

Applications of population sampling to IBNR reserving: Extending and blending the Chain-Ladder with individual reserving methods

Sebastián Calcetero Vanegas, X. Sheldon Lin, Andrei L. Badescu

Department of Statistical Sciences

University of Toronto

Toronto, Ontario

Abstract

This technical report outlines the major findings in the preprint Calcetero-Vanegas et al. (2025), which was partially supported by the Canadian Institute of Actuaries, grant number CS000269. Briefly, this work introduces a novel statistical framework for reserving methods grounded in population sampling theory. We explore how population sampling motivated tools, not explored in reserving, can be employed to enhance the robustness and effectiveness of current reserving models. The methods are illustrated with simulated data along with reproducible code to showcase the implementation of techniques.

1 Introduction

Accurate estimation of reserves of outstanding claims is fundamental to insurance operations, ensuring solvency and compliance with regulatory requirements. The reserving problem in insurance has been extensively studied in actuarial science literature and remains an active area of research due to challenges arising from the data. Outstanding claims reserves are typically divided into two categories: Reported But Not Settled (RBNS) reserves and Incurred But Not Reported (IBNR) reserves, each requiring distinct estimation methodologies. RBNS reserves generally constitute a larger portion of the total reserve but are often of less analytical interest since they can be reasonably estimated using expert judgment, e.g., case estimates, given that the severeness of the associated accidents is known. In contrast, IBNR reserves involve significantly greater uncertainty, as no information is available about the claims they encompass. This makes IBNR reserves more challenging to estimate and a primary focus for both practitioners and researchers. In this discussion, for simplicity, we will focus our analysis on the IBNR reserves. However, we note that the principles outlined here can, with minimal adjustment, be applied to RBNS reserves as well.

The insurance industry has approached claim reserving through two distinct frameworks: macro-level and micro-level methodologies, each with its own theoretical foundations and practical applications (see, e.g., Taylor (2019) for a review of reserving techniques). The macro-level, or aggregate-level, approach—exemplified by the Chain-Ladder (CL) method—estimates reserves at the portfolio level, relying on aggregated data for simplicity and ease of use. However, this approach overlooks individual claim behavior, which can result in inaccuracies. In contrast, the micro-level, or individual-level, approach, as described by Antonio and Plat (2014), utilizes claim-specific data to provide more precise reserve estimates. This precision, however, comes with increased complexity and higher computational demands. Notably, according to a survey by Francis (2016), individual reserving

methods remain virtually absent in practice, with insurance companies worldwide predominantly relying on aggregate models.

Aggregate and individual models have traditionally been built on distinct statistical foundations due to their differing data sources, often resulting in incompatible reserve estimates. This dichotomy has constrained the flexibility and adaptability of reserving methods, impeding the development of models that integrate the strengths of both approaches and slowing the adoption of advanced techniques in the insurance industry. Several studies, including Calcetero-Vanegas et al. (2024), Tee and Käärik (2022), Wahl et al. (2019), Wüthrich (2018), Hiabu (2017), and Charpentier and Pigeon (2016), have sought to address this issue by proposing methods to bridge the aggregate and individual approaches. However, these contributions are often narrowly focused on specific techniques and fail to offer a unified framework that naturally accommodates both micro and macro perspectives of the reserving process.

In response to this challenge, Calcetero-Vanegas et al. (2024) and Calcetero-Vanegas and Lin (2024) recently introduced an inverse probability weighting (IPW) estimator that integrates granular information into the CL method, providing an intermediate structure between aggregate and individual models. This approach was developed by conceptualizing the reserving problem as a population sampling issue, as discussed in Arnab (2017). Population sampling involves selecting and analyzing data subsets to infer characteristics of the entire population, often using sampling designs optimized for the population’s structure. In the context of reserving, this design is similar to Poisson sampling, where each claim is either included or excluded from the sample (i.e., reported or not) based on a Bernoulli-like trial, which depends on the reporting delay time. This process is driven by the nature of the claims, rather than being determined by the analyst, and resembles mechanisms studied in missing data, where only available records are analyzed (e.g., surveys with non-response), and in causal inference, where data is observable only for those who were treated.

While the approach by Calcetero-Vanegas et al. (2024) introduces a novel perspective on claim reserving and expands the CL method, it offers limited exploration of how the population sampling perspective can be fully applied to claim reserving problems. Additionally, it does not adequately address how this perspective can be extended to individual-level methods, thereby providing a comprehensive framework for understanding both micro and macro models. Specifically, Calcetero-Vanegas et al. (2024) focused on the IPW estimator within the so called *design-based* approach to sampling, which does not consider a probabilistic model for the data-generating process. Instead, it depends solely on the randomness of the sampling process, i.e., reporting delays. However, this approach has limitations, as it cannot incorporate distributions or other modeling insights for the claims data, e.g., heavy tails, that are often crucial and widely used in the reserving literature.

Along these lines, this report addresses the gaps in Calcetero-Vanegas et al. (2024) related to the population sampling perspective of the reserving problem, extending it to the *model-based* approach to sampling, in which distributional models for the claims data are also incorporated. This extension is achieved through augmented inverse probability weighting (AIPW) estimators, which build on the IPW framework by integrating a model that includes distributional assumptions in the predictions. We demonstrate that the AIPW estimator provides a unified statistical framework for claim reserving, integrating both aggregate and individual-level approaches. Specifically, it incorporates models such as the Chain-Ladder and Bornhuetter-Ferguson models at the aggregate level, alongside individual models where the severity model satisfies the balance property. This creates a seamless transition between the macro and micro approaches. Thus, the AIPW estimator serves as a natural bridge between aggregate and individual reserving models, enabling the simultaneous leveraging of their respective strengths.

Moreover, this report explores how the population sampling perspective broadens our understanding of the reserving process. In particular, it highlights issues related to sample bias arising from partially observed data—one of the main motivations for using AIPW estimators. Indeed, the sampling bias problem stems from the fact that the behavior of reported claims may differ from that of not reported claims. As a result, inferences based solely on reported claims data may not generalize to not reported claims, leading to poor predictions. This issue is common in insurance data but is often overlooked by actuaries, e.g., American Academy of Actuaries (2023). To address this, we adapt practical statistical approaches from the population sampling literature that can mitigate sampling bias in insurance reserving, thereby improving reserve predictions. In particular, we illustrate methods such as double-robust estimators, weighted estimating equations for model fitting, and synthetic data generation. These techniques not only enhance the precision and robustness of reserve estimates but also introduce statistical methods into the reserving process, which have not yet been used in the field. Our goal is to illustrate the application of the class of sampling-based methods and highlight their benefits for reserving, rather than showcasing the best configuration among them.

The rest of the report is organized as follows: Section 2 introduces the claim reserving problem, defines our notation, and revisits the population sampling perspective. Section 3 presents the AIPW framework as a unifying statistical tool for both aggregate and individual reserving methods. Section 4 explores additional applications of sampling methods in reserving. Section 5 demonstrates how these sampling methods can be applied to IBNR reserves using simulated data along with code, while Section 6 does it with a real Canadian auto insurance data to illustrate their practical applications. Finally, Section 7 concludes with a discussion and directions for future research.

2 A population sampling framework for claim reserving

Suppose an insurance company is evaluating its total liabilities from claims with accident times between $t = 0$ and $t = \tau$, where τ is the valuation time set by the actuary. In general insurance, there is often a delay between an accident occurrence and its reporting time to the company. Moreover, the situation becomes more complex due to payment delays. Reported claims are often paid in several installments as the accident’s impact evolves. At valuation time τ , the insurer knows only the claims that have been reported and the partial amount paid to each one. Therefore the insurer aims to estimate the total amount for not reported claims and the remaining payments for reported claims to calculate the reserve for outstanding claims.

While reported but not settled claims are of interest, insurance companies have more information regarding these as they know the graveness of the accident and can therefore provide an expert estimate of the liability, e.g. case reserves. Along those lines, we will focus on the liability of incurred but not reported claims, which are the claims the insurance company has no information. To address this, we will work with incurred losses rather than paid losses. Incurred losses consider the case reserve per claim, which represents the amount set aside by insurers for estimated unpaid costs of reported claims, based on expert judgment, current payments, and potential future developments. The sum of the case reserve and current payments forms the incurred losses, the best estimate of the total claim cost, which can be used as a reliable proxy for the value of a claim once is settled. From now on, we will work mainly with incurred losses thus avoiding the issue associated with the not settled claims. Similarly, we will refer to the incurred losses when mentioning the terms severity or claim amounts. This approach is common in reserving practices, where actuaries typically use incurred loss triangles rather than paid loss triangles to estimate IBNR reserves, and also in the

literature, e.g., Calcetero-Vanegas and Lin (2024), Fung et al. (2021).

Along those lines, let us describe the claim process as follows:

- $N(\tau)$: Number of claims with accident times before τ . Note that this is partially unobserved before τ .
- Y_i : Sequence of incurred losses per claim, where $i = 1, \dots, N(\tau)$.
- T_i : Sequence of accident times for each claim, where $i = 1, \dots, N(\tau)$.
- R_i : Sequence of associated reporting times, where $i = 1, \dots, N(\tau)$.
- $U_i = R_i - T_i$: Sequence of reporting delay times for each claim, where $i = 1, \dots, N(\tau)$.
- \mathbf{x}_i : Sequence of attributes related to the accident, claim type, policyholder, or payment characteristics, where $i = 1, \dots, N(\tau)$.
- $N^R(\tau)$: Number of claims reported by τ , and $N^{IBNR}(\tau)$: Number of incurred but not reported claims.

The total liability for accidents occurring before τ , denoted $L(\tau)$, is:

$$L(\tau) = \sum_{i=1}^{N(\tau)} Y_i.$$

The portion of liability known to the company by τ (from reported claims), denoted $L^R(\tau)$, is:

$$L^R(\tau) = \sum_{i=1}^{N^R(\tau)} Y_i.$$

One is interested in estimating the liability for not reported claims, denoted $L^{IBNR}(\tau)$, which is:

$$L^{IBNR}(\tau) = L(\tau) - L^R(\tau) = \sum_{i=1}^{N^{IBNR}(\tau)} Y_i.$$

In reserving modeling, there are mainly two approaches explored in the literature in order to have an estimation of the outstanding claims. On one hand, there is the macro-level approach, with the Chain-Ladder (CL) as example, which works with aggregate data and provide estimates without explicitly imposing a probability model for the claim data (e.g. a frequency-severity model). These methods have been seen as ad-hoc procedures lacking statistical justification, and ignore the granular information from individual claims. On the other hand, there is the micro-level approach in which a model is used for claims and their attributes using probability distributions for the frequency, severity and reporting delays. As such, an actuary can derive a distributional model for the IBNR claims, and therefore use it to produce an estimation of the outstanding claims. Nevertheless, the reserving theory has been split mostly in these two, almost mutually exclusive branches, where no single framework provides a statistically sound umbrella for both approaches.

In response, a previous paper Calcetero-Vanegas et al. (2024) illustrated a statistically justified methodology on which it is possible to introduce individual information on the CL method, producing therefore a hybrid aggregate and individual reserving method. As such, it aims to be a starting point to bridge the gap between the micro and macro model. This was achieved by a novel view of the reserving problem as a population sampling method as we review next.

2.1 Claim reserving as a population sampling problem

Claim reserving can be seen as situation in which we want to estimate a population total based solely on a sample. In this context, all $N(\tau)$ claims are treated as the population, with the $N^R(\tau)$ claims reported by the valuation date forming the sample. Here one view this as *Poisson sampling* without replacement, where each claim is a Bernoulli trial, either succeeding (reported by the valuation time) or failing. This is a similar sampling design associated to *missing data* mechanisms. Unlike typical population sampling setups, the sample in reserving is given, rather than selected by an analyst.

Along those lines, the statistical framework proceeds as follows: a membership indicator variable taking the value 1 if a claim is reported, and 0 otherwise can be defined as $\mathbf{1}_i(\tau) = \mathbf{1}_{\{R_i \leq \tau\}}$, with the corresponding inclusion probabilities

$$\pi_i(\tau) = P(R_i \leq \tau \mid \mathbf{x}_i, T_i, Y_i) = P(U_i \leq \tau - T_i \mid \mathbf{x}_i, T_i, Y_i).$$

These probabilities, which depend on the valuation time and vary by claim due to differences in accident types and risk profiles, are unknown and must be estimated from the data. Estimation can be done using techniques such as logistic regression or survival analysis, and is often part of the estimation of reserving models themselves. Indeed, since reporting delay time is a time-to-event random variable with right truncation, survival modeling approaches, such as Cox regression, e.g. Vakulenko-Lagun et al. (2019), George et al. (2014), can be used to estimate the cumulative distribution function, from which the inclusion probabilities can be derived. Proper estimation of these probabilities is crucial from the population sampling perspective as their the main input for all the methods that follow. This procedure will be further illustrated in the numerical study. From now on, we will use the notation $\hat{\pi}_i(\tau)$ to denote the estimated inclusion probabilities based on the selected model..

The IPW estimator

Based on such a sampling design, Calcetero-Vanegas et al. (2024) introduced the unbiased inverse probability weighting (IPW) estimator for the total liability, which scales the reported claims amounts by a factor of $\frac{1}{\hat{\pi}_i(\tau)}$ using the expression:

$$\sum_{i=1}^{N^R(\tau)} \frac{Y_i}{\hat{\pi}_i(\tau)}.$$

Consequently, an unbiased IPW estimator for the IBNR reserve, which utilizes the odds ratio $\frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)}$ that represent the frequency ratio of not reported to reported claims with similar characteristics, via the expression:

$$\sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} Y_i. \quad (1)$$

Although this estimator may not seem intuitive at first glance, it is a natural choice from the perspective of claim reserving as motivated below. The following proposition illustrates how the IPW principle is an ingrained component in the estimation of reserves for not reported claims using reported claims as a starting point. Nevertheless, recall that the IPW principle arises on its own from the population sampling literature, rather than from the claim reserving problem. This proposition serves merely as an illustration and provides an initial result on how population sampling techniques, such as IPW, connect to claim reserving and contribute to a reserving framework.

Proposition 1. Let $\{N(t; \tau), t \leq \tau\}$ be a counting process representing the number of claims occurring before the valuation date τ , with an expected count function given by $\Lambda(t) = E(N(t; \tau))$. Each claim is associated with time-dependent marks $(T_i, Y_i, \mathbf{1}_i(\tau))$. Suppose that the marks Y_i and $\mathbf{1}_i(\tau)$ are conditionally uncorrelated given T_i , and have the following conditional expectations $\mu(t) = E(Y|T = t)$ and $\pi(\tau - t) = E(\mathbf{1}(\tau)|T = t)$.

Let the corresponding compound processes $\{L^R(t; \tau), t \leq \tau\}$ as the thinned process representing the liability of reported claims, i.e., $L^R(t; \tau) = \sum_{i=1}^{N(t; \tau)} Y_i \mathbf{1}_i(\tau)$; and $\{L^{IBNR}(t; \tau), t \leq \tau\}$ as the thinned process representing the liability of incurred but not reported (IBNR) claims, i.e., $L^{IBNR}(t; \tau) = \sum_{i=1}^{N(t; \tau)} Y_i (1 - \mathbf{1}_i(\tau))$.

Then the following relationship holds between the expected values:

$$E(L^{IBNR}(t; \tau)) = \int_0^t \frac{1 - \pi(\tau - s)}{\pi(\tau - s)} dE(L^R(s; \tau)) \quad (2)$$

where this is a Riemann–Stieltjes integral with respect to $E(L^R(t; \tau))$, as a function of t .

