust to changes in the data-generating process.

Table 14. Simulation study results: Claim-level validation of mCube on `SynthETIC` data using the opened-closed-claims approach.

	Chain-Ladder		**mCube (claim-level)**				
Scenario	PE%	SD	Conc.	Spear.	Bias%	MdAPE%	Agg.PE%
Base case	45.6	15.7	0.640	0.405	−23.8	75.0	−18.5
Heavy tails	44.2	17.4	0.636	0.397	−27.1	74.5	−22.1
Long development	43.3	22.1	0.630	0.386	−11.9	85.8	−3.9
High frequency	41.0	14.8	0.676	0.502	−35.7	75.8	−33.7
Non-stationary	37.2	18.8	0.651	0.436	−12.9	76.1	−4.8

Note: For each scenario, 400 closed claims are reopened at a random point in their development and their future cost is simulated via Algorithm 1 with $N_{\text{sim}} = 50$ trajectories per claim. The concordance probability (Conc.) and Spearman correlation (Spear.) measure claim-level discrimination; the mean bias and aggregate percentage error (Agg. PE) measure reserve accuracy. Chain-ladder PE is computed against the true total reserve. Each scenario is replicated 20 times.

7.5. Discussion

The simulation study confirms that mCube generalises to data generated by an independent claims simulator, without requiring access to the proprietary portfolio used in the main case study. The key finding is that mCube provides *claim-level* predictions with a concordance probability of 0.63–0.68 — a capability that is absent from aggregate reserving methods. This is particularly relevant in the context of Solvency II, where individual claim information is increasingly used for risk management and capital allocation. The non-stationary scenario further demonstrates that mCube handles time-varying claim frequencies naturally, whereas the chain-ladder method relies on implicit stationarity assumptions.

The systematic underestimation observed in the claim-level predictions is attributable to the right-censoring structure of the training data and could be addressed by the calibration methodology, which recentres and rescales the predictive distribution using a held-out calibration set. We leave the integration of such a calibration step into the `SynthETIC` simulation framework for future work.

We note that the `SynthETIC` data-generating process has a simpler covariate structure than a typical real insurance portfolio, which limits the discriminative power of the model. With richer covariate information — such as the detailed payment histories and reporting delays available in the case study of Section 6.1 — the concordance probability and Spearman correlation would be expected to improve.

7.6. Sensitivity to hyperparameter choices

To assess the predictive sensitivity and aggregate error of mCube to the choice of hyperparameters, we vary two key settings on the base case scenario: `nMinLev`, the minimum number of observations per bin during the discretisation of continuous predictors, and `nMinMod`, the minimum number of observations required to fit a multinomial model with covariates. Table 15 reports the claim-level metrics across the different settings. The results are completely insensitive to `nMinMod`: Varying it from 200 to 800 produces identical metrics, as the training data comfortably exceeds all thresholds. For `nMinLev`, we observe a bias-variance trade-off: Coarser binning (higher `nMinLev`) improves claim-level discrimination (concordance from 0.640 to 0.685, Spearman from 0.405 to 0.531) by reducing overfitting, but increases aggregate bias (from 18.5% to 31.0%) due to loss of granularity in payment predictions. This is an expected trade-off rather than instability: The practitioner can choose `nMinLev` to balance ranking accuracy against aggregate reserve accuracy depending on the application. The value `nMinLev` = 30 used in the case study of Section 6 represents a balanced default.

Table 15. Sensitivity of mCube claim-level metrics to hyperparameter choices on the base case scenario (20 replications, $N_{sim} = 50$).

nMinLev	nMinMod	Conc.	Spear.	Bias%	Agg.PE%
15	200	0.640	0.405	−23.8	−18.5
30	200	0.669	0.487	−29.8	−26.1
50	200	0.685	0.531	−33.4	−31.0
15	500	0.640	0.405	−23.8	−18.5
15	800	0.640	0.405	−23.8	−18.5

Note: The default setting used in the simulation study is `nMinLev` = 15 and `nMinMod` = 200. The case study of Section 6 uses `nMinLev` = 30.

8. Conclusion

In this article, we have presented a multinomial multi-state micro-level (mCube) model to estimate the reserves of IBNR and RBNS claims. We present a semi-parametric model of the payment distribution, taking into account claim-specific information. On a portfolio level, in the real data study, mCube produces a portfolio level best estimate distribution whose center lies close to the observed reserve. Moreover, the estimates on an individual level are very accurate. In the synthetic study, however, the method exhibits some systematic underestimation at claim level due to right censoring, which motivates future calibration work.

A limitation of the current IBNR model is its sensitivity to nonstationarity in claim reporting patterns. As observed in Section 6.6, the pattern shift in accident years 2011–2012 led to underestimation of IBNR claim counts. A natural extension is to incorporate a Bornhuetter-Ferguson adjustment [3], which blends the data-driven Poisson estimates with an external prior on expected claim counts. This would reduce sensitivity to recent structural breaks while preserving the multinomial framework. The simulation study in Section 7 demonstrates that mCube handles moderate forms of nonstationarity (linearly increasing claim frequency) well, with an aggregate percentage error of only −4.8%. However, more extreme structural breaks would benefit from the Bornhuetter-Ferguson extension.

