

Premium Estimation Under Model Uncertainty: Model Averaging for Left-Truncated Reinsurance Losses

Tatjana Miljkovic^{1a}

¹ Department of Statistics, Miami University, Oxford, Ohio, USA

Keywords: Model averaging, Excess of loss premium, Reinsurance, Mixture models

<https://doi.org/10.66573/001c.141749>

Variance

Vol. 18, 2025

In this paper, model averaging (MA) is proposed as a technique to address model uncertainty when estimating the premium for aggregate excess of loss reinsurance. Despite its intrinsic appeal, the use of MA in premium calculations involving left-truncated reinsurance losses remains largely unexplored. In this study, we examine the effectiveness of MA in estimating excess of loss premiums using both real-world data and Monte Carlo simulations. A model space of finite mixtures based on lognormal and gamma distributions is considered for fitting the left-truncated reinsurance losses. Premium estimates are computed based on all models in the considered model space. For a given retention level, closed-form solutions have been derived for the MA premium estimators based on gamma and lognormal mixtures. We examine how well MA estimators perform in comparison to estimators derived from the best model within a given model space, determined by criteria like the Akaike information criterion and Bayesian information criterion. This investigation offers an alternative approach for actuaries and risk managers to estimate premiums while taking model uncertainty into account.

Address for Correspondence: miljkot@miamioh.edu

1. INTRODUCTION

The aggregate excess of loss reinsurance plays a crucial role in managing risk, maintaining financial stability, and enabling insurers to operate efficiently in a competitive market. In excess of loss reinsurance, risk transfer occurs as the insurer (ceding) company transfers a defined portion of its risk to a reinsurer, which agrees to indemnify the ceding company for losses exceeding a specified retention level up to an agreed-upon limit in exchange for a premium payment. Thus, the premium associated with excess of loss reinsurance represents a critical component of this risk transfer. It represents "the price of risk" (see Deelstra and Plantin 2014). Premium estimation is based on modeling the ceding company's historical loss data, the nature of the risks being insured, and other relevant factors. When actuaries analyze a ceding company's historical losses, they often seek a suitable parametric model. However, in this scenario, there are typically several models that adequately

describe the data, and the actuary may lack a priori reasons to prefer one model over the others. This problem is further compounded by the fact that premium estimates, derived from similarly fitting models, can vary significantly. Since the true model is unknown, there exists model uncertainty in the premium estimation process. Although De Vylder (1996) has highlighted the importance of studying the uncertainty of excess of loss premiums and other distribution functions, this practice has not been widely implemented. We argue that model uncertainty constitutes a significant portion of the uncertainty in premium estimation and should not be overlooked. Therefore, it is crucial to examine the sensitivity of premium estimates to the choice of model, particularly in the context of risk modeling and pricing for aggregate loss premiums. This paper aims to explore this topic in depth.

Over the past few decades, studying model uncertainty risk gained significant attention in statistics, finance, business, social sciences, and many other fields. The interest

^a Tatjana Miljkovic is an associate professor and actuarial science adviser in the Department of Statistics at Miami University. Her research focuses on actuarial science and applied statistics, including mortality modeling, mixture modeling, the EM algorithm, tail risk measures, and insurance capital requirements. She applies statistical methods to health and climate change and collaborates across disciplines. Dr. Miljkovic's research work bridges theory and practice, enriching both her teaching and scholarly contributions. She has published in top actuarial journals and serves as associate editor of the *Journal of Statistical Computation and Simulation*.

in this area of research can be traced to pivotal papers by Madigan and Raftery (1994), Raftery (1995), and Hoeting et al. (1999) that explored the Bayesian model averaging (BMA) approach. By analogy with the Bayesian methods, Buckland, Burnham, and Augustin (1997) advocated the use of information criteria and introduced the frequentist approach to model averaging (MA). Both Bayesian and frequentist approaches recognize that the choice of a single “best” model is often inadequate when quantifying the uncertainty of statistical estimators. MA allows us to control for model uncertainty associated with the quantity of interest by pooling the useful information from all candidate models in the model space, not just a single best model.

BMA works on prior probabilities for a list of candidate models, along with the priors for the parameters of each model, in order to derive the posterior distribution of any parameter of interest. Hoeting et al. (1999) highlighted that BMA provides better average predictive performance in regression setting than any single model that could be selected, demonstrated in a range of applications involving different model classes and types of data. The authors claimed that BMA provides inference about parameters that takes into account an important source of model uncertainty, and it was found that BMA-based confidence intervals are better calibrated than confidence intervals based on a single model. Cairns (2000) used the Bayesian paradigm to discuss parameter and model uncertainty in insurance applications. The author found that the inclusion of parameter uncertainty made a significant contribution to the outcome of the modeling. Further, the quantity of data is found to have a significant impact on the parameter and model uncertainty. In the area of reinsurance, it may not be realistic to assume that a significant amount of data is available, so the problem of parameter and model uncertainty should not be ignored.

The frequentist MA approach deals with model uncertainty not by having the user select a single model from among the set of candidate models according to the Akaike information criterion (AIC; Akaike 1974) or Bayesian information criterion (BIC; Schwarz 1978) but rather by averaging over the set of candidate models in a particular manner using the model weights computed based on the AIC or BIC. This approach was discussed by Buckland, Burnham, and Augustin (1997). Miljkovic and Grün (2021) utilized the same approach in computing model-averaged risk measures such as value-at-risk (VaR) and conditional tail expectation (CTE). The authors investigated a class of 196 composite models for calculating model-average VaR and CTE and found that the performance of the model average estimator of CTE is significantly better than that of the estimator obtained with the best-model approach when considering the full model space. The performance of the model average estimator of VaR using the full model space is comparable to that of the estimator obtained using the best-model approach based on the AIC or BIC; however, it is still recommended to use the model average estimator of VaR because it accounts for model uncertainty. Frequentist MA can be employed to account for model uncertainty in the

excess of premium calculation, which is the focus of this paper.

A majority of papers in the field of insurance loss modeling aim to find a single model or class of models that can provide a better fit to the data than other candidate models or classes. This single model or model class is then used in the actuarial applications related to pricing of reinsurance contracts or risk management. Suitable models have been proposed for the global fitting of insurance losses, capturing both the body and the tail of the loss distribution. These models serve as the basis for estimating the stop-loss premium at different retention levels. Early work in this area by Beirlant, Matthys, and Dierckx (2001) involved fitting the generalized Burr-gamma distribution to aggregate losses to assess a portfolio of risks in both the tails and more central portions of the claim distribution, with the aim of estimating the excess of loss premium under different retention levels. Beirlant et al. (2006) and Klugman, Panjer, and Willmot (2012) considered splicing models, such as the exponential distribution with the Pareto distribution, to produce suitable models for computing the excess of loss premium. Verbelen et al. (2015) demonstrated the calculation of the excess of loss premium when left-truncated aggregated losses are fitted using an Erlang mixture. An extension of this work was considered by Reynkens et al. (2017), which proposed a splicing model with a mixed Erlang distribution for the body and a Pareto distribution for the tail to accommodate global fitting. The authors showed how the excess of loss premium can be computed for different retention levels based on this model in the presence of truncated and censored data. While these models contribute to the loss modeling literature, none have yet addressed the essential aspect of model uncertainty in estimating the excess of loss premium. This highlights the importance of the current study in order to fill the gap in the literature.

Some other approaches for selection of the best model have been explored. Jullum and Hjort (2017) considered both nonparametric and non-nested parametric models to quantify a particular parameter of interest, referred to as the focus parameter (FP). Examples of the FP include a quantile, a standard deviation, kurtosis, interquartile range, etc. Unlike AIC and BIC approaches in selecting the best model, the focus information criterion (FIC) begins by defining a population quantity of interest (the FP) and then proceeds by estimating the mean squared error of different model estimates for this parameter. The goal is to select the model that provides the most accurate inference for a specific parameter or distributional aspect. The same authors also considered an average-weighted version, AFIC, which allows multiple FPs to be analyzed simultaneously. Wang and Hobæk (2019) applied the FIC framework to estimate a single quantile of the claim severity distribution as an FP, an important measure of performance in risk assessment. Their findings indicate that the FIC approach is often comparable to, or significantly better than, the BIC-based model selection—particularly in scenarios where the data exhibit heavy tails, the sample size is relatively small, or the quantile of interest is located far in the tail of the

loss distribution. Motivated by this stream of literature on FIC, one could explore its use in estimating the excess of loss premium by computing FIC-based model weights. In this case, the objective would be to select the model that best estimates the FP of interest (i.e., reinsurance premium at a given retention level). However, this approach presents several challenges. Specifically, selecting an optimal model separately for each retention level can introduce inconsistencies and instability in estimating premiums, making it more complicated to quantify uncertainty. Since FIC-based selection may lead to different models being chosen at different retention levels, it can result in discontinuities in premium estimates, reducing the robustness and reliability of the pricing framework. Thus, AIC- or BIC-based MA should provide a more stable and consistent approach to estimating premiums across different retention levels, while effectively balancing goodness-of-fit and model complexity.

This study aims to incorporate model uncertainty into the pricing of the excess of loss reinsurance while considering a predetermined retention level in the presence of left-truncated aggregate loss data. The study proposes utilizing model average estimators based on AIC and BIC weights for computing the excess of loss premium within a model space that encompasses several finite mixture models with gamma and lognormal component distributions. The MA using AIC or BIC is proposed because it provides stability and robustness in estimating premiums across all levels; a smooth premium function across different retention levels is critical in insurance pricing. With this scenario, the closed-form formulas are derived for the premium calculations. Furthermore, our study will assess the impact of choosing the best model, based on AIC or BIC, on the sensitivity of excess of loss premium estimates when the true model is included or excluded from the model space. The study will compare the performance of the model average estimators with those obtained based on the best model selected using either AIC or BIC, in both an application and an extensive simulation study.

The remainder of the paper is structured as follows. Section 2 explores the methodology employed to develop estimators for the model average excess of loss premium, specifically when dealing with left-truncated loss data. Model weights based on AIC and BIC are presented, as well as the closed-form formulas for estimating premiums given a retention level. Section 3 demonstrates the application of the proposed methodology to actual reinsurance loss data. Section 4 documents a simulation study that evaluates the proposed method across various experimental settings. Finally, Section 5 presents a discussion and concluding remarks.

2. METHODOLOGY

2.1. BACKGROUND

Let X be the total loss random variable for an insurance company that belongs to the set $\mathcal{B} = \mathcal{B}(I; \mu, \sigma)$ of all non-negative random variables with μ , standard deviation σ ,

and support contained in the interval $I = [0, b]$; here $b = +\infty$ is allowed. Under the stop-loss reinsurance agreement, the primary insurance (or ceding) company retains losses below the retention level r and transfers the excess (or ceded) losses to the reinsurer. Let X_I denote the loss amount retained by the insurer, and let X_R denote the loss amount ceded to the reinsurer as part of the stop-loss reinsurance contract. Then, the relationship between X , X_I , and X_R is expressed as follows:

$$X = X_I + X_R = X \wedge r + (X - r)_+$$

where

$$\begin{aligned} X_I &= X \wedge r = \min\{X, r\} \\ &= \begin{cases} X, & X \leq r \\ r, & X > r \end{cases} \end{aligned} \quad (2.1)$$

and

$$\begin{aligned} X_R &= (X - r)_+ = \max\{(X - r), 0\} \\ &= \begin{cases} 0, & X \leq r \\ X - r, & X > r \end{cases} \end{aligned} \quad (2.2)$$

where $0 \leq r \leq b$ is known as the retention level. In return for assuming the risk, the reinsurer charges the reinsurance premium to the cedent. A number of principles have been proposed for determining the appropriate level of premium. One commonly used is the expected value principle, in which the reinsurance premium $\delta(r)$ is determined as $\delta(r) = (1 + \rho)\pi(r)$, where $\rho > 0$ is known as the relative sensitive safety loading. ρ represents a fixed, constant value used to account for the additional premium charged above the expected losses to ensure financial safety (such as for covering unexpected or extreme losses). Typically, ρ and r are independent of each other because ρ is a fixed parameter that does not change with changes in r . The net stop-loss (or excess of loss) premium at a given retention level r is defined as

$$\begin{aligned} \pi(r) &= E[X_R] = E[(X - r)_+] \\ &= \int_r^\infty (X - r)f(x; \theta)dx, \end{aligned} \quad (2.3)$$

where $f(x; \theta)$ is the probability density function (PDF) for modeling the insurance losses. Both $\delta(r)$ and $\pi(r)$ decrease as r increases. The total cost for the insurer in the presence of the stop-loss reinsurance captures the retained loss and the reinsurance premium, while the reinsurer is responsible for paying the claim amount in excess of a given retention.