Proof. A direct application of iterated expectation and Campbell’s theorem leads to:

$$E(L^R(t; \tau)) = E\left(\sum_{i=1}^{N^R(t; \tau)} Y_i \mathbf{1}_i(\tau)\right) = E\left(\sum_{i=1}^{N^R(t; \tau)} \mu(T_i) \pi(\tau - T_i)\right) = \int_0^t \mu(s) \pi(\tau - s) d\Lambda(s)$$

and similarly, $E(L^{IBNR}(t; \tau)) = \int_0^t \mu(s) (1 - \pi(\tau - s)) d\Lambda(s)$. Equivalently, in differential form we obtain:

$$dE(L^R(t; \tau)) = \mu(t) \pi(\tau - t) d\Lambda(t) \quad \text{and} \quad dE(L^{IBNR}(t; \tau)) = \mu(t) (1 - \pi(\tau - t)) d\Lambda(t),$$

and after a simple rearrangement, we obtain

$$dE(L^{IBNR}(t; \tau)) = \frac{1 - \pi(\tau - t)}{\pi(\tau - t)} dE(L^R(t; \tau)),$$

which leads to the integral form shown above. \square

Proposition 1 provides an interpretation of the odds ratio of probabilities as a change-of-measure mechanism, similar to a Radon–Nikodym derivative. Indeed, the IPW principle acts as a weighting function within an integral that transforms expectations of aggregate reported claims into expectations of aggregate not reported claims. This idea can be used to show results associated to the IPW principle, as we will do in the next sections. Moreover, this proposition can be used to motivate the IPW estimator in Equation (1) as an empirical version of Equation 2 where we use the observed path of $L^R(t; \tau)$ as integrator. This assertion can be evidenced as an interim step in Corollary 1.

Corollary 1. The IPW estimator for the IBNR reserve is unbiased.

Proof. Interchanging the integral and expectation on the right-hand side of Equation(2) we obtain

$$E(L^{IBNR}(\tau)) = E(L^{IBNR}(\tau; \tau)) = E\left(\int_0^\tau \frac{1 - \pi(\tau - s)}{\pi(\tau - s)} dL^R(s; \tau)\right) = E\left(\sum_{i=1}^{N^R(\tau)} \frac{1 - \pi_i(\tau)}{\pi_i(\tau)} Y_i\right).$$

The integral simplifies to a sum as $L^R(t; \tau)$ is a step function. Note this proof differs from that from the sampling literature. The latter is more general as does not require a probability model. \square

One of the main attributes of the IPW estimator is that it achieves an estimation in a frequency-severity distribution-free manner, similar to the CL method. In fact, Calcetero-Vanegas et al. (2024) demonstrated that the CL method can be viewed as an IPW estimator where inclusion probabilities are approximated using empirical proportions. By employing a predictive model to estimate inclusion probabilities with granular information, the basic CL method can be extended to incorporate more granular data. We will revisit these results in the next section.

While such a result is remarkable, the discussion on such paper did not fully explore the broader context of the population sampling literature. Specifically, the IPW approach was examined under the *design-based* framework, which relies solely on the randomness of the sampling mechanism (i.e., whether a claim is or not reported) for inference, without accounting for the probabilistic model underlying claim generation process. In contrast, the population sampling literature encompasses more general frameworks, such as the *model-based* approach. This approach incorporates a probabilistic model for claim generation alongside the randomness of the sampling design, thereby leveraging additional information for more accurate modeling and predictions. In this setup, the goal is to combine this model with the IPW to estimate population totals more effectively.

The broader perspective of the *model-based* approach enriches our understanding of reserving and enables more robust methods for prediction. As we will demonstrate in the next sections, sampling-based methods under the *model-based* approach provide a statistically sound foundation for reserving practices, offering valuable insights for improved predictions.

3 Bridging the gap between aggregate and individual reserving using AIPW estimators

This section introduces new augmented inverse probability weighting (AIPW) estimators for the reserve, under the model based approach. Our goals with such estimators are 1) to extend and improve the usual IPW estimator for the reserve with a more robust alternative and 2) to provide a unified statistical framework for macro and micro claim reserving methods. Historically, many aggregate reserving methods were devised as ad-hoc procedures or derived artificially in the stochastic claim reserving literature under non-intuitive assumptions about the claim generation mechanism. In contrast, our framework shows a construction of these methods at an individual level, which provides a more natural interpretation, and an easy pathway to integrate with individual methods. This insight positions the theory of AIPW estimators and population sampling as a cohesive foundation for claim reserving, offering a fresh perspective on these methodologies.

We begin by introducing AIPW estimators as a natural extension of IPW estimators on which a model for the claim data generation is at a play, and then discuss how aggregate and individual reserving approaches are special cases of AIPW estimators, under specific conditions, and therefore bridging the gap for these methodologies.

3.1 Augmented inverse probability weighting estimators for the reserve

Consider a reserving model already selected and fitted to the claim data, which is based on some distributional assumptions, e.g. an individual reserving model or a loss-ratio-based model, etc. This model serves as the starting point for further predictions. Typically, such a model—built using frequency, severity, and reporting delay time models—provides estimates of the total number of claims $\hat{N}(\tau)$, the total number of non-reported claims $\hat{N}^{IBNR}(\tau)$, and the amounts for each claim,

denoted by \hat{Y}_i . The total reserve for unobserved claims is then estimated as $\sum_{i=1}^{\hat{N}^{IBNR}(\tau)} \hat{Y}_i$. Here, the term "model" is used broadly, as these estimates may derive from qualitative or quantitative sources, not necessarily from a statistical model. For instance, the estimates could be informed by expert opinion, such as for case estimates, or by predictive models, as in Antonio and Plat (2014).

As inspired in the literature of population sampling, e.g., Särndal et al. (2003), we propose to extend the IPW estimator to a *model assisted* version via an *augmented inverse probability weighting* (AIPW) estimator (also known as the difference estimator) given by Equation (3).

$$\hat{L}(\tau) = \sum_{i=1}^{\hat{N}(\tau)} \hat{Y}_i + \sum_{i=1}^{N^R(\tau)} \frac{Y_i - \hat{Y}_i}{\hat{\pi}_i(\tau)}, \quad (3)$$

where the first term represents the estimation of the population total based solely on the reserving model through the estimates \hat{Y}_i , and the second term is a scaled version of the prediction errors of the model ($Y_i - \hat{Y}_i$) in the sample. The motivation behind considering this type of estimator is based on applying the IPW principle, but on the errors of the model as in the following idea:

$$L(\tau) = \sum_{i=1}^{N(\tau)} Y_i = \sum_{i=1}^{N(\tau)} \hat{Y}_i + \sum_{i=1}^{N(\tau)} (Y_i - \hat{Y}_i) \approx \sum_{i=1}^{\hat{N}(\tau)} \hat{Y}_i + \sum_{i=1}^{N^R(\tau)} \frac{Y_i - \hat{Y}_i}{\hat{\pi}_i(\tau)}.$$

The intuition here is that the errors $Y_i - \hat{Y}_i$ potentially have less variability than the outcome Y_i , and therefore the associated AIPW estimator would have a lower variance and would lead to better predictions, while also being an unbiased estimator as we show below.

Along those lines, we can derive an unbiased AIPW estimator of the outstanding claims as

$$\hat{L}^{IBNR}(\tau) = \hat{L}(\tau) - L^R(\tau) = \sum_{i=1}^{\hat{N}(\tau)} \hat{Y}_i + \sum_{i=1}^{N^R(\tau)} \frac{Y_i - \hat{Y}_i}{\hat{\pi}_i(\tau)} - \sum_{i=1}^{N^R(\tau)} Y_i \quad (4)$$

$$= \sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} Y_i + \left(\sum_{i=1}^{\hat{N}^{IBNR}(\tau)} \hat{Y}_i - \sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} \hat{Y}_i \right) \quad (5)$$

$$= \sum_{i=1}^{\hat{N}^{IBNR}(\tau)} \hat{Y}_i + \sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} (Y_i - \hat{Y}_i), \quad (6)$$

in which we use the decomposition $\sum_{i=1}^{\hat{N}(\tau)} \hat{Y}_i = \sum_{i=1}^{N^R(\tau)} \hat{Y}_i + \sum_{i=1}^{\hat{N}^{IBNR}(\tau)} \hat{Y}_i$ (abusing the notation of the indices) in the last equality. Note that the regular IPW estimator is recovered by ignoring the terms \hat{Y}_i from the model assisting.

The AIPW estimator acts as a correction mechanism. The first term in Equations (5) and (6) represent the IPW estimator and the micro level estimator for the IBNR reserve respectively, while the associated second term addresses errors in the estimators. Specifically, in Equation (6), the second term corrects the micro-level prediction by adding the mismatch between such predictions and the observed reported claims, i.e., $Y_i - \hat{Y}_i$, after being stretched by the inclusion probabilities odds ratio. In Equation (5), the second term corrects the usual IPW estimator by comparing the predicted and weighted-scaled estimates for not reported claims, focusing on discrepancies caused by the IPW scaling. Thus the reserving model and the IPW estimator correct each other, providing robustness against misspecification.

This interplay between the reserving model and the IPW estimator ensures a desirable property, widely studied in statistics, known as double robustness: if either a) the estimated inclusion probabilities are correctly specified (i.e., the estimated probabilities match the true inclusion probabilities) or b) the prediction of the reserving model of the population total is unbiased, the AIPW estimator provides an unbiased reserve estimate (see, e.g., Seaman and Vansteelandt (2018)). As most of the literature presents this property in the context of causal inference and setups where the population size is deterministic, we establish and demonstrate a version within our framework for completeness in Proposition 2 below.

Proposition 2 (Double robustness of the AIPW). *Under the notation introduced so far, assume also that a) the frequency model is conditionally unbiased given the vector of attributes, i.e., $E(\hat{N}(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}) = E(N(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)})$, and b) that frequency and severity are conditionally independent, given the vector of attributes.*

Therefore the AIPW estimators in Equations (3) and (6) are unbiased, i.e., $E(\hat{L}(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}) = E(L(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)})$ and $E(\hat{L}^{IBNR}(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}) = E(L^{IBNR}(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)})$, provided at least one of the following holds:

1. *The inclusion probabilities are correctly specified, that is, $\hat{\pi}_i(\tau) = \pi_i(\tau) = E(\mathbf{1}_i(\tau) | Y_i, T_i, \mathbf{x}_i)$, where recall $\mathbf{1}_i(\tau)$ is the membership indicator random variable.*
2. *The severity model is conditionally unbiased, that is, $E(\hat{Y}_i | \mathbf{1}_i(\tau), T_i, \mathbf{x}_i) = E(Y_i | \mathbf{1}_i(\tau), T_i, \mathbf{x}_i)$.*

Proof. We prove the assertion only for $\hat{L}(\tau)$ as it would follow immediately for $\hat{L}^{IBNR}(\tau) = \hat{L}(\tau) - L^R(\tau)$ by linearity of the expectation. To do so we consider the two cases separately:

1. Suppose the inclusion possibilities are correctly specified, as $\hat{\pi}_i(\tau) = \pi_i(\tau) = E(\mathbf{1}_i(\tau) | Y_i, T_i, \mathbf{x}_i)$. Then we have,

$$\begin{aligned}
E(\hat{L}(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}) &= E\left(\sum_{i=1}^{\hat{N}(\tau)} \hat{Y}_i + \sum_{i=1}^{N^R(\tau)} \frac{Y_i - \hat{Y}_i}{\hat{\pi}_i(\tau)} | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) \\
&= E\left(\sum_{i=1}^{N(\tau)} \hat{Y}_i | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) + E\left(\sum_{i=1}^{N(\tau)} \frac{Y_i - \hat{Y}_i}{\pi_i(\tau)} \mathbf{1}_i(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) \\
&= E\left(\sum_{i=1}^{N(\tau)} \hat{Y}_i | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) + E\left(\sum_{i=1}^{N(\tau)} \frac{Y_i - \hat{Y}_i}{\pi_i(\tau)} E(\mathbf{1}_i(\tau) | Y_i, T_i, \mathbf{x}_i) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) \\
&= E\left(\sum_{i=1}^{N(\tau)} \hat{Y}_i | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) + E\left(\sum_{i=1}^{N(\tau)} Y_i - \hat{Y}_i | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) \\
&= E\left(\sum_{i=1}^{N(\tau)} Y_i | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) = E(L(\tau) | \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}),
\end{aligned}$$

where we use the assumption that the frequency model is unbiased and the estimated inclusion probability are correctly specified to remove the associated hats in the second equality, apply the iterated law of expectation (conditioning on claim sizes and other attributes) in the third, and cancel out common terms in the fourth. Notably, no assumptions were made about \hat{Y}_i .

2. Suppose that the claim severity model is unbiased, so that $E(\hat{Y}_i \mid \mathbf{1}_i(\tau), T_i, \mathbf{x}_i) = E(Y_i \mid \mathbf{1}_i(\tau), T_i, \mathbf{x}_i)$. Then we have,

$$\begin{aligned}
E\left(\hat{L}(\tau) \mid \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) &= E\left(\sum_{i=1}^{\hat{N}(\tau)} \hat{Y}_i + \sum_{i=1}^{N^R(\tau)} \frac{Y_i - \hat{Y}_i}{\hat{\pi}_i(\tau)} \mid \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) \\
&= E\left(\sum_{i=1}^{N(\tau)} Y_i \mid \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) + E\left(\sum_{i=1}^{N(\tau)} \frac{Y_i - \hat{Y}_i}{\hat{\pi}_i(\tau)} \mathbf{1}_i(\tau) \mid \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) \\
&= E\left(\sum_{i=1}^{N(\tau)} Y_i \mid \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) + E\left(\sum_{i=1}^{N(\tau)} \frac{E(Y_i - \hat{Y}_i \mid \mathbf{1}_i(\tau), T_i, \mathbf{x}_i)}{\hat{\pi}_i(\tau)} \mathbf{1}_i(\tau) \mid \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) \\
&= E\left(\sum_{i=1}^{N(\tau)} Y_i \mid \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right) + 0 = E\left(L(\tau) \mid \{\mathbf{x}_i\}_{i=1}^{N^R(\tau)}\right),
\end{aligned}$$

where we assume the frequency is unbiased to remove the associated hats in the second equality, apply the iterated law of expectation (conditioning on the indicators and other attributes) in the third, and cancel the second term in the fourth, based on the assumption of the severity model being unbiased. Notably, no assumptions were made about $\hat{\pi}_i(\tau)$.

Thus, the AIPW is a doubly robust estimator of the reserve, requiring only one of the models for the claim severity model or the inclusion probabilities to be correctly specified to have unbiasedness. \square

The double-robustness property makes inference based on AIPW estimators more reliable than using a standard IPW estimator or a reserving model alone. The predictive performance of the AIPW estimator falls between that of the reserving model and the IPW estimator. A key takeaway from the sampling approach is that reserving models can benefit from the error correction provided by AIPW estimators, yielding more accurate estimates, as illustrated in the numerical study. However, this is more of a theoretical advantage than a strictly practical one. In reality, all models are misspecified to some extent, and inaccuracies in the reserving model or inclusion probabilities—often due to oversimplified assumptions—may limit the full realization of the double-robustness property. Nevertheless, AIPW estimators typically provide competitive estimates even under these conditions (e.g., Waernbaum and Pazzagli (2023), Kang and Schafer (2007)).

In insurance reserving, many methods implicitly incorporate principles related to the AIPW framework to improve accuracy. These methods are often conceived for practical purposes rather than being purely statistically motivated. In what follows, we demonstrate how the AIPW framework provides a statistical foundation for understanding various reserving methods.

3.2 Aggregate reserving models as AIPW estimators

Aggregate or macro-level reserving methods, such as the CL method and its extensions, rely on aggregate data. A key contribution of Calcetero-Vanegas et al. (2024) was showing that the CL method is a particular case of the IPW estimator under certain homogeneity assumptions regarding the portfolio and inclusion probabilities. Specifically, for claims from a fixed accident year, say the k -th year, the CL estimator for ultimate claims can be expressed as:

$$L^{CL}(\tau) = \frac{L^R(\tau)}{\pi^{CL}(\tau)},$$

where $\pi^{CL}(\tau)$ represents the implied inclusion probability, derived as the inverse of the CL development-to-ultimate factor, i.e., $\pi^{CL}(\tau) = 1/f_k(\tau)$. Note that this idea applies similarly to the other accident years, and so it generalizes immediately. Along those lines, this formulation shows that the CL method is a particular case of the IPW estimator and, therefore, also an AIPW estimator, albeit without incorporating an assisting model.

Within the AIPW framework, this connection extends to other aggregate reserving methods. For example, as demonstrated in Proposition 3, methods in the class of linear credibility models can be shown to be AIPW estimators. This class includes the Bühlmann-Straub model, the Benktander-Hovinen method, the Cape-Cod method and other Bayesian approaches (see Wüthrich and Merz (2008)). This interpretation grants these reserving methods some of the theoretical properties of the AIPW framework, and extends the view of the population sampling beyond the CL method.