Future studies could replace the multinomial models used for the time and payment processes with more flexible machine learning models to obtain a higher predictive power, however, mCube strikes a nice balance between computational efficiency and predictive accuracy. In addition, as can be seen from Table 10, our proposed IBNR model did not handle well the deviation from stationarity that was observed for the two most recent accident years included in the training set. A possible future development might by combining a methodology that automatically detects pattern changes in a reserve triangle on an accident year level, with an extension of our proposed IBNR model by the Bornhuetter-Ferguson method to correct for this violation of stationarity.

Author contributions

Emmanuel Jordy Menvouta: Conceptualization and writing original draft; Emmanuel Jordy Menvouta and Robin Van Oirbeek: Programming; Robin Van Oirbeek and Tim Verdonck: Validation and revising, Methodology and supervision.

Use of Generative-AI tools declaration

In the preparation of this work, the authors did not use Generative AI tools such as ChatGPT at any point.

Acknowledgements

The authors gratefully acknowledge the financial support from the Allianz Research Chair *Prescriptive business analytics in insurance* at KU Leuven.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. S. Dreksler, C. Allen, A. Akoh-Arrey, J. A. Courchene, B. Junaid, J. Kirk, et al., Solvency II technical provisions for general insurers, *Br. Actuar. J.*, **20** (2015), 7–129. https://doi.org/10.1017/S1357321715000021
2. T. Mack, Distribution-free calculation of the standard error of chain ladder reserve estimates, *ASTIN Bull.*, **23** (1993), 213–225. https://doi.org/10.2143/AST.23.2.2005092
3. R. L. Bornhuetter, R. E. Ferguson, The actuary and IBNR, *Proc. Casualty Actuar. Soc.*, **59** (1972), 181–195.
4. P. D. England, R. J. Verrall, Stochastic claims reserving in general insurance, *Br. Actuar. J.*, **8** (2002), 443–518. https://doi.org/10.1017/S1357321700003809
5. M. V. Wüthrich, M. Merz, *Stochastic Claims Reserving Methods in Insurance*, volume 435, John Wiley & Sons, New Jersey, 2008. https://doi.org/10.1002/9780470517819

6. H. Liu, R. Verrall, Predictive distributions for reserves which separate true IBNR and IBNER claims, *ASTIN Bull.*, **39** (2009), 35–60. https://doi.org/10.2143/AST.39.1.2038055

7. M. Merz, M. V. Wüthrich, Paid–incurred chain claims reserving method, *Insur. Math. Econ.*, **46** (2010), 568–579. https://doi.org/10.1016/j.insmatheco.2010.02.004

8. G. Quarg, T. Mack, Munich chain ladder: A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses, *Variance*, **2** (2008), 266–299.

9. T. Verdonck, M. Debruyne, The influence of individual claims on the chain-ladder estimates: Analysis and diagnostic tool, *Insur. Math. Econ.*, **48** (2011), 85–98. https://doi.org/10.1016/j.insmatheco.2010.10.001

10. E. Arjas, The claims reserving problem in non-life insurance: Some structural ideas, *ASTIN Bull.*, **19** (1989), 139–152. https://doi.org/10.2143/AST.19.2.2014905

11. R. Norberg, Prediction of outstanding liabilities in non-life insurance, *ASTIN Bull.*, **23** (1993), 95–115. https://doi.org/10.2143/AST.23.1.2005103

12. M. Pigeon, K. Antonio, M. Denuit, Individual loss reserving with the multivariate skew normal framework, *ASTIN Bull.*, **43** (2013), 399–428. https://doi.org/10.1017/asb.2013.20

13. K. Antonio, R. Plat, Micro-level stochastic loss reserving for general insurance, *Scand. Actuar. J.*, **2014** (2014), 649–669. https://doi.org/10.1080/03461238.2012.755938

14. F. Duval, M. Pigeon, Individual loss reserving using a gradient boosting-based approach, *Risks*, **7** (2019), 79. https://doi.org/10.3390/risks7030079

15. Ł. Delong, M. V. Wüthrich, Neural networks for the joint development of individual payments and claim incurred, *Risks*, **8** (2020), 33. https://doi.org/10.3390/risks8020033

16. Y. Guo, W. Zhou, C. Luo, C. Liu, H. Xiong, Instance-based credit risk assessment for investment decisions in P2P lending, *Eur. J. Oper. Res.*, **249** (2016), 417–426. https://doi.org/10.1016/j.ejor.2015.05.050

17. Y. Guo, Y. Zhai, S. Jiang, Investment decision making for large-scale peer-to-peer lending data: A Bayesian neural network approach, *Int. Rev. Financ. Anal.*, 102 (2025), 104100. https://doi.org/10.1016/j.irfa.2025.103975