Klugman, Panjer, and Willmot (2012) pointed out that the net stop-loss premium is typically calculated using the aggregate claim distribution that requires modeling both frequency and severity of losses. In the actuarial literature, various models have been proposed for modeling excess of loss reinsurance by fitting the severity of losses; see Beirlant et al. (2006) and Verbelen et al. (2015), among others. As previously mentioned, estimates of $\pi(r)$ obtained by solving Equation 2.3 are model-specific, and different models often yield differing estimates. One approach for selecting the most suitable model may involve criteria such as AIC, BIC, or other methods chosen by the researcher. While the best-fitting model aims to closely match observed data with fitted values, a risk manager may prefer alternative model selection techniques. When faced with competing models that adequately fit the data, one may also consider models that produce the lowest or highest estimated pre-

miums. To address model uncertainty comprehensively, our objective is to integrate all models within a small family of finite mixture models into the premium estimation process.

2.2. FITTING LOSSES USING FINITE MIXTURE MODELS

Due to the heterogeneous nature of loss data, finite mixture models offer a flexible framework for capturing complex distributions that may not be well represented by a single parametric family. Parametric mixture models, particularly when estimated using the expectation-maximization (EM) algorithm introduced by Dempster, Laird, and Rubin (1977), have been shown to be effective for modeling insurance severity data with varying policy characteristics (e.g., left truncation, right censoring). Several studies have explored their application, including but not limited to Verbelen et al. (2015), Miljkovic and Grün (2016), Blostein and Miljkovic (2019), Michael, Miljkovic, and Melnykov (2020), and Bae and Miljkovic (2024). In this subsection, we provide a background on finite mixture models as a general approach to loss modeling, as in this paper they are considered the foundation for computing the net stop-loss premium in Equation 2.3.

Suppose X denotes a continuous random variable representing the loss amount, modeled by the PDF of a finite mixture model. This model combines probability distributions and has been applied to fit insurance loss data by Miljkovic and Grün (2016), Blostein and Miljkovic (2019), and Michael, Miljkovic, and Melnykov (2020), among others. The PDF of the K -component finite mixture is expressed as

$$g(\mathbf{x}; \Theta) = \sum_{j=1}^K \omega_j f_j(\mathbf{x}; \theta_j), \quad (2.4)$$

where $f_j(\mathbf{x}; \theta_j)$ represents the j th mixture component with parameters θ_j , and ω_j is the j th mixture weights subject to $0 \leq \omega_j \leq 1$, for $j \in \{1, 2, \dots, K\}$, and $\sum_{j=1}^K \omega_j = 1$. The set of all parameters, denoted by Θ , includes mixing coefficients and parameters of each component distribution: $\Theta = \{\omega_1, \omega_2, \dots, \omega_{K-1}, \theta_1, \theta_2, \dots, \theta_K\}'$. Similarly, the cumulative distribution function (CDF) of this finite mixture is expressed as

$$G(\mathbf{x}; \Theta) = \sum_{j=1}^K \omega_j F_j(\mathbf{x}; \theta_j), \quad (2.5)$$

where $F_j(\mathbf{x}; \theta_j)$ denotes the CDF of the j th component distribution with parameters θ_j .

Typically, both the number of components K and the parameter vector Θ are unknown and estimated using the EM algorithm.

The EM algorithm introduces a new random variable, Z_{ij} , as a missing component representing the identity of the i th observation in the j th component. Z_{ij} will take a value of 1 if the i th observation belongs to the j th component and 0 otherwise. This leads to a complete-data likelihood function

$$L_c(\Theta) = \prod_{i=1}^n \prod_{j=1}^K [\omega_j f_j(x_i; \theta_j)]^{Z_{ij}}. \quad (2.6)$$

The EM algorithm consists of two steps: E -step and M -step. At the E -step, a conditional expectation of the complete-data log-likelihood function $\log L_c(\Theta)$ is obtained, commonly called Q -function. This involves finding the expectation of the latent variable Z_{ij} , which simplifies to finding posterior probabilities ω_{ij} representing the probability of the i th observation belonging to the j th component. On the s th iteration of the EM algorithm, these posterior probabilities are computed as $\omega_{ij}^{(s)} = \frac{\omega_j^{(s-1)} f_j(x_i; \theta_j^{(s-1)})}{\sum_{j=1}^K \omega_j^{(s-1)} f_j(x_i; \theta_j^{(s-1)})}$. The M -step consists of maximizing the Q -function, given as

$$Q(\Theta | \Theta^{(s-1)}) = \sum_{i=1}^n \sum_{j=1}^K \omega_{ij} [\log \omega_j + \log f_j(x_i; \theta_j)], \quad (2.7)$$

which is maximized with respect to parameter vector Θ . At the s th iteration, the estimates of the mixing proportions are updated by using

$$\omega_j^{(s)} = \frac{1}{n} \sum_{i=1}^n \omega_{ij}^{(s)}. \quad (2.8)$$

Remaining parameters in Θ require finding distribution-specific parameters estimates, θ_j , by solving a weighted maximum likelihood estimation problem for each component distribution. The E -step and M -step alternate until the relative increase in the log-likelihood value, at the parameter estimates obtained from consecutive iterations, is no bigger than a small, prespecified tolerance level. To determine the optimal number of components, maximum likelihood estimates and the corresponding log-likelihood values are obtained for each fixed number of components, K , and the best model is selected based on either AIC or BIC. These criteria balance fitness (trying to maximize the likelihood function) and parsimony (using penalties associated with measures of model complexity), trying to avoid overfit. Furthermore, fitting a model with a large number of components requires estimating a very large number of parameters, with a consequent loss of precision in these estimates. Several other information criteria are available in the literature as modified versions of AIC or BIC (Fonseca and Cardoso 2007).

In the actuarial literature, there is no consensus regarding the preferred criterion—AIC or BIC—for use with mixture models. Unlike BIC, AIC imposes a lesser penalty on models with a higher number of parameters, often leading to the selection of models with more mixture components (Steele and Raftery 2010). Research by Roeder and Wasserman (1997) demonstrated that BIC yields a consistent estimator of the mixture density, while Keribin (2000) showed its consistency in choosing the number of components in a mixture model. BIC has been adopted for mixture modeling by Roeder and Wasserman (1997) and has since been widely utilized, particularly in clustering applications (Fraley and Raftery 1998), with promising practical outcomes. Due to its more substantial penalty, BIC typically favors simpler models with fewer parameters.

2.3. FITTING LEFT-TRUNCATED LOSSES USING FINITE MIXTURE MODELS

In this section, we extend the methodology from the previous section to accommodate left-truncated loss data. We define c as the left-truncation point, which represents the threshold below which losses are not observed. This truncation is typically due to policy conditions. The left-truncation c is separate from r (retention level), previously defined in Section 2.1 as part of the excess of loss reinsurance, which is the threshold above which claims are covered by the reinsurer. These two thresholds serve different purposes: c affects data availability and r determines financial risk allocation.

Suppose we are interested in the conditional probability that a single loss exceeds c , the truncation point, given that losses below c have not been observed. The left-truncated PDF, denoted as $g^T(\cdot)$, is derived using conditional probability as follows:

$$\begin{aligned} g^T(x; c, \boldsymbol{\theta}) &= \lim_{d \rightarrow \infty} \frac{P(x < X \leq x+d | X > c)}{d} \\ &= \lim_{d \rightarrow \infty} \frac{P(x < X \leq x+d, X > c)}{dP(X > c)} \\ &= \begin{cases} 0/P(X > c), & x < c \\ \lim_{d \rightarrow \infty} \frac{P(x < X \leq x+d)}{dP(X > c)}, & x \geq c \end{cases} \quad (2.9) \\ &= \begin{cases} 0, & x < c \\ \frac{g(x; \boldsymbol{\theta})}{1-F(c; \boldsymbol{\theta})}, & x \geq c \end{cases} \\ &= \frac{g(x; \boldsymbol{\theta})}{1-F(c; \boldsymbol{\theta})} \mathbf{1}(x \geq c), \end{aligned}$$

where $g(x, \boldsymbol{\theta})$ represents the PDF without truncation, defined for $x > 0$; $F(c, \boldsymbol{\theta})$ is the corresponding CDF evaluated at the truncation point c ; and $\mathbf{1}(\cdot)$ denotes an indicator function. For all other values $x \leq c$, the value of $g^T(x; c, \boldsymbol{\theta})$ is equal to zero. To simplify the notation for the PDF of a left-truncated random variable, the indicator function part of Equation 2.9 will be dropped in the subsequent equations. If we consider fitting the left-truncated losses using a finite mixture model, the left-truncated mixture can be expressed as

$$g^T(x; c, \boldsymbol{\Theta}) = \sum_{j=1}^K \omega_j \frac{g_j(x, \boldsymbol{\theta}_j)}{1 - F_j(c; \boldsymbol{\theta}_j)}, \quad (2.10)$$

where $\boldsymbol{\Theta} = \{\omega_1, \omega_2, \dots, \omega_{K-1}, \boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2, \dots, \boldsymbol{\theta}'_K\}'$. Note that when extending the framework of mixture models to left-truncated data, the left-truncated PDF, $g^T(x; c, \boldsymbol{\Theta})$, replaces the general PDF from Equation 2.4. While this substitution alters the PDF functions used in the E -step and M -step, the overall structure of the EM algorithm remains unchanged. The modifications ensure that parameter estimates properly account for the truncation while preserving the iterative nature and convergence properties of the EM algorithm.

Fitting left-truncated losses through a finite mixture of distributions has been explored in prior research by Verbeulen et al. (2015), Blostein and Miljkovic (2019), and many others. To facilitate this modeling approach, the R package `ltmix`, developed by Blostein and Miljkovic (2021), is designed to accommodate combinations of mixture compo-

nent distributions such as gamma, lognormal, and Weibull distributions.

2.4. PROPOSED METHOD FOR THE EXCESS OF LOSS PREMIUM CALCULATION

Suppose that the left-truncated insurance losses are modeled using Equation 2.10. Now, we consider the computation of the excess of loss premium, which is based on the different retention level, r . Unlike c , which defines the lower boundary of the observed losses (i.e., $[c, \infty]$), r represents the point beyond which the reinsurer assumes responsibility for claims. The excess of loss premium is computed as

$$\begin{aligned} \pi(r) &= E[(X - r)_+ | X > c] \\ &= \int_r^\infty (x - r) g^T(x, c, \boldsymbol{\theta}) dx. \end{aligned} \quad (2.11)$$

This formula assumes that the parametric form $g^T(x, c, \boldsymbol{\theta})$ is known and represents the most suitable model for the available data within the subset of models considered in the model space. The selection of this model typically relies on a model selection criterion, such as AIC or BIC. The best model is chosen based on the minimal value of AIC or BIC among all candidate models in the considered model space. However, it is common for no single model to be clearly superior to others in the model set. Thus, this approach may introduce a model selection bias. If estimated values of certain quantities of interest, such as excess of loss premiums, vary significantly across models, relying solely on the selected best model can be risky. A more prudent approach is to consider basing computations on a weighted estimate of premiums, utilizing some form of MA. This involves assigning model weights to estimated premiums based on the corresponding models in the model space, thereby quantifying model uncertainty and incorporating it into the final premium estimation. Here, we propose an approach to computing the excess of loss premium based on MA.

Consider a model space of $\mathcal{M} = \{\mathcal{M}_1, \dots, \mathcal{M}_M\}$, consisting of M possible mixture models for fitting left-truncated insurance losses. Following the framework outlined in Burnham and Anderson (2003), the probability of selecting model \mathcal{M}_i (where $i = 1, \dots, M$) given the data \boldsymbol{x} is expressed as

$$P(\mathcal{M}_i | \boldsymbol{x}) = \frac{\exp(-\frac{1}{2} \Delta_i)}{\sum_{i=1}^M \exp(-\frac{1}{2} \Delta_i)} = \alpha_i(\boldsymbol{x}), \quad (2.12)$$

where Δ_i represents the difference in information criterion (either AIC or BIC) between model \mathcal{M}_i and the best-fitting model in the set. Specifically, Δ_i corresponds to the AIC or BIC difference and is defined as

$$\Delta_i = \text{AIC}_i - \min(\text{AIC}_1, \dots, \text{AIC}_M), \quad (2.13)$$

or

$$\Delta_i = \text{BIC}_i - \min(\text{BIC}_1, \dots, \text{BIC}_M), \quad (2.14)$$

where AIC_i and BIC_i are the AIC and BIC values, respectively, for model \mathcal{M}_i , and $\min(\text{AIC}_1, \dots, \text{AIC}_M)$ and $\min(\text{BIC}_1, \dots, \text{BIC}_M)$ are the minimum AIC and BIC values, respectively, across all mixture models in the model space \mathcal{M} . The exponential term $\exp(-\frac{1}{2} \Delta_i)$ is proportional to the likelihood of model \mathcal{M}_i given the data, denoted as

$\mathcal{L}(\mathcal{M}_i|\mathbf{x}) \propto \exp(-\frac{1}{2}\Delta_i)$. It reflects the relative strength of evidence for model \mathcal{M}_i based on the data. By normalizing the likelihoods of all models, we obtain a set of positive AIC or BIC weights, $\alpha_i(\mathbf{x})$, such that $\alpha_i(\mathbf{x}) > 0$ and $\sum_{i=1}^M \alpha_i(\mathbf{x}) = 1$. These weights represent the posterior probabilities of the models, with larger Δ_i values leading to smaller probabilities for model \mathcal{M}_i , given the observed data. The weights based on BIC offer an additional interpretation as posterior model probabilities, indicating the probability that a model is the true model given the data. In contrast, AIC-based weights lack such an interpretation. The use of AIC-based weights is supported by the assumption that the true model is unknown and therefore outside of the considered model class (see Buckland, Burnham, and Augustin 1997). The justification for using AIC or BIC weights includes their relative simplicity in calculation, which make them appealing for practitioners.