Proposition 3. *Consider an estimation $\hat{L}_{Cred}(\tau)$ that combines expert opinion with the Chain Ladder method to estimate the ultimate claims using a linear credibility approach, i.e., for a fixed accident year the ultimate claims are estimated as,*

$$\hat{L}_{Cred}(\tau) = Z\hat{L}_{CL}(\tau) + (1 - Z)\hat{L}_E(\tau),$$

where $\hat{L}_{CL}(\tau)$ denotes the Chain Ladder estimate of the ultimate claims, $\hat{L}_E(\tau)$ the expert-based estimate and Z is a credibility factor satisfying $0 \leq Z \leq 1$.

Then $\hat{L}_{Cred}(\tau)$ is an AIPW estimator, where the model assisting is the expert opinion, and the augmentation term is built based on the CL and its implied inclusion probabilities.

Proof. Define $\sum_{i=1}^{\hat{N}^{IBNR}(\tau)} \hat{Y}_i := \hat{L}_E(\tau)$ as the reserving model estimate, and $\hat{L}_E^R(\tau) = \hat{L}_E(\tau) \times \pi^{CL}(\tau)$ as the expected value for the current paid amount by τ , as based on the expert opinion. Defining $\pi^{CL}(\tau) = \pi^{CL}(\tau)/Z$, the credibility-based estimator becomes:

$$\hat{L}_{Cred}(\tau) = \hat{L}_E(\tau) + \frac{\hat{L}_E^R(\tau) - \hat{L}_E(\tau)}{\pi^{CL}(\tau)}.$$

This estimator is in structure an AIPW estimator where the model assisting is the expert opinion, and the augmentation term is built based on credibility adjusted CL inclusion probabilities. Lastly, note that different accident years are treated analogously, so the total reserve for all accident years follows immediately by adding up the results. \square

3.3 Individual reserving models as AIPW estimators

We now delve into individual or micro-level models in insurance, which work with granular data. Micro-level reserving models aim to characterize each component involved in reserve estimation through predictive modeling at an individual level. This includes modeling the number of claims per individual (e.g. using a Poisson process), the corresponding claim amounts (e.g. using a GLM), and their associated reporting delay times (e.g. using survival model). Models are typically selected from flexible classes from the machine learning literature, chosen in a data-driven manner to accurately capture the behavior observed in the data. Consequently, an actuary can derive a distributional model for claims, allowing the estimation of a point estimate \hat{Y}_i for claim amounts, as well as estimates for associated delays and the total number of claims, and non-reported claims $\hat{N}^{IBNR}(\tau)$. We refer to Antonio and Plat (2014) for more details.

To show when a micro-level model can be regarded as a specific case of an AIPW estimator, we first recall the concept of the *balance property* of the severity model, as discussed in Wüthrich (2020)

and Campo and Antonio (2023). In summary, a severity model satisfies the balance property when the average predicted claim sizes align with the average observed values. In terms of totals, this can be expressed as:

$$\sum_{i=1}^{N^R(\tau)} w_i \hat{Y}_i = \sum_{i=1}^{N^R(\tau)} w_i Y_i,$$

where the w_i serve as exposure-related factors, which can take values of $w_i = 1$ for single exposure, or some other values depending on the context. The balance property guarantees that aggregate estimations from a predictive severity model match the losses at the portfolio level, which is the liability of interest. The importance of this property has been highlighted in both ratemaking and reserving contexts extensively. With this in mind we now show how micro-level models can be seen as AIPW estimators.

Proposition 4. *Suppose a micro-level model satisfies a weighted balance property with weights $w_i = \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)}$, then the micro-level prediction is an AIPW estimator.*

Proof. It suffices to verify when the predictions of the micro-level model align with those of the corresponding AIPW estimator. This can be observed in Equation (6), where the difference lies in the second term, the augmentation term. By assumption the model satisfies the weighted balance property, that is:

$$\sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} \hat{Y}_i = \sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} Y_i \Leftrightarrow \sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} (Y_i - \hat{Y}_i) = 0,$$

Hence, the augmentation term is negligible and the AIPW estimation becomes equivalent to the micro-level model. \square

Several actuarial models satisfy the balance property naturally, such as any Generalized Linear Model (GLM) with a canonical link (ee, e.g., Wüthrich and Merz (2023) for further details). In cases where the balance property is not inherently satisfied, actuaries often adjust the severity model post-estimation to enforce this alignment (e.g., Crevecoeur et al. (2022), Campo and Antonio (2023)). Consequently, most micro-level reserving models in practice happen to satisfy the balance property, ensuring their predictions align with those derived from the AIPW estimators, as long as the proper weight is used. The interpretation and significance of this particular weighting scheme will be elaborated upon in the next section. This connection offers a compelling and natural rationale for why achieving the balance property in a model is desirable: it aligns the reserving model with the general properties of the AIPW estimator. Nevertheless, there is no single method to impose the balance property, leading to variations in model predictions and the resulting properties.

3.4 Bridging the gap and hybrid approaches

Since these aggregate and individual reserving models are AIPW estimators, the AIPW framework provides a statistically sound approach to reserve estimation, unifying individual and aggregate models while drawing from population sampling theory. This perspective offers a novel interpretation of reserving methods, complementing existing results in stochastic claim reserving. Although beyond the scope of this report, several noteworthy implications follow. For instance, inference techniques developed for AIPW estimators can now be applied to reserving methods, particularly to aggregate models, where statistical inference has lacked rigor. Additionally, key properties such as unbiasedness, optimality, and other nonparametric guarantees of reserving methods could be directly derived from the established theory of AIPW estimators.

Moreover, alternative AIPW structures can balance the simplicity of the CL method with the complexity of an individual reserving model within a single estimation, leading to hybrid approaches. Example 1 illustrates this idea, and we assess its performance in our numerical study. Other hybrid methods can also be explored depending on analytical constraints. For instance, a standard IPW estimator with a predictive model solely for the inclusion probabilities, incorporating individual information, could serve as a hybrid approach, generalizing the CL method by integrating individual-level data.

Example 1. *A potential structure of interest integrates directly the CL method and a microlevel model into a single estimate. To see how this can be achieved, note that in Equation (5), the IPW component of the AIPW estimator reduces to the Chain-Ladder prediction if we set the inclusion probabilities as $\hat{\pi}_i^{CL}(\tau) = 1/f_i(\tau)$, where $f_i(\tau)$ is the development factor to ultimate from the CL method. The second term then acts as an error correction term based on the micro-level model.*

Under this approach, the resulting AIPW estimator takes the form

$$\sum_{i=1}^{\hat{N}^{IBNR}(\tau)} \hat{Y}_i + \sum_{i=1}^{N^R(\tau)} \frac{1 - \pi_i^{CL}(\tau)}{\pi_i^{CL}(\tau)} (Y_i - \hat{Y}_i),$$

providing a direct trade-off between the micro-level model and the CL estimator. Similarly to the double robustness property of AIPW estimators, this estimator shifts weight toward either the CL estimate or the micro-level estimate, depending on their relative accuracy as reflected in the error correction terms. This balance makes it a compelling structure for reserving, serving as a middle ground between the two approaches.

Note that this estimator may not necessarily outperform an AIPW estimator with model-based inclusion probabilities. However, as model misspecification is common in practice, especially with the challenge of modeling reporting delays, this alternative AIPW estimator may still offer a balanced approach with competitive predictive performance.

Lastly, while AIPW estimators provide a robust approach to reserving by integrating macro- and micro-level methods, they represent just one tool within the broader population sampling framework. Other methodologies can further address the challenges in reserving. In the next section, we explore these complementary techniques and their applications.

4 Further applications of population sampling in reserving

The previous sections demonstrated the versatility of AIPW estimators in reserve estimation. However, more methodologies from the sampling literature can also offer valuable tools for reserving. This section highlights key ideas from population sampling that can improve the claim reserving predictions, providing more flexible and robust estimates.

Before introducing the methods, we first discuss the sampling bias issues associated with the reserving problem more thoroughly so that we can view the situation in an even more general setting. With this notion introduced, we then discuss the overall methods from the sampling literature designed to address such issues, for which the AIPW is a particular case.

4.1 The sampling issues in the reserving problem

In general sampling scenarios, the representativeness of a sample depends on the sampling design. When the sample is not representative of the overall population, the behavior of the sampled data

may significantly differ from that of the population, leading to biased inferences, analysis, model fitting, among others if this is not addressed. This issue is known as sampling bias in the population sampling literature, and similar fields.

In the specific context of reserving, the reporting mechanism divides claim data into two groups: reported and not reported claims. The observed data of reported claims (T_i, Y_i, U_i) come from a truncated distribution, as only claims reported by the valuation time are available, while the rest are treated as missing. Mathematically, the observed data from reported claims are realizations of the conditional distribution:

$$f_{T,Y,U}(t, y, u \mid U \leq \tau - T) = \frac{f_{T,Y,U}(t, y, u) \mathbf{1}_{\{u \leq \tau - t\}}}{P(U \leq \tau - T)}.$$

As a result, the behavior of reported claims may differ from that of not reported claims, i.e., $f_{T,Y,U}(t, y, u \mid U \leq \tau - T) \neq f_{T,Y,U}(t, y, u \mid U > \tau - T)$, particularly if significant dependencies exist among these variables, leading to sampling bias on the reported claims data. Empirical evidence supports this distinction in auto insurance, where reported claims are often more severe than not reported ones.

Along those lines, when estimating reserves and fitting reserving models for components such as claim severity, accounting for conditional distributions is essential to address sampling bias. Unfortunately, this is often overlooked by assuming independence of the claim severity and the reporting delay, e.g., $f_Y(Y_i \mid U_i \leq \tau - T_i) = f_Y(Y_i \mid U_i > \tau - T_i) = f_Y(Y_i)$. This simplification facilitates model fitting but ignores potential discrepancies due to the conditional nature of the data. Imposing a dependence structure between incurred losses and reporting delays would require methods like copulas, which are challenging due to the right-truncation inherent in reporting delays. However, the assumption of independence between claim severity and reporting may oversimplify the problem and lead to poor performance of predictions. If significant dependencies exist, reported claims may differ significantly in severity from not-reported ones, and neglecting these differences can undermine reserve accuracy. A numerical study in Section 6, using real data, illustrates this behavior.

To improve predictions, it is crucial to address discrepancies between reported and not reported claims cost-effectively when estimating reserving model components. This can be achieved by adopting a population sampling perspective and leveraging established techniques from that field. These methods are designed to mitigate sampling bias and enhance predictive accuracy but remain underutilized in actuarial science and reserving, in particular. The following subsection demonstrates how these approaches can be integrated into reserving models.

4.2 Population sampling methods to handle sampling bias

Here, we present an overview of population sampling techniques used to address sampling bias in model predictions. For simplicity, we focus on claim severity, though these methods apply equally to all components of the problem. In this context, we observe data from $f(Y \mid U \leq \tau - T)$ i.e. reported claims, but our goal is to estimate quantities related to not reported claims $f(Y \mid U > \tau - T)$ (e.g. the mean or the whole distribution itself).

The population sampling techniques for addressing biases on a severity model can be categorized based on when they are applied during the estimation process of such a model. Using terminology borrowed from fairness and discrimination bias mitigation (e.g., Charpentier (2024)), we informally classify these methods in order of their prevalence in the literature:

- Post-Processing: Adjustments for the sampling design are applied to predictions after fitting the severity model.
- In-Processing: Adjustments are incorporated into the model fitting process to account for the sampling design.
- Pre-Processing: Modifications are made directly to the data before fitting the model to address the sampling design.

All techniques begin with estimating inclusion probabilities, which define the sampling mechanism in the data and serve as a fundamental input for population sampling methods. In reserving, this involves estimating the probability of a claim being reported or not reported, as detailed in Section 2.1. These probabilities are typically derived as part of the reserving model using survival models, an approach we demonstrate in our numerical case study. We will now discuss several methods under these approaches in detail.

4.3 Post-processing approach: AIPW and weighted balance property

In this approach, the data is treated as if there was no sampling bias, and the severity model is fitted using the data from the reported claims under the independence assumption. The correction associated with the sampling bias is then applied after to the predictions made by this model.

The AIPW, as introduced in the previous section, falls into this category. The augmentation term serves as a post-processing correction to the overall prediction of the reserve to the given individual reserving model. Specifically, this term can be interpreted as measure of the bias of the prediction from the reserving model, as

$$E \left(\sum_{i=1}^{N^R(\tau)} \frac{Y_i - \hat{Y}_i}{\hat{\pi}_i(\tau)} \right) = E \left(\sum_{i=1}^{N(\tau)} \frac{Y_i - \hat{Y}_i}{\pi_i(\tau)} \mathbf{1}_{i(\tau)} \right) = E \left(\sum_{i=1}^{N(\tau)} Y_i - \hat{Y}_i \right) = -E \left(\sum_{i=1}^{N(\tau)} \hat{Y}_i - \sum_{i=1}^{N(\tau)} Y_i \right) = -\text{Bias} \left(\sum_{i=1}^{N(\tau)} \hat{Y}_i \right)$$

and once it is added to construct the AIPW estimator, it subtracts the bias from the prediction of the model.

Other approaches, which may not necessarily be superior, can also be used for this purpose. For example, imposing the balance property on the severity model can be viewed as a post-processing technique aimed at handling the bias in aggregate predictions. However, it is important to note that the way actuaries typically implement the balance property may not adequately address the sampling design, as we explain next.

Given the sampling bias, imposing the balance property solely on the reported data does not guarantee that aggregate quantities from the severity model are unbiased for the entire population of claims. It only ensures unbiasedness within the population of reported claims. To ensure the balance property holds at the overall population level, it is essential to account for the sampling mechanism that determines both reported and not reported claims, rather than focusing exclusively on reported claims.

In this context, we can extend the definition of the balance property of the severity model to different population levels: reported claims, not reported claims, or the entire portfolio, depending on which group is the target. Excluding exposure weights, these balance properties can be formulated as follows:

$$\sum_{i=1}^{N^R(\tau)} \hat{Y}_i = \sum_{i=1}^{N^R(\tau)} Y_i, \quad \sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} \hat{Y}_i = \sum_{i=1}^{N^R(\tau)} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} Y_i, \quad \sum_{i=1}^{N^R(\tau)} \frac{\hat{Y}_i}{\hat{\pi}_i(\tau)} = \sum_{i=1}^{N^R(\tau)} \frac{Y_i}{\hat{\pi}_i(\tau)}$$

The first type of balance property aims to remove the sampling bias in the aggregated claims for the population of reported claims, the second for not reported claims, and the third for the entire population of claims. Each of these properties represents the standard balance property but applies a change of measure using odds ratio weights to account for the specific target populations. In reserving, where the primary focus is on predicting the population of not reported claims, the second balance property is most appropriate, as it directly targets the quantity of interest, which is the aggregate liability of not-reported claims. This approach is expected to lead to more accurate predictions for not-reported claims, and justifies the weighting scheme that shows up in the AIPW estimator in Equation (6).

A simple strategy to enforce this balance property involves adjusting the severity model using a calibration coefficient. Let \hat{Y}_i^{wBP} be the adjusted prediction of the severity model obtained as:

$$\hat{Y}_i^{wBP} = b\hat{Y}_i,$$

where b is determined to guarantee the weighted balance property. Simple algebra shows that this coefficient should be set to:

$$b = \frac{\sum_{i=1}^{N^R(\tau)} \frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} Y_i}{\sum_{i=1}^{N^R(\tau)} \frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} \hat{Y}_i}. \quad (7)$$

This adjustment aims to align the mean of the severity model with the mean of the distribution of not reported claims, thereby improving reserve prediction accuracy. However, the theoretical guarantees depend on the method used to impose the balance property. Alternatives, such as auto-calibration or constrained optimization, are available, but the most effective approach remains uncertain. Indeed, there is a gap on the literature on this regard as well as the general benefits of the balance property.