18. C. A. Hachemeister, A stochastic model for loss reserving, *Trans. 21st Int. Congr. Actuar.*, **1** (1980), 185–194.

19. O. Hesselager, A Markov model for loss reserving, *ASTIN Bull.*, **24** (1994), 183–193. https://doi.org/10.2143/AST.24.2.2005064

20. K. Antonio, E. Godecharle, R. Van Oirbeek, A multi-state approach and flexible payment distributions for micro-level reserving in general insurance, SSRN Electron. J., 2016. https://doi.org/10.2139/ssrn.2777467

21. C. Bettonville, L. d'Oultremont, M. Denuit, J. Trufin, R. Van Oirbeek, Matrix calculation for ultimate and 1-year risk in the semi-Markov individual loss reserving model, *Scand. Actuar. J.*, **2021** (2021), 380–407. https://doi.org/10.1080/03461238.2020.1792539

22. A. L. Badescu, X. S. Lin, D. Tang, A marked Cox model for the number of IBNR claims: Theory, *Insur. Math. Econ.*, **69** (2016), 29–37. https://doi.org/10.1016/j.insmatheco.2016.03.016

23. M. Maciak, O. Okhrin, M. Pesta, Infinitely stochastic micro reserving, *Insur. Math. Econ.*, **100** (2021), 30–58. https://doi.org/10.1016/j.insmatheco.2021.04.007

24. P. D. Allison, Discrete-time methods for the analysis of event histories, *Sociol. Methodol.*, **13** (1982), 61–98. https://doi.org/10.2307/270718

25. W. N. Venables, B. D. Ripley, *Modern Applied Statistics with S*, fourth edition, Springer, New York, 2002. https://doi.org/10.1007/978-0-387-21706-2

26. E. W. Frees, E. A. Valdez, Hierarchical insurance claims modeling, *J. Am. Stat. Assoc.*, **103** (2008), 1457–1469. https://doi.org/10.1198/016214508000000823

27. T. Reynkens, R. Verbelen, J. Beirlant, K. Antonio, Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions, *Insur. Math. Econ.*, **77** (2017), 65–77. https://doi.org/10.1016/j.insmatheco.2017.08.005

28. J. Beirlant, G. Dierckx, A. Guillou, Estimation of the extreme-value index and generalized quantile plots, *Bernoulli*, **11** (2005), 949–970. https://doi.org/10.3150/bj/1137421635

29. L. Montuelle, E. Le Pennec, Mixture of Gaussian regressions model with logistic weights, a penalized maximum likelihood approach, *Electron. J. Stat.*, **8** (2014), 1661–1695. https://doi.org/10.1214/14-EJS939

30. J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, *Statistics of Extremes: Theory and Applications*, John Wiley & Sons, Chichester, 2004. https://doi.org/10.1002/0470012382

31. Ł. Delong, M. Lindholm, M. V. Wüthrich, Collective reserving using individual claims data, *Scand. Actuar. J.*, **2022** (2022), 1–28. https://doi.org/10.1080/03461238.2021.1921836

32. A. Gabrielli, An individual claims reserving model for reported claims, *Eur. Actuar. J.*, **11** (2021), 541–577. https://doi.org/10.1007/s13385-021-00271-4

33. R. Henckaerts, K. Antonio, M. Clijsters, R. Verbelen, A data driven binning strategy for the construction of insurance tariff classes, *Scand. Actuar. J.*, **2018** (2018), 681–705. https://doi.org/10.1080/03461238.2018.1429300

34. J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Stat.*, **29** (2001), 1189–1232. https://doi.org/10.1214/aos/1013203451

35. C. Molnar, B. Bischl, G. Casalicchio, iml: An R package for interpretable machine learning, *J. Open Source Softw.*, **3** (2018), 786. https://doi.org/10.21105/joss.00786

36. C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2019. Available at: `https://christophm.github.io/interpretable-ml-book/`.

37. J. Crevecoeur, J. Robben, K. Antonio, A hierarchical reserving model for reported non-life insurance claims, *Insur. Math. Econ.*, **104** (2022), 158–184. https://doi.org/10.1016/j.insmatheco.2022.02.005

38. T. Gneiting, A. E. Raftery, Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, **102** (2007), 359–378. https://doi.org/10.1198/016214506000001437

39. E. P. Grimit, T. Gneiting, V. J. Berrocal, N. A. Johnson, The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification, *Q. J. R. Meteorol. Soc.*, **132** (2006), 2925–2942. https://doi.org/10.1256/qj.05.235

40. B. Avanzi, G. Taylor, M. Wang, B. Wong, SynthETIC: An individual insurance claim simulator with feature control, *Insur. Math. Econ.*, **100** (2021), 296–308. https://doi.org/10.1016/j.insmatheco.2021.06.004

41. M. J. Pencina, R. B. D'Agostino, Overall c as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation, *Stat. Med.*, **23** (2004), 2109–2123. https://doi.org/10.1002/sim.1802

AIMS Press