The AIC- and BIC-based weights naturally incorporate model uncertainty by assigning higher weights to models with better fit (lower AIC or BIC) and lower weights to models that are less plausible. This is crucial in contexts where no single model is clearly superior. Rather than selecting a single best model, this MA approach provides a way to incorporate the uncertainty across multiple models, leading to more robust estimates. However, both criteria assume that the set of models being compared is adequate and that none of the models is fundamentally incorrect. If all models are misspecified, neither AIC nor BIC will perform well in selecting a truly good model and the model-averaged estimates will also be flawed, potentially leading to poor inferences and predictions. In the subsequent discussion, we will use the terms MA-AIC and MA-BIC to refer to the MA weights derived from the AIC and BIC criteria, respectively.

Suppose each model i in the model space \mathcal{M} produces the excess of loss premium for a given retention level r . Then, the model average excess of loss premium can be obtained by averaging premiums over the individual models as follows:

$$\bar{\pi}(r) = \sum_{i=1}^M \alpha_i(\mathbf{x}) \pi_i(r), \quad (2.15)$$

where $\pi_i(r)$ is the estimated premium based on the model i with the weight $\alpha_i(\mathbf{x})$ obtained using either AIC or BIC approach. According to Madigan and Raftery (1994), the estimator of the form $\bar{\pi}(r)$ has better predictive ability than any estimate of the form $\pi_i(r)$ from the i th candidate model. By combining Equation 2.10 and Equation 2.15, we derive the model-averaged premium as follows:

$$\bar{\pi}(r) = \sum_{i=1}^M \alpha_i(\mathbf{x}) \left[\sum_{j=1}^K \frac{\omega_{ij}}{1 - F_{ij}(c, \boldsymbol{\theta}_{ij})} \int_r^\infty (x-r) g_{ij}(x; c, \boldsymbol{\theta}_{ij}) dx \right], \quad (2.16)$$

where ω_{ij} represents the mixture weight of the i th mixture with j components, $\alpha_i(\mathbf{x})$ denotes the model weight (AIC or BIC) of the i th model in the considered space of finite mixture models, and $\boldsymbol{\theta}_{ij}$ denotes the parameters of the j component in the i th mixture. For finite mixture based on the lognormal or gamma components, considered in this paper, closed-form theoretical results for $\bar{\pi}(r)$ are developed as a result of the following lemmas.

LOGNORMAL CASE

Lemma 1. Suppose that a random variable X follows a left-truncated lognormal distribution with parameters μ and σ and the truncation point c . Then the excess of loss premium at retention level r may be expressed as

$$\pi(r) = \frac{1}{\Phi\left(\frac{\mu - \ln(c)}{\sigma}\right)} \left[e^{\mu + \frac{\sigma^2}{2}} \Phi\left(\frac{\mu + \sigma^2 - \ln(r)}{\sigma}\right) - r \Phi\left(\frac{\mu - \ln(r)}{\sigma}\right) \right].$$

The proof is provided in Appendix A.

Using Lemma 1, we can obtain the model average excess of loss premium by applying the weights over each component of the K -component lognormal mixture as follows:

$$\bar{\pi}(r) = \sum_{i=1}^M \alpha_i \sum_{j=1}^K \frac{\omega_{ij}}{\Phi\left(\frac{\mu_{ij} - \ln(c)}{\sigma_{ij}}\right)} \left[e^{\mu_{ij} + \frac{\sigma_{ij}^2}{2}} \Phi\left(\frac{\mu_{ij} + \sigma_{ij}^2 - \ln(r)}{\sigma_{ij}}\right) - r \Phi\left(\frac{\mu_{ij} - \ln(r)}{\sigma_{ij}}\right) \right]. \quad (2.17)$$

GAMMA CASE

Lemma 2. Suppose that a random variable X follows a left-truncated gamma distribution with the shape parameter α and the scale parameter β and the truncation point c . Then the excess of loss premium at retention level r may be expressed as

$$\pi(r) = \frac{1}{1 - G(c; \alpha, \beta)} \left[\alpha \beta [1 - G(r; \alpha + 1, \beta)] - r [1 - G(r; \alpha, \beta)] \right],$$

where $G(c; \alpha, \beta)$ is the CDF of the gamma random variable.

The proof is provided in Appendix B.

Using Lemma 2, we can obtain the model average excess of loss premium by applying the weights over each component of the K -component gamma mixture as follows:

$$\bar{\pi}(r) = \sum_{i=1}^M \alpha_i \sum_{j=1}^K \frac{\omega_{ij}}{1 - G(c; \alpha_{ij}, \beta_{ij})} \left[\alpha_{ij} \beta_{ij} [1 - G(r; \alpha_{ij} + 1, \beta_{ij})] - r [1 - G(r; \alpha_{ij}, \beta_{ij})] \right], \quad (2.18)$$

where $G(c; \alpha_{ij}, \beta_{ij})$ is the CDF of the gamma random variable associated with the j th component of the i th mixture.

Although this paper focuses on mixtures of lognormal and gamma distributions, two of the most widely cited models for claim severity due to their analytical tractability, these distributions have also been effective in fitting the left-truncated Secura Re dataset (compatible with the R package **Itmix**), which is central to our application. While other distribution families could certainly be considered, they may require numerical integration techniques when closed-form solutions are unavailable. In such cases, an R package like **pracma** or **stats** can facilitate the computation.

3. APPLICATION

In this section, we illustrate the MA approach to the Secura Re dataset, which consists of left-truncated automobile claims. This dataset is accessible through the R package **Itmix**, developed by Blostein and Miljkovic (2021). Secura Re losses include 371 claims recorded by the Belgian reinsurance company for the period 1988 to 2001 in 2002 euros. Claims below 1.2 million euros have not been reported to the reinsurer (see Figure 1). Table 1 presents basic summary statistics for Secura Re. The first five basic statistics in this table are reported in millions. The statistics reveal that the mean (2.609) exceeds the median (2.231), and the skewness coefficient of 2.422 indicates a moderate-to-strong positive skewness in the data.

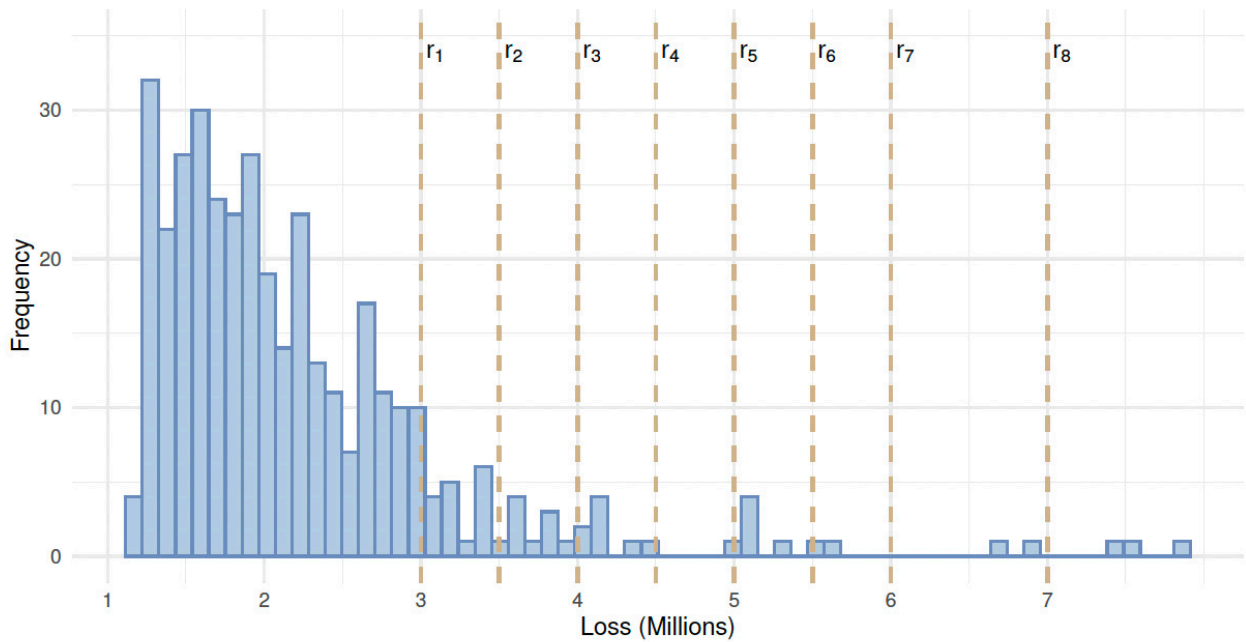


Figure 1. Histogram of Secura Re reinsurance losses with the dashed vertical lines representing eight selected retention levels $r \in \{3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0\}$.

Table 1. Summary statistics for Secura Re reinsurance losses. The first six numbers are reported in millions.

Summary Statistics						
Minimum	1st quartile	Median	Mean	3rd quartile	Maximum	Skewness
1.208	1.573	1.944	2.231	2.609	7.899	2.422

The Secura Re dataset was used in several actuarial applications related to the estimation of the excess of loss premium, risk measures, and confidence intervals for risk measures; see Verbelen et al. (2015), Reynkens et al. (2017), Blostein and Miljkovic (2019), and Grün and Miljkovic (2023), among others. We employ Secura Re to illustrate the proposed MA approach in the calculation of excess of loss reinsurance premiums at various retention levels. Our interest lies in globally fitting a wide range of possible claim outcomes for Secura Re, as opposed to focusing solely on fitting the tail of the distribution (Beirlant, Matthys, and Dierckx 2001). The objective of this paper is to leverage a comprehensive set of fitting models within a defined model space to address model selection uncertainty.

For the analysis of Secura Re data, a model space is defined consisting of six finite mixture models with a fixed number of components. These models include a single-component lognormal (L), a single-component gamma (G), a two-component lognormal (LL), a two-component gamma (GG), a three-component lognormal (LLL), and a three-component gamma (GGG). For each model considered within this model space, the excess of loss premium is calculated at eight retentions $r \in \{3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0\}$, expressed in millions of euros. These retention levels are chosen based on their consistency with the literature on modeling the same dataset (see Verbelen et al.

2015) and to ensure a uniform distribution across the range of given losses. Figure 1 shows the histogram of Secura Re losses, with dashed vertical lines indicating the location of the eight retention levels relative to the data. The MA-AIC and MA-BIC are computed, and the summary of the results is presented in Table 2.

It is important to note that the selection of models and the calculation of AIC or BIC weights are based on the likelihood and penalty of complexity, which are independent of the specific retention levels used for premium calculation. Fitting distributions to the data before selecting retention levels ensures that the models used for premium calculations are well suited to the data. This is particularly relevant for Secura Re data, where the same models have been previously explored in the literature (see Blostein and Miljkovic 2019). Understanding the fitted distributions aids in choosing retention levels that are realistic and manageable from a risk management perspective, rather than relying on arbitrary or less-informed choices.

The choice of lognormal and gamma mixture models in our analysis is guided by their superior fit to the Secura Re dataset, demonstrated in prior research by Blostein and Miljkovic (2019) and Verbelen et al. (2015). While other distribution families may be considered, the empirical evidence from previous studies and our preliminary exploratory analysis confirmed that mixtures of lognormals

and gammas are sufficient to illustrate our proposed methodology on Secura Re data. In general, these mixtures effectively capture the heavy-tailed nature and multimodality observed in the insurance datasets, making them well suited for excess of loss premium calculations.