4.4 In-processing: Fitting models for the population of not reported claims

Another approach to address sampling bias in reported claims and its impact on severity model predictions involves modifying the algorithm for parameter estimation. Typically, parameter estimation for models is achieved by solving an estimating equation, such as $\sum_{i=1}^{N(\tau)} l(Y_i, \theta) = 0$, where the estimating equation could be the gradient of the log-likelihood. To account for sampling bias while estimating the model, we can modify the estimating equation to reflect the sampling design.

A common method in the sampling literature is to use weighted estimating equations, where the weights are the inverse of the inclusion probability $\hat{\pi}_i(t)$, as discussed in studies such as Wooldridge (2007). However, in reserving, the focus shifts from estimating the overall population distribution $f(Y)$ to predicting claims for not reported data, i.e., $f(Y \mid U > \tau - T)$. To achieve this, we propose minimizing a weighted loss function that accounts for the sampling mechanism, with slight modifications to target the population of not reported claims. Based on the IPW principle, the following weighted equation is unbiased for the loss function of the not reported claims population (see Corollary 1):

$$\sum_{i=1}^{N^R} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} l(Y_i, \theta)$$

Here, the parameters θ should be chosen to minimize this weighted loss function instead of the unweighted version, and so providing an estimate of $f(Y | U > \tau - T)$, which is more suitable for the reserving context. This approach is grounded in the literature on missing data and survey designs with non-response. For foundational theory, see generalized estimating equations (e.g., Section 5.4 of Shao (2008)). Recent applications of similar procedures in insurance can be found in Peiris et al. (2024), while Bücher and Rosenstock (2024) discusses related concepts in reserving.

This method provides predictions with clear interpretability, in contrast to post-processing techniques, where model predictions are adjusted in less transparent ways. Additionally, it adjusts all model parameters, not just the intercept, as in the balance property-based method discussed in the previous subsection. The degree to which the model parameters differ from those obtained using the regular unweighted loss function depends on the associations between non-reporting and the explanatory variables. As a result, the fitted model better represents the distribution of not reported claims, yielding more accurate predictions compared to models focused on reported claims or the entire population. We illustrate this advantage with real data in the next section.

Finally, while this report does not explore alternative fitting approaches in detail, other methods exist (e.g., Kim and Shao (2021)). For example, one could use standard estimating equations while imposing the balance property as a constraint. However, such methods tend to be computationally intensive, and their statistical properties are more complex to derive.

4.5 Pre-Processing: Synthetic data sets of not reported claims

A general approach to handling partially observed data involves estimating the population of not reported claims using data augmentation, generation, or imputation methods. By augmenting the data from reported claims, we can create a pseudo-population of not reported claims that follows the distribution $f(Y | U > \tau - T)$. This pseudo-population can then serve as the working data for subsequent analyses, including model fitting. This pre-processing method directly addresses the sampling bias in the data, rather than correcting it in later stages of estimation or prediction.

There are several methods to achieve this, and the following approach, commonly used for bootstrapping in the population sampling literature (e.g., Arnab (2017)), offers a straightforward solution. Briefly, each claim Y_i in the population appears, on average, $\hat{\pi}_i$ times in the sample. Conversely, each Y_i in the sample is expected to appear $1/\hat{\pi}_i$ times in the entire population of claims, or $\frac{1-\hat{\pi}_i}{\hat{\pi}_i}$ times in the population of not reported claims.

Along those lines, synthetic data for not reported claims can be constructed by replicating each reported claim (\mathbf{x}_i, Y_i) in a new dataset a total of $\frac{1-\hat{\pi}_i}{\hat{\pi}_i}$ times. The synthetic data should closely mimic the not reported claims, allowing for accurate reserve estimation and related quantities. The following proposition shows this more formally.

Proposition 5. *The distribution function of the severity of the resulting synthetic data of not reported claims is a strongly consistent estimation of the severity distribution of the population of not-reported claims.*

Proof. The distribution function for the severity of this pseudo-dataset is immediately given as:

$$\hat{F}_Y(y) = \frac{\sum_{i=1}^{N^R(\tau)} \frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} \mathbf{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^{N^R(\tau)} \frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)}},$$

which is a self-normalized IPW estimator (or equivalently a self-normalized importance sampling estimator) of the CDF of the severity of not reported claims. Here the odds ratio acts as the appropriate weighting for the change of measure from the conditional distribution of reported claims to that of not reported claims, as discussed after Proposition 1. This estimator is known to be a strongly consistent estimator (see e.g, Section 10.4 of Ross (2022)) for the conditional distribution, which is $f(Y | U > \tau - T)$. \square

The synthetic data can be used in various ways to aid the estimation of reserves. One approach is to fit the severity model using standard procedures without needing post- or pre-processing corrections, as the synthetic data already reflects the distribution of not reported claims. The resulting fitted model is closely related to one obtained via weighted estimating equations, as the loss function evaluated on the synthetic dataset mirrors the weighted loss function applied to the original data. Consequently, both approaches yield equivalent models. Alternatively, the synthetic data can be used to estimate the reserve directly by aggregating the claims and using the total as a point estimator. This approach aligns closely with the IPW estimator in Equation (1) and produces identical predictions, as both methods use the same weighting scheme. However, working with synthetic data is often more practical, as it allows for the direct use of standard software implementations without modifications. Moreover, they enable analysis of claim size distributions, exploration of attribute associations, and incorporation of uncertainty into calculations as we explain next.

While the process discussed earlier creates a single pseudo-population, multiple synthetic datasets can be generated to account for population uncertainty. For example, in the context of Poisson sampling in reserving, each reported claim can be viewed as a success in a Bernoulli trial, and therefore the number of attempts required to observe such a success follows a geometric distribution.

Along those lines, we propose to generate realization of random variables $Z_i \sim \text{Geom}(\hat{\pi}_i)$, $i = 1, \dots, N^R(\tau)$, to represent the number of not reported claims with the same characteristics as Y_i . The synthetic dataset is then constructed by including each Y_i a total of Z_i times. This approach allows for multiple simulations, where Z_i acts as a random weight, analogous to the fixed weight $\frac{1-\hat{\pi}_i}{\hat{\pi}_i}$. The pseudo-population created with fixed weights can be viewed as the "average" of these simulations. Although this geometric distribution-based method for synthetic data generation is uncommon, it has roots in missing data literature, such as the EM algorithm for truncated data (see Section 2.8 of McLachlan and Krishnan (2007)). The proposition below validates the use of this simulation approach from a Bayesian perspective.

Proposition 6. *Let $M_i(\tau)$ be the total number of claims with the same characteristics as Y_i . Let $M_i^R(\tau)$ and $M_i^{IBNR}(\tau)$ represent, respectively, the number of these claims that are reported before time τ and the number that remain not reported by τ , which arise according to the Poisson sampling scheme with inclusion probability $\pi_i(\tau)$. If $M_i(\tau)$ follows an improper distribution as $P(M_i(\tau) = m) \propto 1/m$, then*

$$M_i^{IBNR}(\tau) | M_i^R(\tau) \sim \text{NegBinom}(r = M_i^R(\tau), p = \pi_i(\tau)).$$

Proof. Because of the Poisson sampling scheme, we have $M_i^R(\tau) | M_i(\tau) \sim \text{Binom}(n = M_i(\tau), p =$

587 $\pi_i(\tau)$). Therefore,

$$\begin{aligned}
P(M_i^{IBNR}(\tau) = k \mid M_i^R(\tau) = r) &= P(M_i(\tau) = k + r \mid M_i^R(\tau) = r) \\
&\propto P(M_i^R(\tau) = r \mid M_i(\tau) = k + r) \times P(M_i(\tau) = k + r) \\
&\propto \binom{k+r}{r} (\pi_i(\tau))^r (1 - \pi_i(\tau))^{(k+r)-r} \times \frac{1}{k+r} \\
&\propto \binom{k+r-1}{k} (1 - \pi_i(\tau))^k (\pi_i(\tau))^r.
\end{aligned}$$

588 The proof finishes by noting that the last line is proportional to the pmf of the negative binomial
589 distribution mentioned earlier. \square

590 Note that while the proposition above refers to a negative binomial distribution, in practice, we
591 use the geometric distribution because each reported claim is treated individually, as they all have
592 entirely different characteristics. Moreover, the use of an improper prior distribution is motivated
593 by Bayesian inference, where is known that improper prior acts as noninformative priors and thus
594 minimizes its influence on the posterior, allowing the data to dominate and making the inference
595 as data-driven as possible. Thus, from the perspective of the resulting posterior distribution, the
596 improper prior is not a strong assumption. Additionally, other improper prior distributions can
597 be used, such as a uniform, which also leads to a negative binomial distribution but with slightly
598 different parameters. Nonetheless, the reciprocal form used in the proposition is widely used in
599 Bayesian statistics and is particularly intuitive in the insurance context, as larger claims become
600 increasingly less likely.

601 This method has several applications in reserving. It can quantify prediction uncertainty, similar to
602 bootstrapping but focused on not reported claims, and generate confidence intervals for predictions
603 and parameters without relying on asymptotic theory. For instance, since the average total claims
604 from these datasets match the IPW estimator, bootstrap confidence intervals for the IPW can
605 be constructed using quantiles from the synthetic totals generated in the simulations. Another
606 application is model testing, as synthetic data provides a model-free method to evaluate prediction
607 quality. Furthermore, for micro-level models that require known claim features (often unavailable),
608 this approach offers a straightforward way to generate claims and their attributes.

609 5 Simulated case study with implementation in R

610 In this section, we illustrate the methods using a simulated data set developed by Wang and
611 Wuthrich (2022). The dataset contains individual claims for non-life insurance reserving studies,
612 containing a range of claim-specific attributes and temporal markers. Each claim is characterized by
613 the accident date, reporting date, and settlement date, alongside payments made over time, claim
614 status, and various covariates such as the claimant's age and the claim type. The dataset includes
615 information on claims closed by a cutoff date, claims reported but not settled, and incurred but not
616 reported (IBNR) claims that need prediction.

617 Key attributes include identifiers such as claim ID, accident and reporting dates in various time
618 units (daily, monthly), reporting delay, and settlement delays. Claims have status indicators that
619 flag them as closed, reported but unsettled (RBNS), or IBNR as of the cutoff date, with cumulative
620 payments recorded for closed and RBNS claims. Additional features include the total claim amount
621 (Ultimate), payment count, age of the injured, and claim types with labels 1 to 6, which may

622 affect claim frequency and severity. Detailed payments data is also available, allowing for tracking
623 individual payment events, which are useful for back-testing claims reserving methods on this
624 dataset.

625 5.1 Generating the Data

626 One has to first download the relevant files in the same working directory, in particular “functions
627 simulation.R,” which can be accessed through the GitHub repository linked in the document:
628 github.com/actuarial-data-science/PackageIndividualClaimsSimulator. This repository includes the
629 R code and associated functions needed to generate and simulate individual claims data. One also
630 must ensure to install all required dependencies, like the SynthETIC R package, which is necessary
631 for running the simulation.

632 The primary function, `data.generation()`, generates the dataset. This produces a list with two
633 objects: a dataset of individual claim details, such as claim ID, type, accident date, reporting date,
634 and cumulative payments up to the cutoff date; and a dataset of claim payments and statuses,
635 identifying each payment event by event month and payment status (open or closed). Below is the
636 code to generate the data, and attributes.

```
##Load packages
library(SynthETIC) # 1.0.0 or higher needed
library(tidyverse)

#### generate claim data
ref_claim <- 1 # currency is 1
time_unit <- 1/24 # we consider biweekly claims development
set_parameters(ref_claim=ref_claim, time_unit=time_unit)
years <- 5 # number of occurrence years
I <- years/time_unit # number of development periods
source("../Tools/functions simulation.R")

# generate individual claims: may take 60 seconds
set.seed(123456789)
claims_list <- data.generation(seed=1000, future_info=F)

# save individual output for easier access
claims <- claims_list[[1]]

# Fix format and create dates variables
claims$Month = format(claims$AccDate, "%m")
claims$Year = as.factor(format(claims$AccDate, "%y"))
claims$Type = as.factor(claims$Type)
claims$RepDelDays = claims$RepDelDays + runif(n=length(claims$RepDelDays))

#Work only with the closed claims as we consider IBNR and not RBNS
claim_payments = claims[claims$Status=="Closed",]
str(claim_payments)
```

637 ## 'data.frame': 20412 obs. of 17 variables:

```

638 ## $ Id      : int  1 2 3 4 5 6 7 8 9 10 ...
639 ## $ Type     : Factor w/ 6 levels "1","2","3","4",...: 2 2 4 4 5 5 1 4 1 4 ...
640 ## $ Age      : int  45 31 27 58 59 43 30 57 30 19 ...
641 ## $ AccDate  : Date, format: "2012-01-05" "2012-01-08" ...
642 ## $ AccMonth : num  1 1 1 1 1 1 1 1 1 1 ...
643 ## $ AccWeekday : Factor w/ 7 levels "Mon","Tue","Wed",...: 4 7 5 5 2 3 5 4 4 6 ...
644 ## $ RepDate   : Date, format: "2012-01-05" "2012-01-08" ...
645 ## $ RepMonth  : num  1 1 1 1 1 1 1 1 1 1 ...
646 ## $ RepDelDays : num  0.45 0.189 1.221 2.083 12.657 ...
647 ## $ SetMonth   : num  62 28 48 37 25 79 22 38 21 38 ...
648 ## $ SetDelMonths: num  61 27 47 36 24 78 21 37 20 37 ...
649 ## $ Ultimate   : num  11946 27040 3829 1754 6443 ...
650 ## $ PayCount   : num  2 3 1 1 3 2 0 2 1 2 ...
651 ## $ Status     : Factor w/ 3 levels "Closed","RBNS",...: 1 1 1 1 1 1 1 1 1 1 ...
652 ## $ CumPaid    : num  11946 27040 3829 1754 6443 ...
653 ## $ Month      : chr  "01" "01" "01" "01" ...
654 ## $ Year       : Factor w/ 10 levels "12","13","14",...: 1 1 1 1 1 1 1 1 1 1 ...

```

655 The data spans the process from 2012-2017, and the reporting delay time is up to 3 years, which
 656 means that at least 3 years of historical data should be used for training. Therefore, we fix the
 657 date of valuation date to be the next business day after the end first month of year 2015. Below
 658 is the code to split the claims into reported and non reported by the valuation date, as well as
 659 the associated amount of the standing liability of nonreported claims, which is the main goal of
 660 estimation of this section.

```

#Valuation date
valuation_date = as.Date("2015-02-02")

#Split data into reported and non-reported (train and test)
#All claims in the period
all_claims = claim_payments %>% filter(AccDate<=valuation_date )

#Reported claims
reported_claims = claim_payments %>% filter(RepDate<=valuation_date,
                                             AccDate<=valuation_date )

#Not reported claims
non_reported_claims = claim_payments %>% filter(RepDate>valuation_date,
                                                  AccDate<=valuation_date )

#Standing claims (i.e. reserve)
sum(non_reported_claims$Ultimate)

```

```

661 ## [1] 13,248,103

```

662 We now proceed to illustrate the implementation of the sampling approach along with the reserving
 663 methods in order to estimate this amount. To do so we first require the estimation of the inclusion
 664 probabilities, which are obtained via the distribution of the reporting delay as described as follows.

5.2 Estimation of inclusion probabilities

We selected a Cox regression model that describes the hazard function of the reporting delay time, $\lambda_{U|x}(u)$, using the attributes of the policyholder \mathbf{x} as covariates, as follows:

$$\log(\lambda_U(u)) = \log(\lambda_0(u)) + \beta_1 \text{ClaimType} + \beta_2 \text{Age} + \beta_3 Y + \beta_4 \text{AccidentDay} + \beta_5 \text{AccidentMonth} + \beta_6 \text{AccidentYear}$$

where \log of the baseline hazard function, $\log(\lambda_0(u))$, is estimated using a B-Spline representation. The fitting is via a generalized additive models with a piecewise exponential that accounts for right-truncation, accessible in the R package `flexsurv` as follows:

```
#calculate truncation time variable (Valuation date-Accident date)
reported_claims$Truncation = as.numeric(valuation_date - reported_claims$AccDate+1)

#Fit the model
library("flexsurv")
fit = flexsurvspline(Surv(RepDelDays) ~ Type+Age+Ultimate+AccWeekday+Month+Year,
                     data = reported_claims,
                     rtrunc = Truncation ) #Estimation with right truncation.

#visualize results
fit
```

```
## Call:
## flexsurvspline(formula = Surv(RepDelDays) ~ Type + Age + Ultimate +
##      AccWeekday + Month + Year, data = reported_claims, rtrunc = Truncation)
##
## Estimates:
```

	data	mean	est	L95%	U95%	se	exp(est)
## gamma0	NA		-5.88e+00	-6.06e+00	-5.69e+00	9.42e-02	NA
## gamma1	NA		9.78e-01	9.59e-01	9.98e-01	9.98e-03	NA
## Type2	4.94e-02		6.96e-02	-7.79e-02	2.17e-01	7.53e-02	1.07e+00
## Type3	3.93e-02		2.75e-03	-1.32e-01	1.37e-01	6.87e-02	1.00e+00
## Type4	2.97e-01		-1.60e-02	-8.05e-02	4.86e-02	3.30e-02	9.84e-01
## Type5	1.33e-01		-1.94e-02	-1.04e-01	6.52e-02	4.32e-02	9.81e-01
## Type6	1.66e-01		-2.84e-02	-1.05e-01	4.81e-02	3.90e-02	9.72e-01
## Age	4.19e+01		8.15e-04	-1.03e-03	2.66e-03	9.40e-04	1.00e+00
## Ultimate	8.12e+03		4.00e-06	-3.81e-05	4.61e-05	2.15e-05	1.00e+00
## AccWeekdayTue	1.41e-01		1.02e-03	-9.34e-02	9.54e-02	4.82e-02	1.00e+00
## AccWeekdayWed	1.47e-01		-5.90e-02	-1.53e-01	3.47e-02	4.78e-02	9.43e-01
## AccWeekdayThu	1.44e-01		-5.89e-02	-1.53e-01	3.50e-02	4.79e-02	9.43e-01
## AccWeekdayFri	1.45e-01		-3.32e-03	-9.69e-02	9.02e-02	4.77e-02	9.97e-01
## AccWeekdaySat	1.36e-01		-5.61e-02	-1.51e-01	3.90e-02	4.85e-02	9.45e-01
## AccWeekdaySun	1.43e-01		-4.60e-02	-1.40e-01	4.81e-02	4.80e-02	9.55e-01
## Month02	9.52e-02		5.95e-02	-5.57e-02	1.75e-01	5.88e-02	1.06e+00
## Month03	9.76e-02		1.17e-02	-1.03e-01	1.26e-01	5.83e-02	1.01e+00
## Month04	9.24e-02		1.70e-01	5.43e-02	2.85e-01	5.88e-02	1.18e+00
## Month05	9.51e-02		1.70e-01	5.60e-02	2.85e-01	5.84e-02	1.19e+00
## Month06	8.41e-02		2.08e-01	9.01e-02	3.27e-01	6.04e-02	1.23e+00
## Month07	7.67e-02		2.97e-01	1.75e-01	4.18e-01	6.20e-02	1.35e+00

```

698 ## Month08          7.53e-02   3.60e-01   2.38e-01   4.83e-01   6.25e-02   1.43e+00
699 ## Month09          7.06e-02   4.52e-01   3.27e-01   5.76e-01   6.36e-02   1.57e+00
700 ## Month10          6.57e-02   4.62e-01   3.34e-01   5.89e-01   6.51e-02   1.59e+00
701 ## Month11          6.98e-02   4.12e-01   2.86e-01   5.38e-01   6.41e-02   1.51e+00
702 ## Month12          7.32e-02   4.77e-01   3.52e-01   6.02e-01   6.37e-02   1.61e+00
703 ## Year13           3.57e-01   3.70e-01   3.11e-01   4.28e-01   2.99e-02   1.45e+00
704 ## Year14           2.25e-01   1.18e+00   1.11e+00   1.25e+00   3.61e-02   3.26e+00
705 ## Year15           3.96e-03   3.50e+00   3.09e+00   3.91e+00   2.08e-01   3.30e+01
706 ##
707 ## N = 6816,  Events: 6816,  Censored: 0
708 ## Total time at risk: 1363016
709 ## Log-likelihood = -41887, df = 29
710 ## AIC = 83833

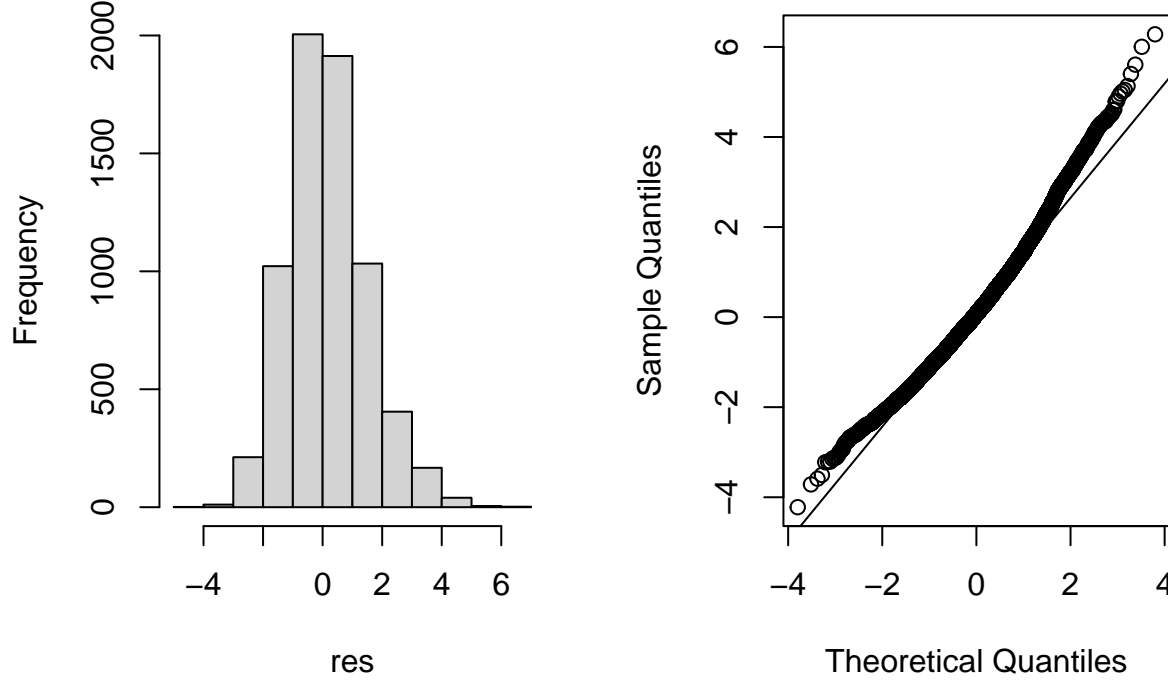
```

711 Note that reporting delay seem to change based on the type of claim as well as the accident date.
 712 We now validate the model and estimate the associated inclusion probabilities. The plots of pseudo
 713 residuals below illustrates that these are close to follow a normal approximation, therefore the model
 714 adjust the data properly.

```

#Predict inclusion prob and pseudo residuals
#We divide and predcit on groups, to make the calculation less expensive
K= 100 #
indexes = 1:length(reported_claims$RepDelDays)
split_indices <- split(indexes, cut(indexes, K, labels = FALSE))
#Vectors that store the values of the inclusion probabilities and the residuals
probs= numeric(0)
residuals = numeric(0)
for (i in 1:K ){
  #Predict probabilities
  a = predict(fit,type = "survival",
              times = reported_claims[split_indices[[i]],]$Truncation,
              newdata = reported_claims[split_indices[[i]],] )
  b = predict(fit,type = "survival",
              times = reported_claims[split_indices[[i]],]$RepDelDays,
              newdata = reported_claims[split_indices[[i]],] )
  #Concatenate
  probs = rbind(probs,
                cbind(split_indices[[i]],a$.pred[[1]]) )
  residuals = rbind(residuals,
                    cbind(split_indices[[i]],b$.pred[[1]]) )
}
#GOF Residual analysis
#Adjust residuals based on truncation probs and normal quantile
res = qnorm((1-residuals$.pred_survival)/(1-probs$.pred_survival))
#Plot them
par(mfrow = c(1, 2))
hist(res, main= "")
qqnorm(res, main= " " )
qqline(res)

```



715

716 5.3 Fitting a micro-level model

717 We illustrate the sampling methods along with a micro-level model. We first describe a simplistic
 718 micro-level model for this data, and fit it without using directly the population sampling based
 719 methods.

720 For the frequency, as we do not have the overall portfolio of policyholders but only the reported
 721 claims, we estimate a model using pseudo observations of the nonreported claims, which are
 722 constructed via the IPW estimate of the number of non-reported claims, $\frac{1-\pi_i(\tau)}{\pi_i(\tau)}$. This gives a model
 723 with the predictive rate for the number of non reported claims per number of reported claims of a
 724 given type. Statistically, this approach with pseudo observations is in spirit the same as the one
 725 from an EM algorithm, but we will not further elaborate on this.

726 For illustration, consider a Poisson regression model as follows. Note that this can be fitted using
 727 the glm function as follows.

$$N_i^{IBNR}(\tau) \sim \text{Poisson}(\lambda_i)$$

728

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{ClaimType} + \beta_2 \text{Age} + \beta_3 \text{AccidentDay} + \beta_4 \text{AccidentMonth} + \beta_5 \text{AccidentYear}$$

```
#Proxy for non-reported frequency
reported_claims$N_IBNR = probs$.pred_survival/(1-probs$.pred_survival)
```



```

#Fit frequency model
fit_freq = glm(formula = N_IBNR ~ Type+Age+AccWeekday+Month+Year,
               family = poisson,
               data = reported_claims)
#Predict frequency
predicted_freq=predict(fit_freq,
                      type = "response")
#Summarize the model
summary(fit_freq)

```

```

729 ##
730 ## Call:
731 ## glm(formula = N_IBNR ~ Type + Age + AccWeekday + Month + Year,
732 ##      family = poisson, data = reported_claims)
733 ##
734 ## Deviance Residuals:
735 ##      Min        1Q    Median        3Q        Max
736 ## -2.635   -0.234   -0.015    0.166    4.586
737 ##
738 ## Coefficients:
739 ##              Estimate Std. Error z value Pr(>|z|)
740 ## (Intercept)  -4.571363   0.207648  -22.01 < 2e-16 ***
741 ## Type2        -0.051003   0.144476   -0.35  0.72407
742 ## Type3         0.132385   0.135410    0.98  0.32824
743 ## Type4         0.093108   0.075657    1.23  0.21845
744 ## Type5         0.283368   0.092609    3.06  0.00221 **
745 ## Type6        -0.038229   0.088153   -0.43  0.66454
746 ## Age          -0.000125   0.002150   -0.06  0.95353
747 ## AccWeekdayTue -0.236417   0.106008   -2.23  0.02574 *
748 ## AccWeekdayWed -0.073760   0.102375   -0.72  0.47123
749 ## AccWeekdayThu -0.167640   0.105032   -1.60  0.11047
750 ## AccWeekdayFri -0.192253   0.103020   -1.87  0.06202 .
751 ## AccWeekdaySat -0.367805   0.107429   -3.42  0.00062 ***
752 ## AccWeekdaySun -0.112256   0.099432   -1.13  0.25891
753 ## Month02       1.354132   0.146961    9.21 < 2e-16 ***
754 ## Month03       1.019078   0.188644    5.40  6.6e-08 ***
755 ## Month04       1.008567   0.190729    5.29  1.2e-07 ***
756 ## Month05       1.081507   0.185369    5.83  5.4e-09 ***
757 ## Month06       1.171987   0.191149    6.13  8.7e-10 ***
758 ## Month07       1.482833   0.185408    8.00  1.3e-15 ***
759 ## Month08       1.680618   0.180069    9.33 < 2e-16 ***
760 ## Month09       1.921810   0.175628   10.94 < 2e-16 ***
761 ## Month10       2.065882   0.172003   12.01 < 2e-16 ***
762 ## Month11       2.332177   0.170011   13.72 < 2e-16 ***
763 ## Month12       2.547152   0.168291   15.14 < 2e-16 ***
764 ## Year13        0.727039   0.106679    6.82  9.4e-12 ***
765 ## Year14        2.363186   0.094851   24.91 < 2e-16 ***

```

```

766 ## Year15          6.670898   0.168271   39.64 < 2e-16 ***
767 ## ---
768 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
769 ##
770 ## (Dispersion parameter for poisson family taken to be 1)
771 ##
772 ##      Null deviance: 3630.06  on 6815  degrees of freedom
773 ## Residual deviance:  991.64  on 6789  degrees of freedom
774 ## AIC: Inf
775 ##
776 ## Number of Fisher Scoring iterations: 5

```

For the severity, we note that some claims may lead to no payment at all. As such, we use a Tweedie distribution as described follows. Note that this can also be fitted easily with the glm package as follows.

$$Y_i \sim \text{Tweedie}(\mu_i, p)$$

```

780 log( $\mu_i$ ) =  $\beta_0$  +  $\beta_1$ ClaimType +  $\beta_2$ Age +  $\beta_3$ AccidentDay +  $\beta_4$ AccidentMonth +  $\beta_5$ AccidentYear

```

```

library(statmod) #Has Tweedie distribution

#Fit severity model
fit_sev = glm(formula = Ultimate ~ Type+Age+AccWeekday+Month+Year,
              family= tweedie(link="log"),
              data = reported_claims)

#Predict values
predicted_sev=predict(fit_sev,
                     type = "response")

#Summarize results
summary(fit_sev)

```

```

781 ##
782 ## Call:
783 ## glm(formula = Ultimate ~ Type + Age + AccWeekday + Month + Year,
784 ##      family = tweedie(link = "log"), data = reported_claims)
785 ##
786 ## Deviance Residuals:
787 ##      Min       1Q   Median       3Q      Max
788 ## -52507   -3898   -1768    1649   275821
789 ##
790 ## Coefficients:
791 ##              Estimate Std. Error t value Pr(>|t|)
792 ## (Intercept)    8.400792   0.083791  100.26 < 2e-16 ***
793 ## Type2          2.068574   0.050490   40.97 < 2e-16 ***
794 ## Type3          0.453415   0.099184    4.57 4.9e-06 ***
795 ## Type4        -0.604514   0.102505   -5.90 3.9e-09 ***
796 ## Type5          0.771170   0.059966   12.86 < 2e-16 ***
797 ## Type6        -0.387238   0.108819   -3.56 0.00038 ***
798 ## Age            0.006432   0.000958    6.71 2.1e-11 ***

```

```

799 ## AccWeekdayTue 0.083967 0.046707 1.80 0.07226 .
800 ## AccWeekdayWed 0.061759 0.048959 1.26 0.20719
801 ## AccWeekdayThu 0.005145 0.049003 0.10 0.91639
802 ## AccWeekdayFri 0.112244 0.045992 2.44 0.01469 *
803 ## AccWeekdaySat -0.010668 0.052750 -0.20 0.83974
804 ## AccWeekdaySun 0.055143 0.049301 1.12 0.26339
805 ## Month02 0.012809 0.056791 0.23 0.82156
806 ## Month03 -0.124948 0.063706 -1.96 0.04988 *
807 ## Month04 0.006786 0.057313 0.12 0.90575
808 ## Month05 -0.182957 0.070241 -2.60 0.00921 **
809 ## Month06 -0.012272 0.065991 -0.19 0.85248
810 ## Month07 -0.220978 0.072909 -3.03 0.00245 **
811 ## Month08 0.133728 0.055057 2.43 0.01517 *
812 ## Month09 0.073201 0.065957 1.11 0.26712
813 ## Month10 -0.014423 0.065937 -0.22 0.82685
814 ## Month11 0.158102 0.057697 2.74 0.00616 **
815 ## Month12 0.226149 0.055529 4.07 4.7e-05 ***
816 ## Year13 -0.030494 0.030162 -1.01 0.31204
817 ## Year14 0.015982 0.034479 0.46 0.64301
818 ## Year15 -0.006777 0.241226 -0.03 0.97759
819 ## ---
820 ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
821 ##
822 ## (Dispersion parameter for Tweedie family taken to be 1.9e+08)
823 ##
824 ## Null deviance: 1.9739e+12 on 6815 degrees of freedom
825 ## Residual deviance: 1.2902e+12 on 6789 degrees of freedom
826 ## AIC: NA
827 ##
828 ## Number of Fisher Scoring iterations: 9

```

829 We can therefore combine these two components to provide an estimate of the reserve using the
830 model fitted on the pre-processing:

```

#Predict reserve: IBNR frequency x severity
ML = sum(predicted_freq*predicted_sev)
ML

```

```
831 ## [1] 10,809,767
```

832 We now proceed to illustrate how such working model can be adjusted with the use of the population
833 sampling techniques.