The lower portion of [Table 2](#) shows the relative errors in the estimated premium when compared to the empirical results. Four different methods are employed for premium estimation: (i) selecting the best model using the AIC approach (BEST-AIC), (ii) selecting the best model using the BIC approach (BEST-BIC), (iii) using the MA-BIC approach, and (iv) using the MA-AIC approach.

The relative error of each estimation approach compared to its corresponding empirical value is reported, and the results are discussed as follows. The empirical results serve as the benchmark, consistent with previous literature (see Verbelen et al. 2015; Reynkens et al. 2017). However, an alternative benchmark could be selected, such as the best model based on the specified model selection criterion. The AIC and BIC approaches lead to the selection of two different best models. While the BIC approach favors a single lognormal (L), the AIC approach prefers a GG mixture. These results align with previous literature findings, where Blostein and Miljkovic (2019) reported that the lognormal distribution is the best fit for Secura Re based on the BIC criterion, while Verbelen et al. (2015) reported that a two-component Erlang mixture provided the best fit for the same dataset when the AIC criterion was employed in the model selection. It is worth noting that the two-component Erlang mixture can be viewed as a special case of a GG mixture with both components having different shape parameters while sharing a common scale parameter.

The BIC criterion exhibits a preference for a simpler model with fewer parameters, imposing a larger penalty on models with more parameters. As a result, the top model based on BIC weights is the single L model, carrying approximately 96% weight, and the second best is a single-component gamma (G), with about 4% weight. In contrast, the AIC criterion tends to distribute model weights across multiple models. The four models with significant AIC weights are the GG model (47%), the L model (34%), the LL model (13%), and the G model (2%).

Based on the relative error results, the MA-BIC approach demonstrates a slight underestimation of premiums relative to the nonparametric results for all retention levels. Conversely, the MA-AIC approach slightly overestimates the premiums relative to the nonparametric results for the same retention levels. Larger relative errors in premium results are observed as the retention level increases. This is expected because the fit of the tail varies across different models, with some models better suited for the main body of the data and others for the tail. All methods exhibit considerable variability at $r = 7.0M$, the retention level close to the end of the data range (i.e., maximum observation is 7.899 million).

On average, the MA-AIC premium results closely approximate the nonparametric results across all retention levels except at $r = 3.0M$. The MA-BIC premium results are closer to the nonparametric results for the retention levels

$r = 3.0M$, $r = 3.5M$, and $r = 7.0M$ compared to those generated by the MA-AIC approach. However, the MA-AIC premium results are closer to nonparametric results for the retention levels $r = 4.0M$, $r = 4.5M$, $r = 5.0M$, $r = 5.5M$, and $r = 6.0M$ compared to those generated by the MA-BIC approach.

For the sake of further investigation, we will assume that the best model selected by the BIC approach (L) is excluded from the model space. Thus, the model space under consideration includes only five models: G, GG, GGG, LL, and LLL. The question of interest is: which of the model average estimates will yield results that are closer to the empirical results? [Table 3](#) provides a summary of the results for the five models fitted to Secura Re, along with the model average estimates based on BIC and AIC weights. [Table 3](#) displays the relative errors for the estimated premium when compared to the nonparametric results. Once again, four different methods are employed for premium estimation: (i) BEST-BIC, (ii) BEST-AIC, (iii) MA-BIC, and (iv) MA-AIC approaches. The following discussion delves into the results.

When excluding the L model from the model space of the six models, the G model is selected as the best model by the BIC criterion, while the AIC criterion designates the GG model as the best model. Again, we arrive at two distinct types of models under consideration. Interestingly, the result of the model selection based on the AIC criterion remains the same as when the six original models were used in the model space. The results of the relative errors indicate that the difference between the MA-AIC approach and the nonparametric approach is somewhat the smallest for most retention levels except the first two. Once again, it is observed that the estimates for the highest retention level of 7.0 million have a large relative error and may not be reliable, irrespective of the method employed.

The relative errors based on both the MA-BIC and BEST-BIC methods indicate an underestimation of the premium relative to the empirical results across all retention levels. Conversely, the MA-AIC and BEST-AIC methods tend to overestimate the premium relative to the empirical results.

In summary, it is crucial to emphasize that the MA approach for Secura Re data is demonstrated within a specific class of models, namely a limited space of finite mixture models. This class includes the best-fitting model identified in previous research. If a different class of models, such as composite models, is to be explored, the same MA approach can be easily applied. In fact, several attempts have been made to fit Secura Re data using composite model classes or splicing models (e.g., Reynkens et al. 2017; Grün and Miljkovic 2019). Further, the model space of mixture models can be enlarged to include length- or size-biased mixtures to fit left-truncated data, such as those proposed by Bae and Ko (2020) and Bae and Miljkovic (2024). However, none of the previous research related to alternative model classes has incorporated model uncertainty risk in its applications.

Table 2. Summary of excess of loss premiums for Secura Re at different retention levels $r = \{3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0\}$. Model weights ω_{bic} and ω_{aic} (based on BIC and AIC) guide six mixture models: G, GG, GGG, L, LL, and LLL for premium calculation. MA-BIC and MA-AIC represent model average premiums with BIC and AIC weights respectively. The relative errors are computed for (i) BEST-BIC model vs. Empirical premium, (ii) BEST-AIC model vs. Empirical premium, (iii) MA-BIC vs. Empirical premium, and (iv) MA-AIC vs. Empirical premium.

Summary of the Excess of Loss Premium for Various Retention Levels (6 Models)										
Model	ω_{bic}	ω_{aic}	3.0M	3.5M	4.0M	4.5M	5.0M	5.5M	6.0M	7.0M
Empirical			161,728	108,837	74,696	53,312	35,888	24,420	17,211	4,785
G	0.0387	0.0136	156,257	89,552	50,866	28,680	16,071	8,958	4,971	1,513
GG	0.0038	0.4730	193,325	133,531	95,656	68,639	47,949	32,033	20,279	6,842
GGG	0.0000	0.0222	194,564	130,224	91,256	64,599	44,618	29,488	18,512	6,243
L	0.9564	0.3355	151,716	89,533	53,570	32,541	20,074	12,572	7,990	3,363
LL	0.0011	0.1335	178,258	117,249	79,443	54,952	38,575	27,384	19,620	10,313
LLL	0.0000	0.0222	183,636	123,580	87,057	61,482	42,350	28,179	18,108	6,829
MA-BIC			152,078	89,730	53,653	32,552	20,044	12,522	7,932	3,312
MA-AIC			176,663	115,704	78,475	53,909	36,714	24,428	15,773	6,052
(i) Relative Error: BEST-BIC Model (L) and Empirical Premium Estimates										
			-0.0619	-0.1774	-0.2828	-0.3896	-0.4407	-0.4852	-0.5357	-0.2971
(ii) Relative Error: BEST-AIC Model (L) and Empirical Premium Estimates										
			0.1954	0.2269	0.2806	0.2875	0.3361	0.3118	0.1783	0.4299
(iii) Relative Error: MA-BIC and Empirical Premium Estimates										
			-0.0597	-0.1756	-0.2817	-0.3894	-0.4415	-0.4872	-0.5391	-0.3078
(iv) Relative Error: MA-AIC and Empirical Premium Estimates										
			0.0923	0.0631	0.0506	0.0112	0.023	-0.0003	-0.0836	0.2649

Table 3. Summary of the excess of loss premiums for Secura Re at various retention levels $r = \{3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0\}$ when model L is excluded from the model space. Model weights ω_{bic} and ω_{aic} (based on BIC and AIC) guide five mixture models: G, GG, GGG, LL, and LLL for premium calculation. MA-BIC and MA-AIC represent model average premiums with BIC and AIC weights. The relative errors are computed for (i) BEST-BIC model vs. Empirical premium, (ii) BEST-AIC model vs. Empirical premium, (iii) MA-BIC vs. Empirical premium, and (iv) MA-AIC vs. Empirical premium.

Summary of the Excess of Loss Premium When Model L Is Excluded from the Model Space (5 Models)										
Model	ω_{bic}	ω_{aic}	3.0M	3.5M	4.0M	4.5M	5.0M	5.5M	6.0M	7.0M
Empirical			161,728	108,837	74,696	53,312	35,888	24,420	17,211	4,785
G	0.8875	0.0188	156,159	89,477	50,812	28,642	16,046	8,942	4,960	1,509
GG	0.0785	0.5917	195,018	134,070	94,737	66,879	46,268	31,059	20,151	7,653
GGG	0.0000	0.1088	189,428	130,018	90,015	61,749	41,705	27,694	18,091	7,386
LL	0.0339	0.2558	185,896	124,747	86,025	60,142	42,274	29,751	20,930	10,342
LLL	0.0000	0.0249	179,941	120,208	83,312	57,884	39,789	27,018	18,162	8,047
MA-BIC			160,221	94,177	55,458	32,715	19,310	11,385	6,695	2,292
MA-AIC			190,970	130,061	90,884	63,654	44,020	29,842	19,791	8,206
(i) Relative Error: BEST-BIC Model (G) and Empirical Premium Estimates										
			-0.0344	-0.1779	-0.3197	-0.4627	-0.5529	-0.6338	-0.7118	-0.6846
(ii) Relative Error: BEST-AIC Model (GG) and Empirical Premium Estimates										
			0.2058	0.2318	0.2683	0.2545	0.2892	0.2719	0.1708	0.5994
(iii) Relative Error: MA-BIC and Empirical Premium Estimates										
			-0.0093	-0.1347	-0.2576	-0.3863	-0.4619	-0.5338	-0.6110	-0.5211
(iv) Relative Error: MA-AIC and Empirical Premium Estimates										
			0.1808	0.1950	0.2167	0.1940	0.2266	0.2220	0.1499	0.7150

4. SIMULATION

In this section, a Monte Carlo (MC) simulation study is performed to comprehensively evaluate the performance of the proposed MA approach and to quantify model uncertainty under varying settings and data generating mechanisms similar to those observed in Secura Re. The data generating process is based on the two-component Erlang mixture: a specific form of the GG mixture with shape parameters $\alpha = (5, 16)$, a common scale parameter $\theta = 360096$, and mixing weights $\omega = (0.97, 0.03)$, reported by Verbelen et al. (2015) to be the best-fitting model. Through this simulation study, the excess of loss premium is computed under four estimation methods for various settings that consider two sample sizes, two truncation points, seven retention levels, and the two simulation scenarios. These simulation settings are summarized as follows.

- **Estimation methods:** The four estimation methods considered are BEST-AIC, BEST-BIC, MA-AIC, and MA-BIC. The former two methods employ AIC and BIC as model selection criteria, while the latter two methods are grounded in the proposed MA.
- **Sample size:** The sample size is denoted by n and takes values in the set $n = \{500, 1,000\}$.
- **Retention levels:** The retention level spans values in the set $R = \{3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0\}$, expressed in millions.
- **Simulation scenarios:** Two model spaces are considered: \mathcal{M}_1 includes six models $\mathcal{M}_1 = \{G, GG, GGG, L, LL, LLL\}$, and \mathcal{M}_2 encompasses five models $\mathcal{M}_2 = \{G, GGG, L, LL, LLL\}$, excluding the GG model. Note that the mixture models with a fixed number of components are considered, similar to those used in modeling Secura Re data.
- **Truncation points:** The truncation point, denoted by c , takes values in the set $c = \{0.8, 1.2\}$, expressed in millions.
- **Repetitions:** The number of MC runs is set to 1,000, ensuring robust statistical assessment.

To evaluate the four estimation methods, the relative error (RE) is used as the performance measure. For the i th level of sample size, the r th retention level, the t th truncation point, the k th method of estimation, and the l th repetition in the MC runs, the RE_{irtkl} is calculated using the formula

$$RE_{irtkl} = \frac{\hat{\pi}_{irtkl} - \pi_{rt}}{\pi_{rt}}, \quad (4.1)$$

where $\hat{\pi}_{irtkl}$ denotes simulated premium value, while π_{rt} is the premium based on the true-data generating process. For each combination of simulation settings, a distribution of the RE of the premium is generated relative to the true excess of loss premium based on 1,000 simulation runs. The true excess of loss premium is computed based on the assumed data generating process.