834 5.4 Post-Processing

835 For the post processing we illustrate the use of the AIPW estimator. This can easily implemented
836 once the micro-level model has been fitted as suaul.

```

#Estimate IPW weights
reported_claims$Weight =probs$.pred_survival/(1-probs$.pred_survival)

```

```

#Calculate augmentation term and AIPW
error_term = sum((reported_claims$CumPaid-predicted_sev)*reported_claims$Weight)
AIPW = ML+error_term

#Display results
cbind(ML_model = ML,
      Augmentation_term =error_term,
      AIPW = AIPW)

```

```

837 ##      ML_model Augmentation_term      AIPW
838 ## [1,] 10,809,767           1,387,960 12,197,727

```

839 5.5 In-Processing

840 For the in-processing approach we consider the same micro-level model, but this is fitted using
 841 weighted estimating equations instead. In this case, for the glm, we fit an alternative severity model
 842 but using a weighted maximum likelihood approach.

```

#Fit severity model using weighted likelihood
fit_sev_wl = glm(formula = Ultimate ~ Type+Age+AccWeekday+Month+Year,
                 family= tweedie(link="log"),
                 data = reported_claims,
                 weights = Weight)

#Predict values
predicted_sev_wl = predict(fit_sev_wl,
                          type = "response")

#Summarize estimation
summary(fit_sev_wl)

```

```

843 ##
844 ## Call:
845 ## glm(formula = Ultimate ~ Type + Age + AccWeekday + Month + Year,
846 ##      family = tweedie(link = "log"), data = reported_claims, weights = Weight)
847 ##
848 ## Deviance Residuals:
849 ##      Min       1Q   Median       3Q      Max
850 ## -57409    -948    -100     697   140525
851 ##
852 ## Coefficients:
853 ##              Estimate Std. Error t value Pr(>|t|)
854 ## (Intercept)    7.57037    0.11625   65.12 < 2e-16 ***
855 ## Type2         2.39560    0.06869   34.87 < 2e-16 ***
856 ## Type3        -0.23497    0.17452   -1.35  0.17821
857 ## Type4        -0.50533    0.13245   -3.82  0.00014 ***
858 ## Type5         1.29141    0.06820   18.94 < 2e-16 ***
859 ## Type6         1.09129    0.06962   15.67 < 2e-16 ***

```

```

860 ## Age                0.00800    0.00113    7.09  1.5e-12 ***
861 ## AccWeekdayTue      0.47904    0.05073    9.44  < 2e-16 ***
862 ## AccWeekdayWed      0.29396    0.06437    4.57  5.0e-06 ***
863 ## AccWeekdayThu      0.13830    0.06682    2.07  0.03851 *
864 ## AccWeekdayFri      0.57777    0.05917    9.76  < 2e-16 ***
865 ## AccWeekdaySat      0.11569    0.06910    1.67  0.09417 .
866 ## AccWeekdaySun      0.89427    0.04533   19.73  < 2e-16 ***
867 ## Month02            -0.54585    0.06401   -8.53  < 2e-16 ***
868 ## Month03            -0.05083    0.07862   -0.65  0.51792
869 ## Month04            -0.10565    0.08135   -1.30  0.19410
870 ## Month05            -0.27463    0.09088   -3.02  0.00252 **
871 ## Month06            -0.35845    0.10961   -3.27  0.00108 **
872 ## Month07            -0.42564    0.09243   -4.60  4.2e-06 ***
873 ## Month08             0.10238    0.06648    1.54  0.12362
874 ## Month09            -0.41975    0.10395   -4.04  5.4e-05 ***
875 ## Month10            -0.20821    0.06952   -3.00  0.00275 **
876 ## Month11            -0.37691    0.08432   -4.47  8.0e-06 ***
877 ## Month12             0.28942    0.06063    4.77  1.8e-06 ***
878 ## Year13              0.10597    0.05438    1.95  0.05140 .
879 ## Year14              0.07516    0.05009    1.50  0.13351
880 ## Year15              0.75500    0.07095   10.64  < 2e-16 ***
881 ## ---
882 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
883 ##
884 ## (Dispersion parameter for Tweedie family taken to be 4e+07)
885 ##
886 ##      Null deviance: 4.3335e+11  on 6815  degrees of freedom
887 ## Residual deviance: 2.7372e+11  on 6789  degrees of freedom
888 ## AIC: NA
889 ##
890 ## Number of Fisher Scoring iterations: 17

```

891 The fitted model display similarities with the previous fitted model for the severity (which was
892 fitted using the data of reported claims without any adjustments), yet there are some considerable
893 differences in the estimated parameters. This illustrates the different behavior in the claim severity
894 between the two populations.

895 We can therefore combine the frequency model with the alternative severity model to provide an
896 estimate of the reserve:

```

#Predict reserve: IBNR frequency x severity
ML_WL = sum(predicted_freq*predicted_sev_wl)
ML_WL

```

```

897 ## [1] 11,713,328

```

898 5.6 Pre-Processing

899 While one can use the weights $\frac{1-\pi_i(\tau)}{\pi_i(\tau)}$ for the number of repetition, we create the synthetic data set
900 of non-reported claims by generating realizations of weights according to a geometric distribution

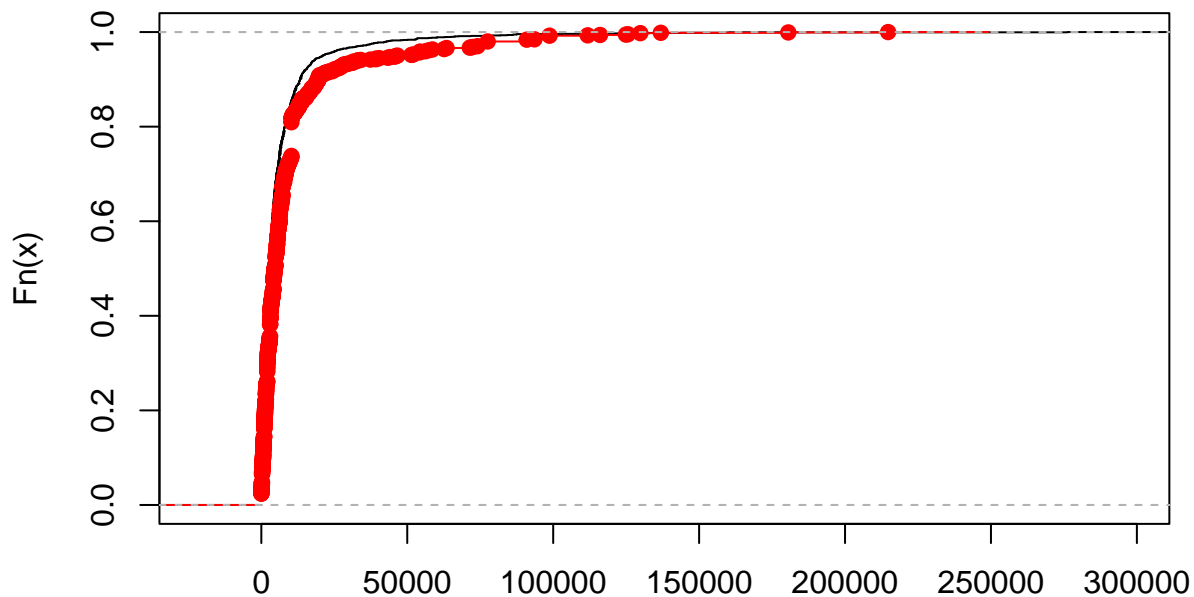
901 with success probability equal to the inclusion probability. This accounts for randomness in the
 902 data generation process. To illustrate the similarities of the synthetic non-reported claims and the
 903 real non-reported claims, we observe that the CDF of the distribution of the claim amount from the
 904 synthetic data set resembles considerably the one of the true non reported claims data.

```
#Create data set of synthetic claims
synthetic_non_reported = reported_claims %>%
  mutate(Rcount = rgeom(n= length(reported_claims$Ultimate),
                          prob = 1-probs$.pred_survival )) %>%
  uncount(weights = Rcount)

#Distribution of the non-reported claims
ecdf_nr = ecdf(non_reported_claims$Ultimate)

#Distribution of the synthetic non-reported claims
ecdf_s = ecdf(synthetic_non_reported$Ultimate)

#Plot distributions
plot(ecdf_nr, main=" ",
     xlab= "Claim amount of non-reported claims - Black: True data, Red: Synthetic data.")
lines(ecdf_s, col="red")
```



Claim amount of non-reported claims – Black: True data, Red: Synthetic data.

905
 906 Lastly, note that while we could directly fit the mcirelevel model on this data and have another
 907 prediction, we remark that the pre-processing approach leads automatically to an estimate of the

908 reserve by directly using the total amount from the synthetic data. We focus on the latter approach
 909 to illustrate the versatility of the methodology.

```
#Using the synthetic data set we can estimate the reserve right away
SYN = sum(synthetic_non_reported$Ultimate)
SYN
```

910 ## [1] 13,347,650

911 5.7 Discussion

912 The summary of the results are illustrated in the table below.

```
# True value and point estimates
true_value <- sum(non_reported_claims$Ultimate)
point_estimates <- c(true_value,ML, AIPW,ML_WL, SYN)
methods <- c("True value","Micro Level", "AIPW", "Weighted", "Synthetic")

# Create a data frame to hold the information
comparison_table <- data.frame(
  Method = as.character(methods),
  Estimate = point_estimates,
  Error = point_estimates - true_value,
  Percent_Error = (point_estimates - true_value) / true_value * 100
)

print(comparison_table)
```

913 ##	Method	Estimate	Error	Percent_Error
914 ## 1	True value	13,248,103	0	0.00
915 ## 2	Micro Level	10,809,767	-2,438,336	-18.41
916 ## 3	AIPW	12,197,727	-1,050,376	-7.93
917 ## 4	Weighted	11,713,328	-1,534,775	-11.58
918 ## 5	Synthetic	13,347,650	99,547	0.75

919 We can observe that all methodologies provide reasonable estimates of the IBNR reserve. In
 920 particular, it is to note that the estimates obtained using the aid of the population sampling
 921 methods provide estimates very close to the actual value of the reserve, which are better than the
 922 one obtained from the micro-level model alone. In particular, we note that regular frequency-severity
 923 model performs an under-estimation, but the AIPW augmentation term adjust this estimation
 924 up to the actual level of the reserve. Similarly, it is to note how the synthetic data approach
 925 provides an incredibly close estimate of the reserve, specially as this approach does not use any
 926 frequency-model, but only the associated inclusion probabilities. Lastly, in the case of the weighted
 927 likelihood approach, this provides an estimation that resembles the original frequency-severity model,
 928 nevertheless the correction in the estimated severity makes the prediction to be closer to the true
 929 value of the reserve.

6 Illustration with real data

In this section, we examine the performance of the various sampling-based approaches described earlier and compare them with traditional methods using a real dataset from a Canadian automobile insurance company. Because this dataset is proprietary and cannot be shared, we have also prepared a small simulation study—including the corresponding code for implementation—to illustrate the methodology. This is available at <https://htmlpreview.github.io/?https://github.com/sfcalcetero/AIPW/blob/main/CIA-report--F.html>. This simulation is intended for practical demonstration rather than serving as a comprehensive case study.

The Canadian dataset for this section contains detailed records of Physical Damage (PD) claims from January 2014 to December 2016, documenting individual claims, policyholder demographics, and automobile specifications. The information includes vehicular attributes such as age, horsepower, weight, and market value; policyholder characteristics like gender and age; and accident-specific details, including the time of occurrence, time of reporting, and the claim amount.

A key characteristic of this dataset is the significant difference in the claim amounts between reported and not reported claims, with not reported claims generally having lower severity. This discrepancy is common in auto insurance and is largely due to reporting behaviors. Larger, more severe claims are often reported immediately, leading to an over-representation of high-severity claims among those reported. In contrast, smaller claims, which may involve minor damages, are more likely to be delayed or not reported, resulting in a lower average severity for not reported claims. Figure 1 illustrates the distributions of the claim amounts for both groups, showing that reported claims tend to have a longer tail than not reported ones.

Along those lines, loss estimation methods based solely on reported claims tend to overestimate reserves if this difference is not accounted for. To preview how the population sampling approach can address this issue, we also plot in Figure 1 the distribution of a pseudo-population of not reported claims, synthetically generated using the methodology described in Section 4.5. This synthetic data set more closely resembles the distribution of true not reported claims than what would be obtained from reported claims alone.

In the following, we will estimate the reserves of not reported claims on a monthly basis, end of the month, using the period of time 2014 and 2015 as training and 2016 as the testing period.

6.1 Model fitting

To effectively illustrate the improvements gained from the sampling methods, we need a model to assist in the estimation. For this purpose, we first fit a micro-level model to our data. The model described below yielded the best goodness-of-fit among various distributional assumptions. For practical application, we refit the model every month to incorporate the latest available data, which means the model parameters change over time. The parameter values presented here are for the first month of the testing period, serving as an illustration.

Reporting delay time and inclusion probabilities

Here we estimate a single model that would be used for both the reporting delay time and the construction of the inclusion probabilities. As mentioned earlier, inclusion probabilities are a critical component for the application of the population sampling methods, so special attention is given to this step.

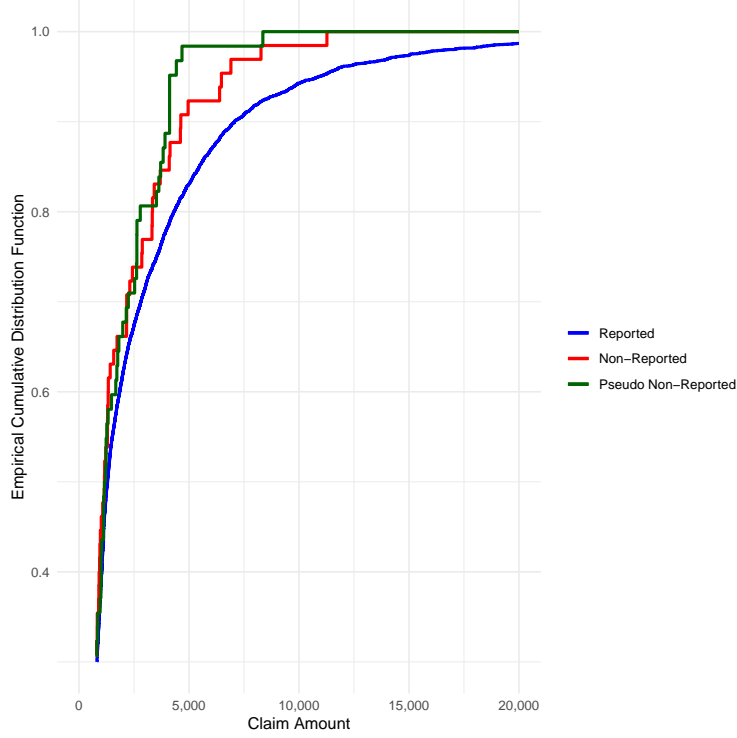


Figure 1: Empirical CDFs of the claim severity of reported and not reported claims for the first month in the testing set

971 Along those lines, we use a flexible model that accounts for the right-truncation of the reporting
 972 delay time. The structure selected is the one of a Cox regression model that describes the hazard
 973 function of the reporting delay time, $\lambda_{U|\mathbf{x}}(u)$, using the attributes of the policyholder \mathbf{x} as covariates,
 974 as follows:

$$\log(\lambda_U(u)) = \log(\lambda_0(u)) + \beta_1 \text{Car_Age} + \beta_2 \text{Claim_Count} + \beta_3 \log(\text{Horse_Power}) + \beta_4 \log(\text{Car_Weight}) \\ + \beta_5 \log(\text{Car_Price}) + \beta_6 \text{Gender} + \beta_7 \text{Driver_Age} + S_8(\text{Accident_day}) + S_9(\text{Claim_Amount}).$$

975 The log of the baseline hazard function, $\log(\lambda_0(u))$, is estimated using a B-Spline representation.
 976 Additionally, we incorporate nonlinear effects for the covariates "Accident_day" and "Claim_Amount"
 977 with the terms $S_8(\text{Accident_day})$ and $S_9(\text{Claim_Amount})$ respectively, which are also estimated
 978 using a B-Spline representation. The implementation is done via a generalized additive model with
 979 a piecewise exponential function that accounts for right-truncation. This is available in R packages
 980 such as `flexsurvreg`, `pamtools` or `reservr`.