Figure 2 presents box plots illustrating the distributions of RE under various simulation settings, maintaining a fixed truncation point at 1.2. The organization of box plots into four rows corresponds to different combinations of

sample size and model space. The first row represents a sample size of $n = 500$ with model space \mathcal{M}_1 , the second row denotes $n = 1,000$ with \mathcal{M}_1 , the third row pertains to $n = 500$ with \mathcal{M}_2 , and the fourth row aligns with $n = 1,000$ with \mathcal{M}_2 . Each row includes box plots for four methods: MA-BIC (soft lavender), MA-AIC (light coral), BEST-BIC (pale green), and BEST-AIC (pale blue). The pale vertical zero-line serves as the reference point, where—ideally, when the estimated premium matches the true premium—RE is equal to zero. Two distinct groups of box plots emerge based on the method: MA-AIC and BEST-AIC exhibit similarities, while MA-BIC and BEST-BIC also share similarities, although these two groups differ in their locations. In many instances, the vertical zero-line is close to one side of the box plot representing MA-AIC and BEST-AIC methods, while it does not intersect MA-BIC and BEST-BIC box plots. Variability is evident with an increase in retention levels and the emergence of outliers. Similarly, a panel plot of box plots in Figure 3 is constructed for the distribution of RE under different simulation settings, with the truncation point fixed at 0.8. Here it is observed that the vertical zero-line intersects both MA-AIC and BEST-AIC box plots for all retention levels, in contrast to MA-BIC and BEST-BIC box plots, which are notably shifted from the vertical reference line.

The magnitude of RE shows a decrease when the truncation point is lowered from 1.2 to 0.8, corresponding to a reduction in the left tail area of the distribution. Notably, when the true model is encompassed within the model space and the sample size is small ($n = 500$), the MA-AIC method yields slightly superior results compared to the BEST-AIC approach. However, with a larger sample size ($n = 1,000$), MA-AIC outperforms BEST-AIC at certain retention levels, particularly in the right tail. It is observed that the variability in the distribution of REs increases with a higher retention level. Notably, smaller variability is noted for retention level 3.0M, while the largest variability is reported for 7.0M (tail area).

The tabulated values for the average RE, denoted as \overline{RE} , are summarized in Table 6 in Appendix C. The standard deviations of the RE for each distribution across the combination of settings are presented in Table 7 (Appendix C). The standard deviation increases with an increase in retention level across all other simulation settings and serves as an indicator of the uncertainty associated with the premium estimates. Specifically, the standard deviation of the RE distribution can be used to quantify the uncertainty in premium estimation at each retention level. It is also important to note that the MA-AIC and MA-BIC methods exhibit lower variability, on average, across all simulation settings, compared to their counterparts (BEST-AIC and BEST-BIC).

Table 4 illustrates the frequency of the correct model being selected over 1,000 MC runs when the true model is included in the model space. The AIC appears to select the correct model (GG) most of the time, whereas the BIC tends to favor the simpler model (L) regardless of the truncation point and the sample size.

When the true model is excluded from the model space, the frequency of the best model selected is summarized in

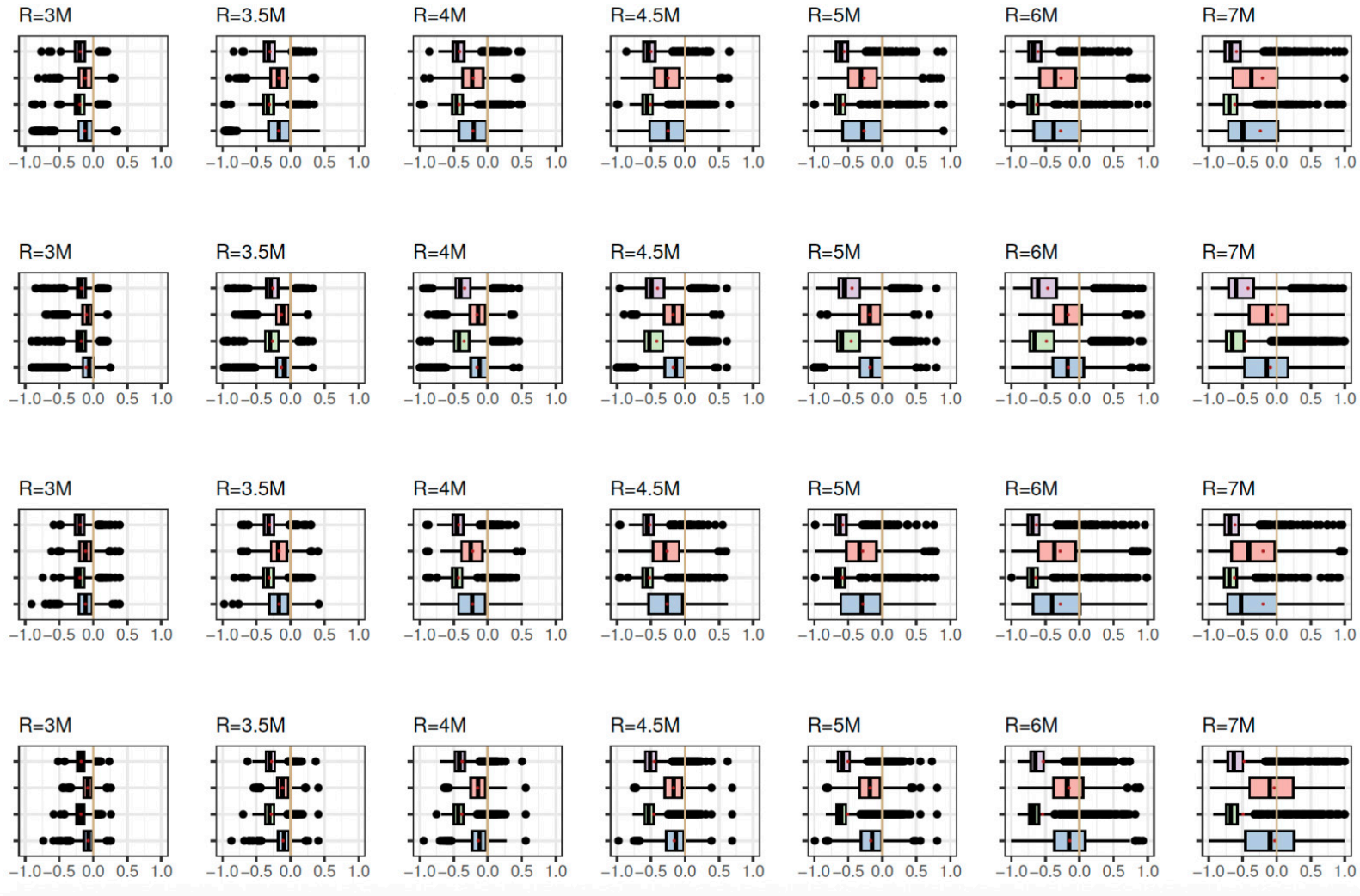


Figure 2. Panel plot illustrates the summary of relative errors for $c = 1.2$. The plot is organized in four rows, each representing a combination of sample size and model space. The first row corresponds to a sample size of $n = 500$ with \mathcal{M}_1 , the second row to $n = 1,000$ with \mathcal{M}_1 , the third row to $n = 500$ with \mathcal{M}_2 , and the fourth row to $n = 1,000$ with \mathcal{M}_2 . Each row contains box plots for four methods: MA-BIC (soft lavender), MA-AIC (light coral), BEST-BIC (pale green), and BEST-AIC (pale blue).

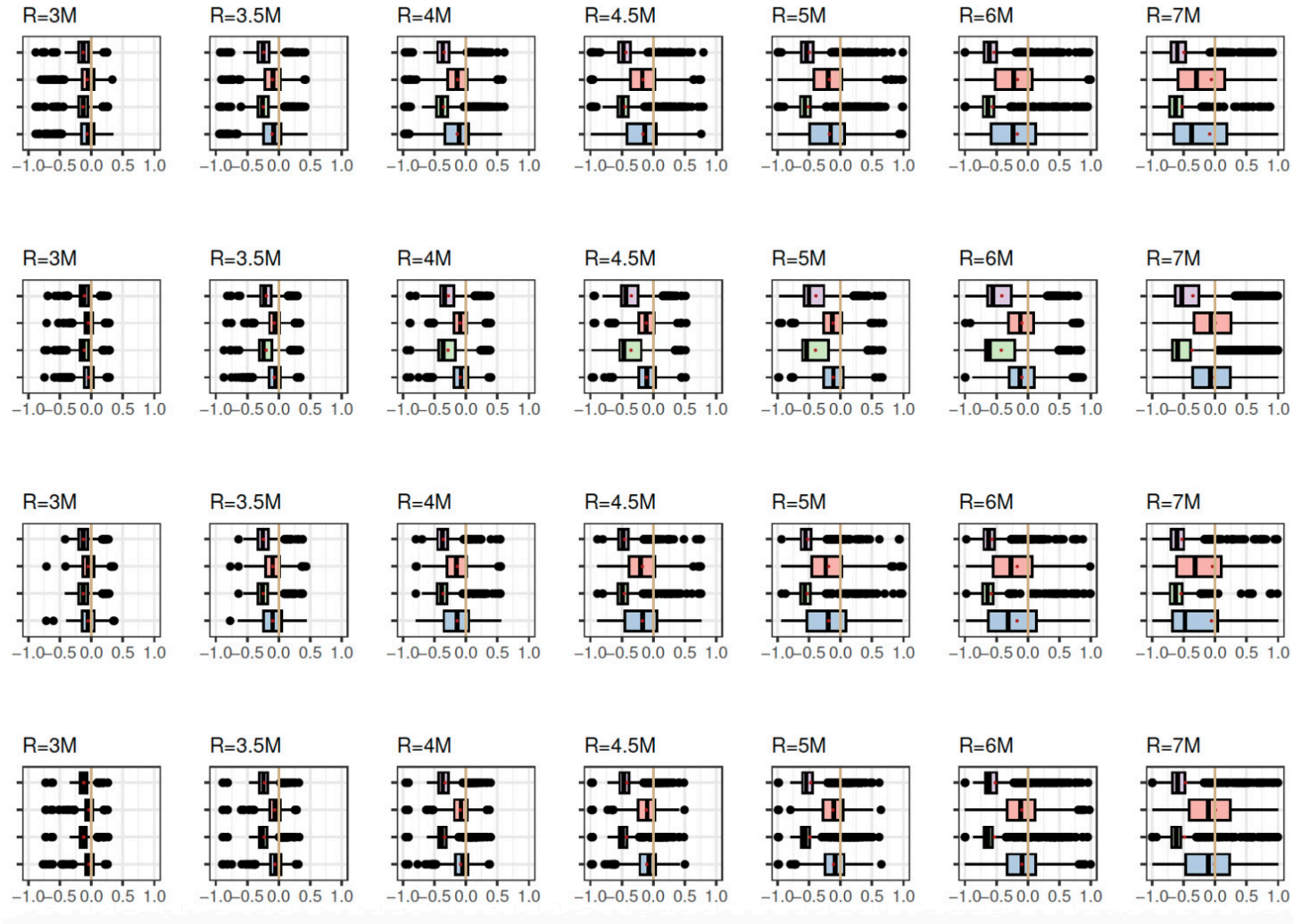


Figure 3. Panel plot illustrates the summary of relative errors for $c = 0.8$. The plot is organized in four rows, each representing a combination of sample size and model space. The first row corresponds to a sample size of $n = 500$ with \mathcal{M}_1 , the second row to $n = 1,000$ with \mathcal{M}_1 , the third row to $n = 500$ with \mathcal{M}_2 , and the fourth row to $n = 1,000$ with \mathcal{M}_2 . Each row contains box plots for four methods: MA-BIC (soft lavender), MA-AIC (light coral), BEST-BIC (pale green), and BEST-AIC (pale blue).

Table 4. Frequency table showing the number of times the correct model selected over six models.

c	n	G	GG	GGG	L	LL	LLL
AIC: Model Space \mathcal{M}_1							
0.8	500	1	444	99	221	142	93
	1,000	0	593	117	46	151	93
1.2	500	0	363	91	261	187	98
	1,000	0	554	98	47	196	105
BIC: Model Space \mathcal{M}_1							
0.8	500	1	61	0	897	35	6
	1,000	0	227	0	689	79	5
1.2	500	1	39	0	907	45	8
	1,000	0	195	1	695	103	6

Table 5. Frequency table showing the number of times the correct model selected over five models.

c	n	G	GGG	L	LL	LLL
AIC: Model Space \mathcal{M}_2						
0.8	500	1	199	335	306	159
	1,000	0	342	102	288	268
1.2	500	1	145	328	388	138
	1,000	0	238	90	474	198
BIC: Model Space \mathcal{M}_2						
0.8	500	1	1	937	55	6
	1,000	0	1	854	138	7
1.2	500	1	2	919	65	13
	1,000	0	1	809	188	2

Table 5 for both AIC and BIC model selection criteria. The BIC consistently selects a single-component lognormal as the best model regardless of the truncation point and the sample size. The AIC tends to select several candidate models, among which a single lognormal is frequently chosen, particularly when the truncation point is 0.8 and 1.2 with a sample size of 500. A single-component gamma is never selected by either AIC or BIC. The AIC tends to select the GGG model more frequently than the BIC approach, even though the true model GG, which is excluded from \mathcal{M}_2 , consists of the same component distributions.

Figure 4 shows the panel plot consisting of multiple line plots of \overline{RE} under various simulation settings. The organization of line plots into four rows corresponds to different combinations of sample size, model space, and truncation point. The first column represents a sample size of $n = 500$, while the second column represents sample size $n = 1,000$. The first two rows correspond to truncation point 0.8, while the last two rows show the results for truncation point 1.2. Sample space is labeled as “in” for \mathcal{M}_1 and “out” for \mathcal{M}_2 . There are four line styles denoting one of the four estimation methods: BEST-AIC (solid), BEST-BIC (dashed), MA-AIC (dotted), and MA-BIC (dot-dashed). This panel plot reveals a common trend among all estimation methods: none of them yields an \overline{RE} greater than zero. The results validate

the analysis of real data where the best-selected models seem to slightly underestimate the empirical results. MA-AIC and BEST-AIC show significantly smaller \overline{RE} s across all retention levels compared to the BEST-BIC and MA-BIC methods. Additionally, from Figure 4 we observe that an increase in sample size seems to decrease \overline{RE} . Lowering the left-truncation point has an impact of reducing the \overline{RE} . This is expected, since lowering the truncation point would allow the inclusion of more observations that were previously excluded due to being below the original truncation threshold. This inclusion of additional data points can provide more information about the lower tail of the distribution as well as the overall distribution of the data, leading to improved accuracy of the premium estimates based on this distribution.

Figure 5 displays the standard deviation of RE values across retention levels, organized similarly to Figure 4. Overall, the standard deviation of RE tends to increase with higher retention levels—especially noticeable with smaller sample sizes ($n = 500$). For these, the variability in RE is notably lower for BEST-BIC (dashed line) and MA-BIC (dot-dashed line) than for BEST-AIC (solid line) and MA-AIC (dotted line). MA-BIC shows slightly lower variability than BEST-BIC. For larger sample sizes ($n = 1,000$), there is less variability in RE across methods. Considering \mathcal{M}_1 , MA-AIC

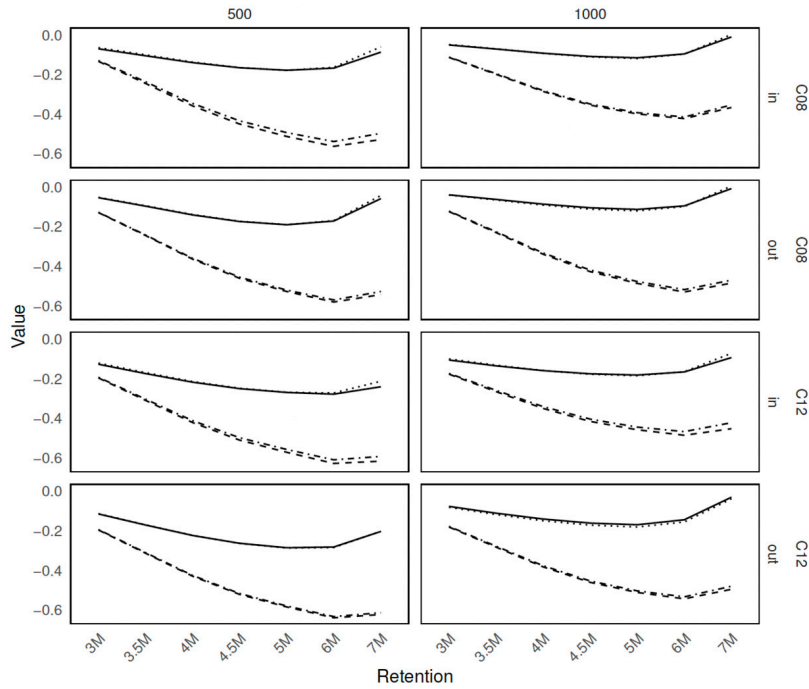


Figure 4. Line plots of \overline{RE} organized for all combinations of the simulation settings as follows: sample size (in columns): 500, 1,000; method of estimation: BEST-AIC (solid), MA-AIC (dotted), BEST-BIC (dashed), and MA-BIC (dotdashed); truncation point denoted as C08 and C12 for 0.8 and 1.2, respectively (across two rows); two scenarios labeled as “in” corresponding to \mathcal{M}_1 and “out” corresponding to \mathcal{M}_2 (by row).

shows the lowest standard deviation across most retention levels, while for \mathcal{M}_2 , MA-BIC demonstrates the smallest variability in RE.

To assess the performance of the proposed approach with a small sample size, additional simulation runs ($N = 1,000$) were conducted with $n = 50$ and a truncation point of $c = 1.2$. The data generating process remained consistent with the original setup, based on a two-component GG mixture with shape parameters $\alpha = (5, 16)$, a common scale parameter $\theta = 360096$, and mixing weights $\omega = (0.97, 0.03)$. Two scenarios were considered: “in” corresponding to \mathcal{M}_1 and “out” corresponding to \mathcal{M}_2 . The results are summarized in Appendix C. [Table 8](#) presents the mean of the RE distribution across different settings, while [Table 9](#) reports the standard deviation. Compared to the results obtained with moderate and large sample sizes, we observe substantial variability across all retention levels, regardless of the premium estimation approach. [Figure 6](#) displays box plots organized into two rows corresponding to the two model spaces. The variability increases with higher retention levels, accompanied by a growing number of outliers. For retention levels in the tail of the distribution, the variability becomes so large that the results may be considered unreliable. Moreover, small sample sizes introduce additional uncertainty in parameter estimation, a well-known phenomenon in statistical inference, further compounding the challenges of working with limited data.

It is important to note that a simulation study can be used to quantify model uncertainty by analyzing the standard deviation of the RE distribution associated with the premium estimates. However, directly quantifying model

uncertainty with real data is more challenging, as the true values are not available for comparison. In such cases, one potential approach is to simulate data under conditions similar to those in the real data and apply the same methods used in the simulation study, as demonstrated in this section of the paper. This approach allows for assessing how the model behaves under various scenarios and better quantifying the uncertainty associated with the MA process in a real-world context.

5. CONCLUSION AND DISCUSSION

The field of loss modeling is extensive, encompassing a wide array of model classes for modeling insurance losses. However, there has been relatively little research on MA to address the uncertainty inherent in selecting the best-fitted models, which can significantly impact critical risk management decisions.

This paper introduces an approach for addressing uncertainty in model selection and improving the accuracy of estimated excess of loss premiums. The proposed method involves MA for estimating the excess of loss premium when finite mixture models, utilizing gamma and lognormal component distributions, are considered. Furthermore, we have derived closed-form formulas for computing the model average premium using AIC and BIC weights, which are based on the gamma and lognormal mixtures, fitted to left-truncated insurance losses.

When estimating the excess of loss premium using Secura Re data, we found that the model average estimators based on either AIC or BIC weights perform slightly better

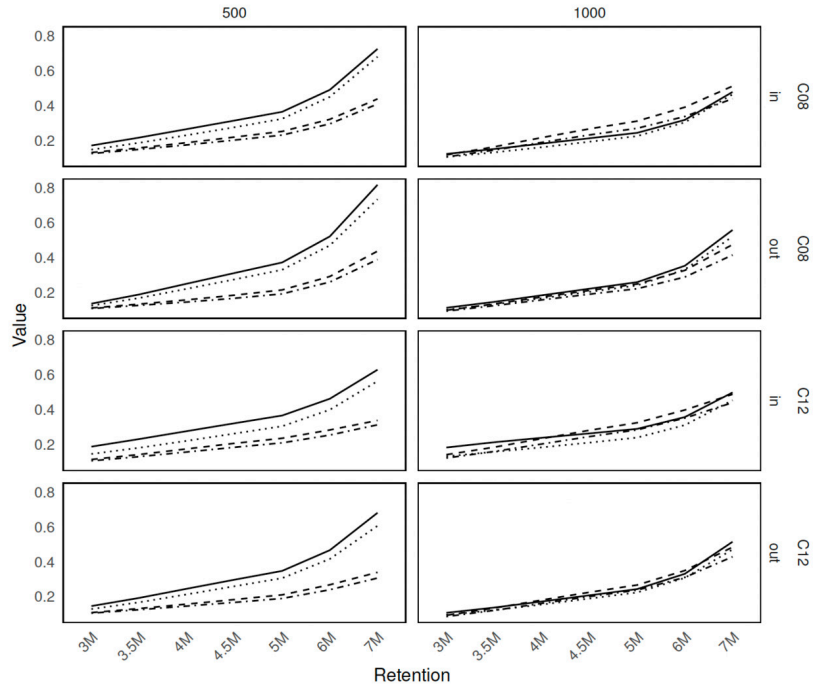


Figure 5. Line plots show the standard deviation of RE values organised for all combinations of the simulation settings as follows: sample size (in columns): 500, 1,000; method of estimation: BESTAIC (solid), MA-AIC (dotted), BEST-BIC (dashed), and MA-BIC (dotdashed); truncation point denoted as C08 and C12 for 0.8 and 1.2, respectively (across two rows); two scenarios labeled as “in” corresponding to \mathcal{M}_1 and “out” corresponding to \mathcal{M}_2 (by row).

than individual models selected solely based on AIC or BIC criteria. Most importantly, model average estimators account for model uncertainty, an important factor when multiple models are involved in analyzing loss data.

Through extensive simulations, we discovered that the MA-BIC method exhibits the least variability across all retention levels when the sample size is small. On average, both MA-AIC and AIC tend to approximate the true model more closely across all simulation settings, with MA-AIC showing slight improvements in the results. Our simulation study validates findings observed in modeling real datasets, as the shape of the simulated data closely resembles that of Secura Re data.

The reader is advised that while this study focuses on lognormal and gamma mixture models due to their strong empirical fit to the Secura Re dataset, other distribution families can also be considered based on the specific characteristics of the real data. It is important to assess the distributional fit of the data before selecting the model, as different datasets may exhibit different behaviors that could be better captured by alternative distributions. As such, we encourage practitioners to explore a wide range of distribution families when analyzing their own datasets, adjusting their model selection based on the specific features of the data at hand. Based on the considered distribution families, the excess of loss premium can be computed using the proposed MA approach to account for model uncertainty and improve the reliability of the premium estimates, offering a more robust solution compared to relying on a single model.

Other methods for averaging models can be explored; for instance, for small n a corrected version of AIC (AIC_C) exists that provides stronger penalty for small sample sizes and stronger than BIC for very small sample sizes. AIC_C is reported by Burnham and Anderson (2004) to have better small-sample behavior. Further developments are being made, and new forms of information criteria have been proposed. The Watanabe-Akaike (or widely applicable) information criterion (WAIC; see Watanabe and Opper 2010) is an attempt from Bayesian statistics to emulate a cross-validation analysis. While WAIC is preferred to the deviance information criterion, a direct leave-one-out criterion is now recommended (see Spiegelhalter et al. 2002 and Vehtari, Gelman, and Gabry 2016, among others). These developments highlight the ongoing efforts to refine and improve MA techniques.

There is significant potential to extend the methodology proposed in this paper to encompass premium calculations for other types of reinsurance contracts. For example, in surplus share reinsurance, which covers a percentage of each loss up to a certain limit, one can utilize the MA approach. By averaging the loss ratios from various models using AIC or BIC weights, the reinsurance premium can be determined while accounting for model uncertainty. Another potential application is in parametric (or index-based) insurance, which pays out based on predefined trigger events, such as rainfall levels or earthquake magnitudes. The premium can be estimated using model-averaged predictions of the trigger event’s probability. In this context, AIC or BIC weights allow for averaging the

predictions of the event's occurrence to determine the premium. Additionally, catastrophe bonds (cat bonds) are used by insurers to transfer risk to the capital markets. The premium for these bonds can be influenced by models estimating the frequency and severity of catastrophic events. The expected loss rate used in premium calculations can be derived from model-averaged estimates of catastrophe risk, with AIC or BIC weights applied to average predictions of catastrophe loss from various models. The methodology in this paper can also be extended to a layer reinsurance contract, where the reinsurer covers losses within a specific range, between a lower retention point and an upper retention point. The premium for such a contract can be derived by integrating over the range of losses that the reinsurer is liable for.

One avenue for future work would be to explore the use of the FIC approach in calculation of the excess of loss premium. Future research could explore modifications to FIC to enhance its applicability to premium estimation, including the potential development of a smoothed or regularized approach that could impose constraints to ensure consistency across different retention levels. Further investigation into the asymptotic properties of FIC in the context of excess loss pricing, particularly under different loss distribution assumptions, could also provide valuable insights. Exploring these directions would help refine the use of FIC in actuarial applications.