981 The fitted parameters are shown in Table 1 and Figure 2. All regression coefficients are statistically
 982 significant. The baseline hazard function indicates that most claims are reported within the first two
 983 months, but there is a long tail for claims with extended reporting times. Furthermore, the claim
 984 size has an almost increasing effect, suggesting that larger claims tend to be reported more quickly.
 985 This is consistent with the observed greater severity for reported claims than for not reported.
 986 Finally, the effect of accident day on the hazard function reveals a quarterly seasonality.

987 Lastly, we use the model to obtain the inclusion probabilities, which can be computed by plugging

Coefficient	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Value	7.08	-17.41	4.28	-38.62	-255.88	36.46	-2.03
Std. Err.	0.07	0.38	1.66	2.09	1.06	0.49	0.01
p-val	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Table 1: Estimated regression coefficients of the Cox regression model for the reporting delay time, scaled by $\times 10^3$

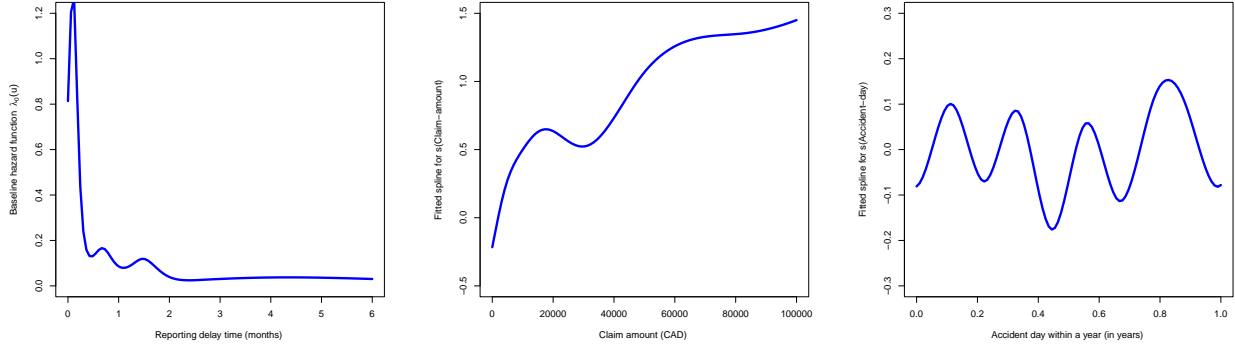


Figure 2: Fitted baseline hazard function for the reporting delay time (left panel) and fitted effects of the claim amount (center panel) and the accident day on the hazard of the reporting delay time (right panel)

988 in the estimated hazard function into the expression below:

$$\hat{\pi}_i(\tau) = Pr(U_i \leq \tau - T_i \mid \mathbf{x}_i, T_i, Y_i) = 1 - \exp\left(-\int_0^{\tau-T_i} \lambda_U(u) du\right). \quad (8)$$

989 Frequency model

990 We found that the usual Poisson process assumption for the number of claims per policyholder did
991 not fit well due to significant overdispersion and an excess of zero claims. To address this, we used
992 a practical approach by fitting a model for the random variable representing the number of claims
993 up to the valuation time τ and applied thinning to derive the distribution of not reported claims.

In our dataset, the Zero-Inflated Negative Binomial (ZINB) model provided a good fit. The ZINB has the following probability mass function

$$P_{ZINB}(k; q, \theta, r) := q\mathbf{1}_{\{k=0\}} + (1-q) \binom{k+r-1}{k} \left(\frac{\theta}{\theta+r}\right)^k \left(\frac{1}{\theta+r}\right)^r,$$

994 where q is the probability of zero inflation, θ is associated with the mean of the negative binomial
995 distribution, and r is a dispersion parameter. Along those lines, the frequency model is the following:

- Let the total number of claims from the j -th contract up to τ , $N_j(\tau)$, follow a ZINB regression:

$$P(N_j(\tau) = n) = P_{ZINB}(n; q_j, \xi_j \theta_j, r),$$

$$\begin{aligned} \text{logit}(q_j) = & \beta_0 + \beta_1 \text{Car_Age} + \beta_2 \text{Claim_Count} + \beta_3 \log(\text{Horse_Power}) + \beta_4 \log(\text{Car_Weight}) \\ & + \beta_5 \log(\text{Car_Price}) + \beta_6 \text{Gender} + \beta_7 \text{Driver_Age} + \beta_8 \xi, \end{aligned}$$

$$\begin{aligned} \log(\theta_j) = & \beta_0 + \beta_1 \text{Car_Age} + \beta_2 \text{Claim_Count} + \beta_3 \log(\text{Horse_Power}) + \beta_4 \log(\text{Car_Weight}) \\ & + \beta_5 \log(\text{Car_Price}) + \beta_6 \text{Gender} + \beta_7 \text{Driver_Age} \end{aligned}$$

where r is a common dispersion parameter, and ξ_j is a known exposure term.

- The arrival time T of a claim has cumulative distribution function $F_T(t)$, which may depend on covariates. In our dataset, and commonly observed in reserving, accident times are uniformly distributed over the life of the contract, so we use this distribution.
- Assuming independence of delays for each claim, the marginal distributions for reported claims and not reported claims are given by:

$$P(N_j^R(\tau) = k) = P_{ZINB}(k; q_j, p_j(\tau) \xi_j \theta_j, r) \quad (9)$$

$$P(N_j^{IBNR}(\tau) = k) = P_{ZINB}(k; q_j, (1 - p_j(\tau)) \xi_j \theta_j, r) \quad (10)$$

where $p_j(\tau) = \int_0^\tau \hat{\pi}_i(\tau - t) dF_T(t)$ is the average inclusion probability. For uniform accident times, this simplifies to $p_j(\tau) = \frac{1}{\tau} \int_0^\tau \pi_j(\tau - t) dt$.

- For prediction, we can use the conditional distribution of not reported claims given reported claims. By Bayes' rule, we find that

$$P(N_j^{IBNR}(\tau) = k \mid N_j^R(\tau)) = P_{ZINB}(k; \tilde{q}_j, (1 - p_j(\tau)) \xi_j \tilde{\theta}_j, \tilde{r}_j)$$

where $\tilde{q}_j = \frac{q_j \mathbf{1}_{\{N_j^R(\tau)=0\}}}{q_j + (1 - q_j)(1 + \theta_j \xi_j p_j(\tau))^{-r}}$, $\tilde{\theta}_j = \frac{\theta_j}{1 + \theta_j \xi_j p_j(\tau)}$, and $\tilde{r}_j = r + N_j^R(\tau)$. Thus, the expected number of not reported claims for the j -th policyholder, denoted as λ_j^{IBNR} , is:

$$\lambda_j^{IBNR} = (1 - \tilde{q}_j) \tilde{r}_j (1 - p_j(\tau)) \xi_j \tilde{\theta}_j \quad (11)$$

This approach has not been used in reserving before and offers a flexible alternative to the Poisson process while remaining analytically tractable. Unlike the Poisson process, this method incorporates the current number of reported claims into the prediction, similar to the dynamic claim score by Yanez et al. (2024), which considers historical claiming behavior.

Using the reported claims, we fit the ZINB regression for reported claims as in Equation (9) via maximum likelihood, and then plug the estimated parameters in Equation (11) for predictions. This can be done with R packages for GLMs, such as `pscl`, where the term $\xi_j p_j(\tau)$ acts as an offset. The fitted model for the first month in the testing period is:

Coefficient		β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8 & r
$\text{logit}(q)$	Value	1.13	0.10	-0.25	0.28	0.74	-1.09	0.19	0.01	1.31
	Std. Err.	2.68	0.02	0.09	0.43	0.61	0.35	0.12	0.00	0.06
	p-val	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
$\log(\theta)$	Value	-1.82	-0.02	0.05	0.13	0.22	-0.42	0.12	0.00	16.37
	Std. Err.	1.61	0.01	0.05	0.25	0.34	0.19	0.07	0.00	-
	p-val	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	-

Table 2: Estimated regression coefficients of the ZINB model for the number of claims

Severity models

This is the main component of the reserving model where the sampling methods play a crucial role. As noted earlier, the severity distribution differs significantly between reported and not reported claims. Therefore, we fit three versions of the severity model under different configurations: one using the regular approach (i.e., without incorporating the sampling design into the estimation), and the other two using population sampling adjustments.

For the general regression model, we found that a lognormal regression was well-suited to our data, as specified below. We will also assume that claim amounts are iid for a given contract.

$$\begin{aligned} \log(Y_i) &\sim \text{Norm}(\nu_i, \sigma^2) \\ \nu_i &= \beta_0 + \beta_1 \text{Car_Age} + \beta_2 \text{Claim_Count} + \beta_3 \log(\text{Horse_Power}) + \beta_4 \log(\text{Car_Weight}) \\ &\quad + \beta_5 \log(\text{Car_Price}) + \beta_6 \text{Gender} + \beta_7 \text{Driver_Age} \end{aligned}$$

This model can be easily fitted using a least-squares procedure, such as `lm` in R, on the log-scale of the claim amounts. The prediction, given by the mean, is then:

$$\hat{Y}_i = \exp(\nu_i + \sigma^2/2).$$

Now, we proceed to estimate the model and the respective adjustments based on sampling theory. Conceptually, we are working with the same model specification, i.e., the same distribution and regression formula as mentioned above. However, these variations will differ in the estimated parameters as follows:

- The first estimation assumes that severity and reporting are independent. In this case, we use maximum likelihood as usual, which coincides with minimization using unweighted least squares. We denote the point prediction as \hat{Y}_i . The fitted parameters are:

Coefficient	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	σ
Value	7.10	-0.03	-0.04	0.01	-0.44	0.36	0.03	-0.002	0.73
Std. Err.	0.68	0.00	0.02	0.10	0.14	0.07	0.03	0.0001	
p-val	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	

Table 3: Estimated regression coefficients of the log-normal model for the claim amount

- The second model is estimated using the weighted loss function approach, as proposed in Section 4.4. This no longer imposes independence with reporting on the estimated model. In this case, instead of maximizing directly the log-likelihood, we maximize its weighted version, i.e. $\sum_{i=1}^{N^R} \frac{1 - \hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} \log f(Y_i)$, which coincides with a weighted least squares procedure for our selected model. The fitted parameters are as follows:

Coefficient	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	σ
Value	7.98	-0.01	-0.01	0.31	-0.99	0.47	0.03	-0.0001	0.84
Std. Err.	0.48	0.003	0.02	0.07	0.09	0.05	0.05	0.001	
p-val	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	

Table 4: Estimated regression coefficients of the log-normal model for the claim amount using weighted least squares

Although the estimated parameters show similarities with the first model, the values differ. This highlights the differences between the severity of reported and not reported claims. Recall that this estimated model is equivalent to the one that would be estimated if we used the synthetic data set. As such, we do not make a distinction and will denote the point prediction as \hat{Y}_i^{WL} . Nevertheless, the synthetic data set approach may be used to construct bootstrapped versions of the coefficients.

- The third model is a variation of the first, where we impose the weighted balance property as described in Section 4.3. We denote the point prediction as \hat{Y}_i^{wBP} . The fitted parameters are the same as those of the first model, except the intercept of the models, which changes via the multiplicative coefficient from Equation (7) as:

$$\hat{Y}_i^{wBP} = \left(\frac{\sum_{i=1}^{N^R(\tau)} \frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} Y_i}{\sum_{i=1}^{N^R(\tau)} \frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} \hat{Y}_i} \right) \hat{Y}_i = 0.70 \hat{Y}_i$$

As the coefficient is less than 1, the adjusted model will predict a reduced severity when compared to the model that assumes independence. Note that this is consistent with the fact that the severity of not reported claims tends to be lower than that of reported claims.

6.2 Estimation of the reserve

Here we present the reserve estimations during the testing period. We compare the actual reserve values, both in terms of amount and number, with the predictions from various models detailed in Table 5. These models incorporate different combinations of components from the previous section, and the results are shown in Figures 4 and 3, as well as Table 6.

The models selected for comparison in Table 5 are categorized based on the nature of their estimation methods. The primary objective is to demonstrate how the sampling-motivated reserve estimators can outperform traditional methods, highlighting any potential improvements. To this end, the first two models are classic reserving methods, serving as a baseline for comparison. The third, fourth and fifth models showcase the main sampling-motivated modifications for addressing bias, including the AIPW estimator. The last two models explore alternative structures of the AIPW that represent an intermediate approach between micro- and macro-models. The predictions based solely on the synthetic dataset of not reported claims yield point estimations equivalent to the regular IPW estimator and are therefore not listed separately in the analysis.

Figure 3 shows the point estimates for the number of not reported claims compared to the actual values for each month. These are obtained by setting $Y_i = \hat{Y}_i = 1$ in all the formulas. As some of the methods listed in Table 5 differ only in their severity component, they are not listed as they lead to identical predictions for the number of claims. The results suggest that the various methods are competitive in predicting the number of not reported claims during the testing period, with no consistent pattern of over- or underestimation. However, the CL method consistently underestimates the number of not reported claims. Overall, the methods generally capture the trend in fluctuations of not reported claims, with the ML model exhibiting behavior closest to the actual data. These findings imply that the fitted models for frequency and inclusion probabilities are appropriately specified, effectively capturing the nuances of claim counts and reporting delays.

Figure 4 presents the point estimates for the reserve generated by all methods compared to the true values for each month. The results indicate that the ML reserve consistently overestimates the reserve throughout the testing period. Given that the models for frequency and reporting delay

Model	Description	Expression for Prediction
<i>Traditional reserving models</i>		
ML	Prediction using the components of the micro-level model.	$\sum_{j=1}^M \lambda_j^{IBNR} \hat{Y}_j$
CL	Prediction using the traditional development factors $(f_k(\tau))$ approach.	$\sum_{k=1}^m \left(\sum_{i T_i=k} Y_i \right) \times (f_k(\tau) - 1)$
<i>AIPW estimators</i>		
AIPW	Prediction of ML along with the augmentation term	$\sum_{j=1}^M \lambda_j^{IBNR} \hat{Y}_j + \sum_{i=1}^{N^R(\tau)} \frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} (Y_i - \hat{Y}_i)$
IPW	Prediction using Equation (1), with probs from the Cox model. a.k.a Micro-level CL.	$\sum_{i=1}^{N^R(\tau)} \frac{1-\hat{\pi}_i(\tau)}{\hat{\pi}_i(\tau)} Y_i$
AIPW-CL	Same as AIPW, but using probabilities implied by the CL	$\sum_{j=1}^M \lambda_j^{IBNR} \hat{Y}_j + \sum_{i=1}^{N^R(\tau)} \frac{1-\pi_i^{CL}(\tau)}{\pi_i^{CL}(\tau)} (Y_i - \hat{Y}_i)$
<i>Other sampling methods</i>		
ML-wBP	Same as ML, but with modified intercept for balance property as in Equation (7)	$\sum_{j=1}^M \lambda_j^{IBNR} \hat{Y}_j^{wBP}$
ML-WL	Same as ML, but severity estimated with weighted likelihood	$\sum_{j=1}^M \lambda_j^{IBNR} \hat{Y}_j^{WL}$

Table 5: Models in the numerical comparison: Traditional Reserving Models, AIPW estimators, and other methods. Note that ML-wBP can be thought of as an AIPW estimator as the associated augmentation term would be 0.

appropriately fit the number of claims, this overestimation is therefore attributed to the severity component of the model. This finding aligns with the earlier observation that the mean severity of reported claims (which are used to fit the model) is higher than that of not reported claims.