Finally, the proposed framework is based on the expected premium principle, but it can be carefully extended to other premium principles, such as the variance premium principle (which adds a variance term to adjust the premium for risk variability) and the standard deviation premium principle (which considers the standard deviation of the loss distribution). Challenges may arise when combining the mean and variance (or standard deviation), particularly in terms of computational complexity. Additionally, careful consideration should be given to the model weights, as ideally they would reflect not only the goodness of fit for the expected loss but also the variability captured by the model (i.e., the variance and standard deviation).

.....

ACKNOWLEDGMENT

The author sincerely appreciates the thoughtful comments and suggestions from the editor and two anonymous reviewers, which have helped enhance the quality of this manuscript. The author is also grateful to Dr. Nikita Barabanov and Dr. Seonjin Kim for their insightful feedback on an earlier version of this work.

Submitted: September 06, 2024 EDT. Accepted: May 25, 2025 EDT. Published: July 30, 2025 EDT.

REFERENCES

- Akaike, H. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* 19 (6): 716–23. <https://doi.org/10.1109/TAC.1974.1100705>.
- Bae, T., and B. Ko. 2020. "On the Mixtures of Length-Biased Weibull Distributions for Loss Severity Modeling." *Journal of the Korean Statistical Society* 49:422–38. <https://doi.org/10.1007/s42952-019-00021-z>.
- Bae, T., and T. Miljkovic. 2024. "Loss Modeling with the Size-Biased Lognormal Mixture and the Entropy Regularized EM Algorithm." *Insurance: Mathematics and Economics* 117:182–95. <https://doi.org/10.1016/j.insmatheco.2024.05.003>.
- Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels. 2006. *Statistics of Extremes: Theory and Applications*. Wiley.
- Beirlant, J., G. Matthys, and G. Dierckx. 2001. "Heavy-Tailed Distributions and Rating." *ASTIN Bulletin* 31 (1): 37–58. <https://doi.org/10.2143/AST.31.1.993>.
- Blostein, M., and T. Miljkovic. 2019. "On Modeling Left-Truncated Loss Data Using Mixtures of Distributions." *Insurance: Mathematics and Economics* 85:35–46. <https://doi.org/10.1016/j.insmatheco.2018.12.001>.
- . 2021. Package "Ltmix: Left-Truncated Mixtures of Gamma, Weibull, and Lognormal Distributions." *R Package Version 0.2.1*. <https://CRAN.R-project.org/package=ltmix>.
- Buckland, S. T., K. P. Burnham, and N. H. Augustin. 1997. "Model Selection: An Integral Part of Inference." *Biometrics* 53 (2): 603–18. <https://doi.org/10.2307/2533961>.
- Burnham, K. P., and D. R. Anderson. 2003. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 11th Edition*. Springer.
- . 2004. "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods & Research* 33 (2): 261–304. <https://doi.org/10.1177/0049124104268644>.
- Cairns, A. J. 2000. "A Discussion of Parameter and Model Uncertainty in Insurance." *Insurance: Mathematics and Economics* 27 (3): 313–30. [https://doi.org/10.1016/S0167-6687\(00\)00055-X](https://doi.org/10.1016/S0167-6687(00)00055-X).
- De Vylder, F. E. 1996. *Advanced Risk Theory: A Self-Contained Introduction*. Éditions de l'Université de Bruxelles.
- Deelstra, G., and G. Plantin. 2014. *Risk Theory and Reinsurance*. Springer. <https://doi.org/10.1007/978-1-4471-5568-3>.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM-Algorithm." *Journal of the Royal Statistical Society, Series B* 39 (1): 1–38. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- Fonseca, J. R., and M. G. Cardoso. 2007. "Mixture-Model Cluster Analysis Using Information Theoretical Criteria." *Intelligent Data Analysis* 11 (2): 155–73. <https://doi.org/10.3233/IDA-2007-11204>.
- Fraley, C., and A. E. Raftery. 1998. "How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis." *Computer Journal* 41 (8): 578–88. <https://doi.org/10.1093/comjnl/41.8.578>.
- Grün, B., and T. Miljkovic. 2019. "Extending Composite Loss Models Using a General Framework of Advanced Computational Tools." *Scandinavian Actuarial Journal* 2019 (8): 642–60. <https://doi.org/10.1080/03461238.2019.1596151>.
- . 2023. "The Automated Bias-Corrected and Accelerated Bootstrap Confidence Intervals for Risk Measures." *North American Actuarial Journal* 27 (4): 731–50. <https://doi.org/10.1080/10920277.2022.2141781>.
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science* 14 (4): 382–417.
- Jullum, M., and N. L. Hjort. 2017. "Parametric or Nonparametric: The FIC Approach." *Statistica Sinica* 27 (3): 951–81.
- Keribin, C. 2000. "Consistent Estimation of the Order of Mixture Models." *Sankhyā: The Indian Journal of Statistics, Series A* 62 (1): 49–66.
- Klugman, S. A., H. H. Panjer, and G. E. Willmot. 2012. *Loss Models: From Data to Decisions*. 4th ed. Wiley. <https://doi.org/10.1002/9781118787106>.

- Madigan, D., and A. E. Raftery. 1994. "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window." *Journal of the American Statistical Association* 89 (428): 1535–46. <https://doi.org/10.1080/01621459.1994.10476894>.
- Michael, S., T. Miljkovic, and V. Melnykov. 2020. "Mixture Modeling of Data with Multiple Partial Right-Censoring Levels." *Advances in Data Analysis and Classification* 14 (2): 355–78. <https://doi.org/10.1007/s11634-020-00391-x>.
- Miljkovic, T., and B. Grün. 2016. "Modeling Loss Data Using Mixtures of Distributions." *Insurance: Mathematics and Economics* 70:387–96. <https://doi.org/10.1016/j.insmatheco.2016.06.019>.
- . 2021. "Using Model Averaging to Determine Suitable Risk Measure Estimates." *North American Actuarial Journal* 25 (4): 562–79. <https://doi.org/10.1080/10920277.2021.1911668>.
- Raftery, A. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111–63. <https://doi.org/10.2307/271063>.
- Reynkens, T., R. Verbelen, J. Beirlant, and K. Antonio. 2017. "Modelling Censored Losses Using Splicing: A Global Fit Strategy with Mixed Erlang and Extreme Value Distributions." *Insurance: Mathematics and Economics* 77:65–77. <https://doi.org/10.1016/j.insmatheco.2017.08.005>.
- Roeder, K., and L. Wasserman. 1997. "Practical Bayesian Density Estimation Using Mixtures of Normals." *Journal of the American Statistical Association* 92 (439): 894–902. <https://doi.org/10.1080/01621459.1997.10474044>.
- Schwarz, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6 (2): 461–64. <https://doi.org/10.1214/aos/1176344136>.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society, Series B* 64 (4): 583–639. <https://doi.org/10.1111/1467-9868.00353>.
- Steele, R. J., and A. E. Raftery. 2010. "Performance of Bayesian Model Selection Criteria for Gaussian Mixture Models." *Frontiers of Statistical Decision Making and Bayesian Analysis* 2:113–30.
- Vehtari, A., A. Gelman, and J. Gabry. 2016. "Efficient Implementation of Leave-One-Out Cross-Validation and WAIC for Evaluating Fitted Bayesian Models." Preprint, arXiv, arXiv:1507.04544.
- Verbelen, R., L. Gong, K. Antonio, A. Badescu, and S. Lin. 2015. "Fitting Mixtures of Erlangs to Censored and Truncated Data Using the EM Algorithm." *ASTIN Bulletin* 45 (3): 729–58. <https://doi.org/10.1017/asb.2015.15>.
- Wang, Y., and I. Hobæk. 2019. "Focussed Selection of the Claim Severity Distribution." *Scandinavian Actuarial Journal* 2019 (2): 129–42. <https://doi.org/10.1080/03461238.2018.1519847>.
- Watanabe, S., and M. Opper. 2010. "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory." *Journal of Machine Learning Research* 11:3571–94.

APPENDIX

A. PROOF OF LEMMA 1

The excess of loss premium for the retention level r in the case where X follows a left-truncated lognormal distribution with parameters μ and σ and a truncation point c is defined as

$$\begin{aligned} \pi(r) &= E((X - r)_+ | X > c) \\ &= \int_r^\infty (x - r) \frac{f(x, \mu, \sigma)}{1 - \Phi\left(\frac{\ln(c) - \mu}{\sigma}\right)} dx, \end{aligned} \quad (\text{A.1})$$

where $f(x, \mu, \sigma) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2}}$ is the lognormal PDF, and $\Phi\left(\frac{\ln(c) - \mu}{\sigma}\right)$ is the lognormal CDF evaluated at c , with $\Phi(\cdot)$ denoting the CDF of the standard normal distribution.

Using the linearity property of integration, we can split the integral into two parts and pull out the constant term, resulting in

$$\begin{aligned} \pi(r) &= \frac{1}{1 - \Phi\left(\frac{\ln(c) - \mu}{\sigma}\right)} \left[\int_r^\infty \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2}} dx - r \int_r^\infty \frac{1}{x\sqrt{2\pi}} e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2}} dx \right]. \end{aligned} \quad (\text{A.2})$$

The first integral in Equation A.2 can be solved using the substitution $z = \frac{\ln(x) - \mu}{\sigma}$, leading to the following solution:

$$e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\ln(r) - \mu}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z - \sigma)^2} dz. \quad (\text{A.3})$$

After another substitution $t = z - \sigma$, we obtain

$$\begin{aligned} e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\ln(r) - \mu - \sigma^2}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \\ = e^{\mu + \frac{\sigma^2}{2}} \left[1 - \Phi\left(\frac{\ln(r) - \mu - \sigma^2}{\sigma}\right) \right]. \end{aligned} \quad (\text{A.4})$$

To solve the second integral in Equation A.2, we apply the same substitution $z = \frac{\ln(x) - \mu}{\sigma}$, leading to

$$\int_{\frac{\ln(r) - \mu}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = 1 - \Phi\left(\frac{\ln(r) - \mu}{\sigma}\right). \quad (\text{A.5})$$

Considering the symmetry of the standard normal distribution, the final solution is

$$\pi(r) = \frac{1}{\Phi\left(\frac{\mu - \ln(c)}{\sigma}\right)} \left[e^{\mu + \frac{\sigma^2}{2}} \Phi\left(\frac{\mu + \sigma^2 - \ln(r)}{\sigma}\right) - r \Phi\left(\frac{\mu - \ln(r)}{\sigma}\right) \right].$$

B. PROOF OF LEMMA 2

The excess of loss premium for the retention level r in the case where X follows a left-truncated gamma distribution with shape parameter α , scale parameter β , and a truncation point c is defined as

$$\begin{aligned} \pi(r) &= E((X - r)_+ | X > c) \\ &= \int_r^\infty (x - r) \frac{f(x; \alpha, \beta)}{1 - G(c; \alpha, \beta)} dx, \end{aligned} \quad (\text{B.1})$$

where $f(x; \alpha, \beta) = \frac{1}{x\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^\alpha e^{-\frac{x}{\beta}}$ is the gamma PDF, and $G(c; \alpha, \beta)$ is the gamma CDF evaluated at c . That is, $G(c; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \frac{c}{\beta})$. Note, the gamma CDF is represented in terms of the gamma function $\Gamma(\alpha)$ and lower incomplete gamma function $\gamma(\alpha, \frac{x}{\beta})$; that is, $G(x; \alpha, \beta) = \frac{\gamma(\alpha, \frac{x}{\beta})}{\Gamma(\alpha)}$. The incomplete gamma function is defined as $\gamma(\alpha, \frac{x}{\beta}) = \int_0^{\frac{x}{\beta}} t^{\alpha-1} e^{-t} dt$.

Using the linearity property of integration, we can split the integral in Equation B.1 into two integrals and simplify it further as

$$\begin{aligned} \pi(r) &= \frac{1}{1 - G(c; \alpha, \beta)} \left[\int_r^\infty \frac{1}{\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^\alpha e^{-\frac{x}{\beta}} dx - r \int_r^\infty \frac{1}{x\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^\alpha e^{-\frac{x}{\beta}} dx \right]. \end{aligned} \quad (\text{B.2})$$

Using the property of the gamma distribution, the first integral above can be expressed with CDF of gamma. That is, $\int_r^\infty \frac{1}{\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^\alpha e^{-\frac{x}{\beta}} dx = \alpha\beta \int_r^\infty \frac{1}{x\Gamma(\alpha+1)} \left(\frac{x}{\beta}\right)^{\alpha+1} e^{-\frac{x}{\beta}} dx = \alpha\beta[1 - G(r; \alpha + 1, \beta)]$. The second integral above is equivalent to $1 - G(r; \alpha, \beta)$. Thus, the final solution of Equation B.1 is expressed as follows:

$$\begin{aligned} \pi(r) &= \frac{1}{1 - G(c; \alpha, \beta)} \left[\alpha\beta[1 - G(r; \alpha + 1, \beta)] - r[1 - G(r; \alpha, \beta)] \right]. \end{aligned} \quad (\text{B.3})$$

C. SIMULATION TABLES AND FIGURES

Table 6. Summary of the mean of REs for each combination of the simulation settings

c	n	Method	3M	3.5M	4M	4.5M	5M	6M	7M
Scenario 1: Model Space \mathcal{M}_1									
1.2	500	BEST-AIC	-0.127	-0.174	-0.217	-0.250	-0.269	-0.278	-0.240
		BEST-BIC	-0.196	-0.309	-0.421	-0.510	-0.571	-0.628	-0.616
		MA-AIC	-0.121	-0.169	-0.214	-0.248	-0.268	-0.272	-0.212
		MA-BIC	-0.192	-0.303	-0.412	-0.498	-0.557	-0.609	-0.592
	1,000	BEST-AIC	-0.105	-0.135	-0.159	-0.175	-0.181	-0.165	-0.093
		BEST-BIC	-0.177	-0.265	-0.350	-0.415	-0.458	-0.485	-0.452
		MA-AIC	-0.099	-0.131	-0.158	-0.177	-0.185	-0.164	-0.071
		MA-BIC	-0.173	-0.258	-0.341	-0.404	-0.444	-0.467	-0.422
Scenario 2: Model Space \mathcal{M}_2									
1.2	500	BEST-AIC	-0.115	-0.171	-0.223	-0.263	-0.285	-0.282	-0.203
		BEST-BIC	-0.196	-0.313	-0.429	-0.521	-0.585	-0.639	-0.623
		MA-AIC	-0.113	-0.170	-0.223	-0.263	-0.287	-0.284	-0.202
		MA-BIC	-0.193	-0.310	-0.425	-0.517	-0.580	-0.633	-0.614
	1,000	BEST-AIC	-0.076	-0.111	-0.140	-0.161	-0.169	-0.144	-0.030
		BEST-BIC	-0.181	-0.282	-0.382	-0.460	-0.511	-0.544	-0.495
		MA-AIC	-0.081	-0.118	-0.149	-0.171	-0.181	-0.155	-0.038
		MA-BIC	-0.178	-0.278	-0.376	-0.453	-0.503	-0.533	-0.480
Scenario 1: Model Space \mathcal{M}_1									
0.8	500	BEST-AIC	-0.069	-0.105	-0.139	-0.165	-0.178	-0.167	-0.086
		BEST-BIC	-0.133	-0.244	-0.358	-0.450	-0.512	-0.562	-0.528
		MA-AIC	-0.064	-0.099	-0.137	-0.164	-0.178	-0.163	-0.060
		MA-BIC	-0.129	-0.236	-0.346	-0.434	-0.493	-0.538	-0.497
	1,000	BEST-AIC	-0.050	-0.070	-0.091	-0.108	-0.115	-0.095	-0.009
		BEST-BIC	-0.113	-0.198	-0.285	-0.354	-0.398	-0.422	-0.366
		MA-AIC	-0.048	-0.069	-0.092	-0.110	-0.118	-0.096	0.002
		MA-BIC	-0.112	-0.195	-0.281	-0.349	-0.392	-0.414	-0.353
Scenario 2: Model Space \mathcal{M}_2									
0.8	500	BEST-AIC	-0.053	-0.096	-0.141	-0.174	-0.190	-0.171	-0.058
		BEST-BIC	-0.129	-0.244	-0.364	-0.461	-0.527	-0.581	-0.543
		MA-AIC	-0.052	-0.093	-0.138	-0.173	-0.190	-0.169	-0.042
		MA-BIC	-0.127	-0.242	-0.360	-0.455	-0.520	-0.570	-0.528
	1,000	BEST-AIC	-0.039	-0.062	-0.085	-0.104	-0.113	-0.094	-0.007
		BEST-BIC	-0.124	-0.230	-0.339	-0.427	-0.487	-0.530	-0.485
		MA-AIC	-0.040	-0.065	-0.091	-0.111	-0.120	-0.097	0.005
		MA-BIC	-0.121	-0.225	-0.333	-0.419	-0.477	-0.518	-0.470

Table 7. Summary of standard deviations REs for each combination of the simulation settings

c	n	Method	3M	3.5M	4M	4.5M	5M	6M	7M
Scenario 1: Model Space \mathcal{M}_1									
1.2	500	BEST-AIC	0.188	0.232	0.277	0.322	0.367	0.463	0.629
		BEST-BIC	0.115	0.143	0.175	0.207	0.236	0.284	0.338
		MA-AIC	0.146	0.182	0.221	0.263	0.305	0.400	0.564
		MA-BIC	0.107	0.131	0.158	0.185	0.210	0.255	0.313
	1,000	BEST-AIC	0.183	0.213	0.239	0.264	0.290	0.358	0.499
		BEST-BIC	0.142	0.185	0.234	0.282	0.325	0.399	0.489
		MA-AIC	0.133	0.159	0.184	0.211	0.240	0.312	0.454
		MA-BIC	0.123	0.161	0.204	0.246	0.285	0.352	0.441
Scenario 2: Model Space \mathcal{M}_2									
1.2	500	BEST-AIC	0.146	0.193	0.245	0.297	0.348	0.466	0.681
		BEST-BIC	0.109	0.132	0.157	0.183	0.210	0.268	0.339
		MA-AIC	0.129	0.168	0.213	0.260	0.306	0.416	0.606
		MA-BIC	0.105	0.125	0.146	0.167	0.190	0.240	0.307
	1,000	BEST-AIC	0.107	0.138	0.172	0.208	0.243	0.331	0.515
		BEST-BIC	0.093	0.134	0.181	0.225	0.267	0.350	0.484
		MA-AIC	0.095	0.122	0.155	0.189	0.224	0.308	0.471
		MA-BIC	0.085	0.121	0.162	0.201	0.238	0.311	0.429
Scenario 1: Model Space \mathcal{M}_1									
0.8	500	BEST-AIC	0.171	0.216	0.265	0.314	0.364	0.490	0.725
		BEST-BIC	0.131	0.157	0.187	0.219	0.252	0.321	0.438
		MA-AIC	0.147	0.186	0.230	0.276	0.324	0.450	0.680
		MA-BIC	0.125	0.149	0.174	0.202	0.231	0.295	0.409
	1,000	BEST-AIC	0.123	0.151	0.182	0.213	0.244	0.318	0.478
		BEST-BIC	0.117	0.163	0.216	0.267	0.311	0.391	0.511
		MA-AIC	0.105	0.132	0.161	0.193	0.225	0.304	0.464
		MA-BIC	0.106	0.144	0.189	0.232	0.270	0.337	0.439
Scenario 2: Model Space \mathcal{M}_2									
0.8	500	BEST-AIC	0.136	0.189	0.251	0.312	0.372	0.522	0.817
		BEST-BIC	0.111	0.134	0.158	0.185	0.215	0.292	0.437
		MA-AIC	0.125	0.169	0.221	0.275	0.330	0.470	0.734
		MA-BIC	0.108	0.126	0.145	0.167	0.192	0.260	0.389
	1,000	BEST-AIC	0.113	0.146	0.184	0.222	0.260	0.353	0.558
		BEST-BIC	0.098	0.135	0.174	0.212	0.249	0.327	0.475
		MA-AIC	0.102	0.132	0.167	0.204	0.241	0.332	0.521
		MA-BIC	0.093	0.125	0.158	0.190	0.222	0.289	0.415

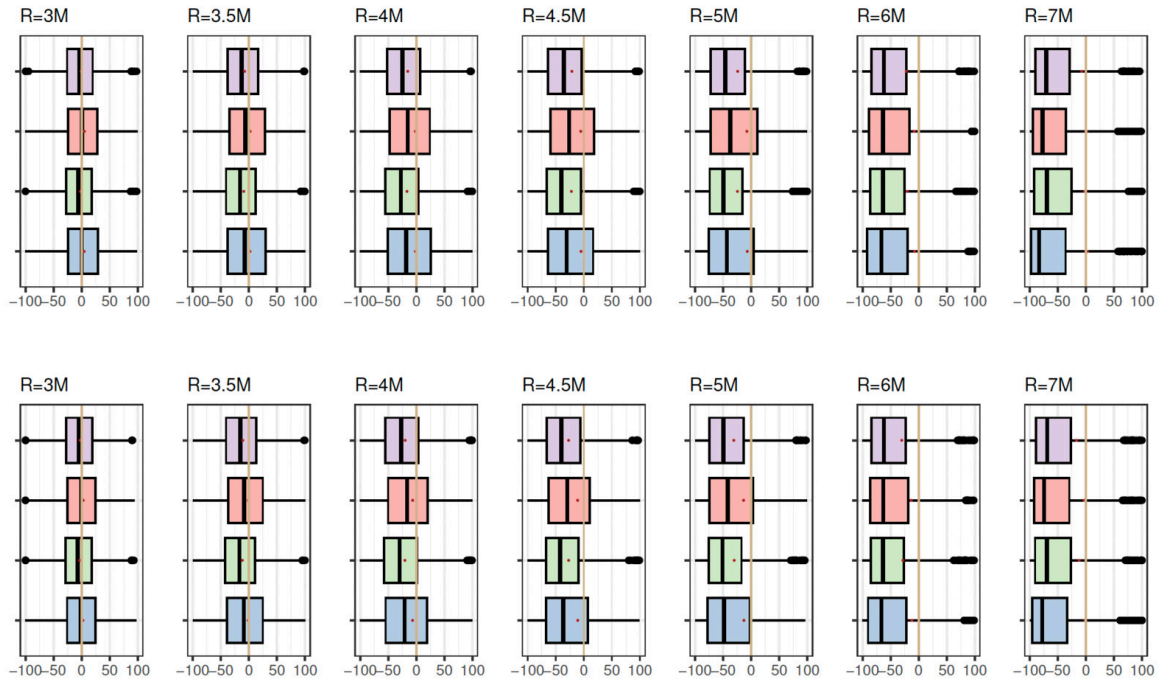


Figure 6. Panel plot illustrates the summary of relative errors for $c = 1.2$. The plot is organized in two rows, each representing a combination of sample size and model space. The first row corresponds to a sample size of $n = 50$ with \mathcal{M}_1 , the second row corresponds to a sample size $n = 50$ with \mathcal{M}_2 . Each row contains box plots for four methods: MA-BIC (soft lavender), MA-AIC (light coral), BEST-BIC (pale green), and BEST-AIC (pale blue).

Table 8. Summary of the mean of REs for each combination of the simulation settings when sample size is $n = 50$

c	n	Method	3M	3.5M	4M	4.5M	5M	6M	7M
Scenario 1: Model Space \mathcal{M}_1									
1.2	50	BEST-AIC	4.66	1.74	-1.79	-4.99	-7.05	-7.18	2.44
		BEST-BIC	-2.03	-9.35	-16.8	-22.2	-24.5	-20.5	-2.61
		MA-AIC	5.00	1.91	-1.96	-5.44	-7.67	-7.94	1.27
		MA-BIC	-0.67	-7.83	-15.4	-21.2	-24.2	-22.1	-7.57
Scenario 2: Model Space \mathcal{M}_2									
1.2	50	BEST-AIC	2.42	-1.81	-6.77	-10.8	-13.0	-12.0	-1.81
		BEST-BIC	-3.55	-11.8	-20.5	-27.1	-30.4	-28.1	-12.2
		MA-AIC	2.98	-1.40	-6.66	-11.1	-13.6	-13.2	-3.43
		MA-BIC	-2.52	-10.9	-20.0	-27.1	-31.0	-30.1	-16.9

Table 9. Summary of standard deviations of REs for each combination of the simulation settings when sample size is $n = 50$

c	n	Method	3M	3.5M	4M	4.5M	5M	6M	7M
Scenario 1: Model Space \mathcal{M}_1									
1.2	50	BEST-AIC	41.6	50.9	61.8	74.3	88.3	127.0	198.0
		BEST-BIC	36.5	44.1	52.2	61.5	72.9	108.0	180.1
		MA-AIC	39.3	47.8	57.4	68.9	82.1	120.1	189.2
		MA-BIC	35.6	42.6	49.6	58.0	68.6	102.0	169.0
Scenario 2: Model Space \mathcal{M}_2									
1.2	50	BEST-AIC	42.3	51.5	61.9	73.4	86.4	120	174.0
		BEST-BIC	36.1	42.4	48.6	55.5	64.2	91.3	147.1
		MA-AIC	40.1	48.5	57.8	68.0	79.9	112.0	165.0
		MA-BIC	35.7	41.4	46.7	52.4	60.1	84.7	135.2