Metric	ML	CL	AIPW	ML-wBP	ML-WL	IPW	AIPW-CL
ME	-37,605	28,400	9,563	7,874	5,924	-6,691	11,668
MAE	49.155	49,454	31,016	31,921	33,476	43,482	35,131
RMSE	57,356	55,288	40,669	40,974	42,240	50,896	52,027
MAPE	34.9%	29.3 %	19.6%	20.4%	21.6%	28.2%	19.6%

Table 6: Error metrics for the total of the reserves over the testing period. ME: Mean error, RMSE: Root mean square error, MAE: Mean absolute error, MAPE: Mean absolute percentage error. Bolded text indicates the model with the best performance on each row.

The performance of the ML model adjustments using the AIPW, ML-wBP, and ML-WL methods is notable. Figure 4 and Table 6 show that these approaches yield nearly identical estimates, with slight differences in some months. This similarity, expected given their shared goal of bias correction, highlights the significant bias in the original ML model. The unbiased estimates are closer to the predictions of these methods, which effectively detect and adjust the ML model’s positive bias. In most cases, the correction is downward, except in month 8, where an upward adjustment accounts for a peak in the reserve.

A key feature of these adjustments is their ability to enhance not only the point prediction but also the reserve’s variability. While the ML model provides stable but overly smooth estimates over time, the adjusted methods capture the reserve’s fluctuations more accurately. This ability to reflect

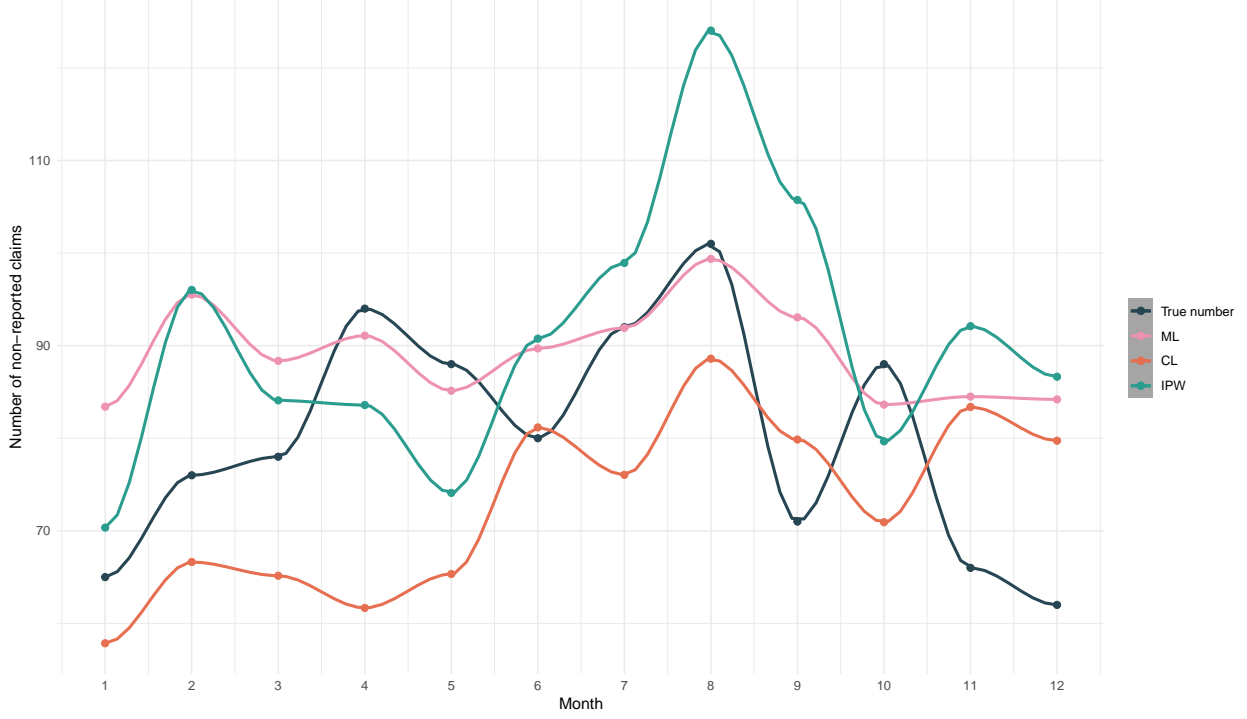


Figure 3: Estimation for the number of not reported claims per month

the reserve’s dynamic behavior underscores the value of integrating sampling theory-inspired bias correction techniques with micro-level models. Table 6 quantitatively demonstrates the improvement, with the AIPW estimator achieving the most significant error reduction.

Turning to the CL, IPW, and AIPW-CL methods, the CL approach underestimates the reserve but captures its overall trend, which is valuable from a macro-level perspective. The IPW estimator provides a closer approximation to the actual reserve while maintaining a similar trend. These two methods behave similarly because the CL approach simplifies the IPW estimator. By incorporating claim and policyholder attributes into the inclusion probabilities and development factors, the IPW method significantly improves the CL estimates. Notably, these methods do not rely on frequency/severity models, yet their reserve predictions remain competitive.

The AIPW-CL method combines elements of the ML model’s stability with the CL method’s ability to capture fluctuations. Its estimates, blending the strengths of both models, deliver more desirable predictions than either approach individually. This illustrates how the AIPW estimator integrates the corrective strengths of both methods into a unified prediction. As discussed in Section 3.4, the IPW and AIPW-CL methods act as intermediaries between micro-level and macro-level models, leveraging the advantages of both frameworks. The AIPW estimator, striking a balance between the trend behavior of macro-level approaches and the accuracy of micro-level predictions, offers enhanced predictive power overall.

7 Conclusions

A key barrier to adopting micro-level reserving is its significant methodological difference from macro-level models, such as the Chain-Ladder, and the effort required to develop it. The transition

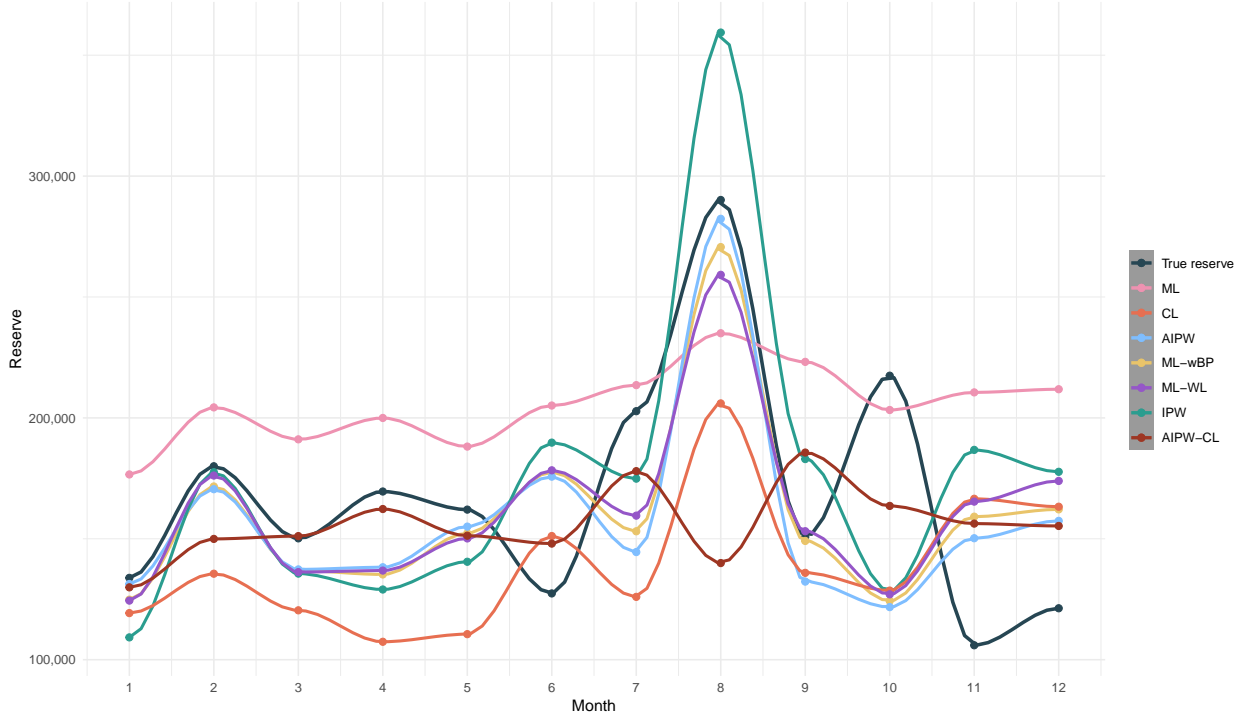


Figure 4: Estimation for the outstanding claims reserve per month

from macro to micro-level models is challenging for insurance companies and difficult for regulators to validate, as the foundational principles differ greatly from those of the Chain-Ladder and development factors. This gap between the two reserving methodologies hinders the adoption of micro-level models in the insurance sector.

To face this issue, this report introduces a novel perspective to claim reserving by applying principles from population sampling theory, specifically through the use of the augmented inverse probability weighting (AIPW) estimator. The key theoretical contribution of this work is that the AIPW estimator integrates the aggregate perspective of macro-level models with the detailed insights of micro-level models, offering a unified statistical framework that enhances both the robustness and accuracy of reserve estimates. This approach effectively bridges the gap between traditional macro-level and micro-level methodologies. Additionally, the sampling framework can provide new theoretical insights into the statistical inference of reserving methods, including those traditionally viewed as ad-hoc procedures, such as the Chain-Ladder method.

Moreover, adopting a population sampling perspective for reserving also offers a new understanding of the reserving process, emphasizing the importance of addressing the sampling bias caused by the partially observed claims. The population sampling literature provides practical and theoretically sound methods to address the diverse challenges posed by the sampling bias. Specifically, the use of augmentation terms for correction, weighted balance properties, weighted estimation equation for model estimation, and generation of synthetic data sets have the potential to improve reserve estimation for a given reserving model. All of this is corroborated empirically

The numerical results are promising, demonstrating that the AIPW estimator, along with other sampling-based estimators, improves the precision of reserve estimates and better captures the variability and uncertainty inherent in claims data, with the AIPW estimator showing potential to

1130 outperform traditional models in terms of both accuracy and adaptability.

1131 Future research could explore other methods that address sampling bias and their applications to
1132 reserving problems. For example, Equation (2) can be used as an alternative to combine reserving
1133 models for the reported claims along with the IPW principle. Similarly, other techniques designed
1134 for missing data or those from causal inference contexts may offer valuable insights. Specifically,
1135 the process of adjusting model predictions based on observed errors in reported claims aligns with
1136 Bayesian approaches using posterior distributions. Thus, incorporating Bayesian models into claim
1137 reserving may provide an interesting and valuable perspective. Similarly, this report demonstrates
1138 the benefits of employing sampling-based methods rather than focusing solely on finding the optimal
1139 configuration. Future research could further explore the best structures for AIPW estimators, as
1140 well as for other methods to have optimal reserving predictions.

References

- American Academy of Actuaries (2023). An Actuarial View of Data Bias: Definitions, Impacts, and Considerations. *American Academy of Actuaries Risk Management and Financial Reporting Council*. Available at: https://www.actuary.org/sites/default/files/2023-07/risk_brief_data_bias.pdf.
- Antonio, K. and Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014(7):649–669.
- Arnab, R. (2017). *Survey sampling theory and applications*. Academic Press.
- Bücher, A. and Rosenstock, A. (2024). Combined modelling of micro-level outstanding claim counts and individual claim frequencies in non-life insurance. *European Actuarial Journal*, pages 1–33.
- Calcetero-Vanegas, S., Badescu, A. L., and Lin, X. S. (2024). Claim reserving via inverse probability weighting: a micro-level chain-ladder method. *European Actuarial Journal*, pages 1–47.
- Calcetero-Vanegas, S., Badescu, A. L., and Lin, X. S. (2025). A population sampling framework for claim1 reserving in general insurance. *arXiv preprint arXiv:2502.15598*.
- Calcetero-Vanegas, S. and Lin, X. S. (2024). Applications of population sampling to insurance ratemaking and reserving. *Japanese Journal of Statistics and Data Science*, 7(2):1155–1170.
- Campo, B. D. and Antonio, K. (2023). Insurance pricing with hierarchically structured data an illustration with a workers’ compensation insurance portfolio. *Scandinavian Actuarial Journal*, 2023(9):853–884.
- Charpentier, A. (2024). *Insurance, biases, discrimination and fairness*. Springer.
- Charpentier, A. and Pigeon, M. (2016). Macro vs. micro methods in non-life claims reserving (an econometric perspective). *Risks*, 4(2):12.
- Crevecoeur, J., Robben, J., and Antonio, K. (2022). A hierarchical reserving model for reported non-life insurance claims. *Insurance: Mathematics and Economics*, 104:158–184.
- Francis, L. (2016). Astin working party releases report on reserving practices for general insurance worldwide. <https://ar.casact.org/astin-working-party-releases-report-on-reserving-practices-for-general-insurance-worldwide>. Accessed: 2023-06-04.
- Fung, T. C., Badescu, A. L., and Lin, X. S. (2021). A new class of severity regression models with an application to ibnr prediction. *North American Actuarial Journal*, 25(2):206–231.
- George, B., Seals, S., and Aban, I. (2014). Survival analysis and regression models. *Journal of nuclear cardiology*, 21(4):686–694.
- Hiabu, M. (2017). On the relationship between classical chain ladder and granular reserving. *Scandinavian Actuarial Journal*, 2017(8):708–729.
- Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data.
- Kim, J. K. and Shao, J. (2021). *Statistical methods for handling incomplete data*. Chapman and Hall/CRC.

1178 McLachlan, G. J. and Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley & Sons.

1179 Peiris, H., Jeong, H., Kim, J.-K., and Lee, H. (2024). Integration of traditional and telematics data
1180 for efficient insurance claims prediction. *ASTIN Bulletin: The Journal of the IAA*, 54(2):263–279.

1181 Ross, S. M. (2022). *Simulation*. academic press.

1182 Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer
1183 Science & Business Media.

1184 Seaman, S. R. and Vansteelandt, S. (2018). Introduction to double robust methods for incomplete
1185 data. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 33(2):184.

1186 Shao, J. (2008). *Mathematical statistics*. Springer Science & Business Media.

1187 Taylor, G. (2019). Loss reserving models: Granular and machine learning forms. *Risks*, 7(3):82.

1188 Tee, L. and Käärik, M. (2022). Estimating ibnr claim counts using different levels of data aggregation.
1189 *Communications in Statistics-Simulation and Computation*, 51(6):3013–3036.

1190 Vakulenko-Lagun, B., Mandel, M., and Betensky, R. A. (2019). Inverse Probability Weighting
1191 Methods for Cox Regression with Right-Truncated Data. *Biometrics*, 76(2):484–495.

1192 Waernbaum, I. and Pazzagli, L. (2023). Model misspecification and bias for inverse probability
1193 weighting estimators of average causal effects. *Biometrical Journal. Biometrische Zeitschrift*,
1194 65(2):e2100118.

1195 Wahl, F., Lindholm, M., and Verrall, R. (2019). The collective reserving model. *Insurance:
1196 Mathematics and Economics*, 87:34–50.

1197 Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems.
1198 *Journal of econometrics*, 141(2):1281–1301.

1199 Wüthrich, M. V. (2018). Neural networks applied to chain–ladder reserving. *European Actuarial
1200 Journal*, 8:407–436.

1201 Wüthrich, M. V. (2020). Bias regularization in neural network models for general insurance pricing.
1202 *European Actuarial Journal*, 10(1):179–202.

1203 Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*. John
1204 Wiley & Sons.

1205 Wüthrich, M. V. and Merz, M. (2023). *Statistical foundations of actuarial learning and its
1206 applications*. Springer Nature.

1207 Yanez, J. S., Boucher, J.-P., and Pigeon, M. (2024). Modeling payment frequency for loss reserves
1208 based on dynamic claim scores. *North American Actuarial Journal*, 28(3):491–512.