Empirical Bayes Estimation of Structure Variables in the Collective Risk Model for Reserve Risk

Alessandro Ricotta and **Edoardo Luini**
Department of Statistics, University "La Sapienza" Rome
`{alessandro.ricotta, edoardo.luini}@uniroma1.it`

Abstract

The evaluation of reserve risk, which represents a fundamental part of underwriting risk for non-life insurers, can be achieved through a wide range of stochastic approaches, including the Collective Risk Model. This paper proposes a Bayesian technique to evaluate the standard deviation of structure variables embedded into the Collective Risk Model. We adopt prior distributions whose parameters are measured using the available data set making use of Mack's formula linked to bootstrap methodology. Moreover, correlation between structure variables is investigated with a Bayesian method, where a dependent bootstrap approach is adopted. Finally, a case study is carried out.

keywords: stochastic claims reserving, collective risk model, structure variables, Bayesian approach, bootstrap.

1 Introduction

Stochastic claims reserving models allow the assessment of the standard deviation or the probability distribution of claims reserve necessary to quantify the capital charge from a solvency point of view [12]. Moreover, evaluate the riskiness of outstanding claims plays a fundamental role in managing insurance companies. A variety of stochastic methodologies exist in literature. Mack proposed a first approach ([18], [19], [20]), which provides the prediction variance related to Chain-Ladder estimate; the variability of the reserve is herein split into Process Variance and Estimation Variance. Furthermore, other methodologies like Bootstrap ([8], [10]) and Generalized Linear Models ([9]) are used to determine the claims reserve distribution. In recent years, in stochastic claims reserving, Bayesian methods have become increasingly important and adopted. Their main advantages consist in the possibility to investigate distributions of model parameters and the chance to incorporate external information rigorously into actuarial models. Without being exhaustive, the principal deterministic methods developed under the Bayesian framework are Chain-Ladder ([32], [36], [25], [28]), Bornhuetter-Ferguson ([32], [36], [11]) and Overdispersed Poisson Model ([11]). Additionally, in [31] different Bayesian approaches to estimate claim frequency are presented and in [37], [1], [22], [38] it is shown a range of other Bayesian models for both incurred and paid loss data. Furthermore, Meyers ([21]) developed a Bayesian Collective Risk Model where the structure of parameters is based on the deterministic method called Cape Code; the expected loss ratio and the incremental paid loss development factor, which represent model parameters, are evaluated in a Bayesian manner. Within Bayesian theory, Empirical Bayes methods are procedures for statistical inference in which the prior distribution is estimated from the data. They are often thought of as a bridge between classical and Bayesian inference and are popularly employed by researchers and practitioners ([26]). A detailed introduction to the topic can be found in [5] and in [4]. In actuarial science this approach has been

mainly adopted in Credibility theory (see for example [23], [16] and [15]) and together with Chain-Ladder Model ([34]).
The Collective Risk Model (CRM) to assess claims reserve was proposed by different authors such as the International Actuarial Association ([14]), Meyers ([21]) and Savelli and Clemente ([29], [30]). This approach was extended by Ricotta and Clemente ([27]) assuming that incremental payments to be estimated in the run-off triangle are a compound mixed Poisson process, where the uncertainty on claim size is introduced with a multiplicative structure variable. The model considers, therefore, structure variables on claim count and claim size in order to describe parameter uncertainty on both random variables. In addition, linear dependence between different development and accidental years is addressed. The authors obtained the exact expressions of mean, standard deviation and skewness of the claims reserve distribution and showed the non-negligible impact that structure variables have on it. Finally, they present a method based on Mack's formula to quantify the standard deviation of structure variables.
The aim of this paper is to introduce a Bayesian procedure to estimate the standard deviation of the structure variables related to the Collective Risk Model as described in [27]. In addition, the dependence between model parameters is taken into account, i.e. claim count and claim size, caused by the deterministic average cost method; linear correlation, evaluated according to the Bayesian framework, is introduced in the CRM through structural risk factors.
Concerning the Bayesian approach adopted to quantify the standard deviation of structure variables, the bootstrap methodology jointed to the Mack's formula is used to enforce both the likelihood function and the prior distributions of Bayes' rule. Run-off triangles of different accounting years are considered with the aim to acquire all the accessible historical information available to the insurance company. The Bayesian method applied to evaluate correlation between structure variables is built on Mack's formula joined to a dependent bootstrap approach. The bootstrap methodology is herein carried out by jointly resampling in a dependent manner the data into the run-off triangles of claim count and average claim cost, namely entries that fill the same position in the respective run-off triangles. We use priors whose parameters are calibrated according to available data, following the empirical parametric scheme, for avoiding introduction of any sort of expert judgment. It is worth pointing out that the procedures herein described for determining priors can be implemented using other coherent run-off triangles data in order to incorporate external evidence, like market information, coming from different sources.
Model parameters different from the structure variables are calibrated by using a data set of individual claims and an average cost method; the deterministic Frequency-Severity method, based on the Chain-Ladder mechanics, is adopted to separately calculate the number of claims and the average cost for each cell of the bottom part of the run-off triangle.
Monte Carlo method is performed to simulate the claims reserve distribution according to the whole lifetime of insurer obligations. Furthermore, with regards to a one-year time horizon evaluation, we adapt the “re-reserving” method ([7], [24]) and estimate both the uncertainty of claims development result and the reserve risk capital requirement.
The paper is organized as follows. Section 2 introduces the Collective Risk Model as proposed in [27] and displays how to estimate parameters other than structure variables. In Section 3, the Bayesian approach is presented and performed to estimate the standard deviation of structure variables; at the same time we report results acquired according to the Metropolis-Hastings algorithm with respect to two non-life insurers. Moreover, the exact moments of structure variables are acquired. Section 4 refers to Pearson correlation coefficient between structure variables; results related to the two data sets are also reported. A case study on two non-life insures is shown in Section 5 where the Collective Risk Model is enforced to evaluate claims reserve distribution concerning both a total run-off and a one-year time horizon. In addition, we investigate the effect of linear correlation magnitude between structure variables on both claims reserve and the average Pearson correlation coefficient

affecting outstanding claims of different accident and development years. Conclusions follow.

2 Collective Risk Model

This section reports the main features of the Collective Risk Model developed in [27]. This model, based on the Collective Risk Theory, aims to assess claims reserve in a stochastic way. Here the claims reserve is represented through the run-off triangle: available data is reported in rectangular table of dimension $N \times N$ where rows ($i = 1, ..., N$) represent the claims accident years (AY), whereas columns ($j = 1, ..., N$) are the development years (DY) related to the number or the amount of claims. Data linked to observed incremental payments fill the upper triangle $D = \{X_{i,j}; i + j \leq N + 1\}$, where $X_{i,j}$ denotes incremental payments of claims in the cell (i, j), namely claims incurred in the generic accident year i and paid after $j - 1$ years of development. Analogously, the observed number of claims $n_{i,j}$ in the upper triangle is defined as $D^n = \{n_{i,j}; i + j \leq N + 1\}$.
Future numbers or amounts of payments must be assessed for each cell of the lower triangle. The scope is to investigate the random variable (r.v.)[1] of future incremental payments $\widetilde{X}_{i,j}$. The CRM represents incremental payments for each cell to be estimated as follows:

$$\widetilde{X}_{i,j} = \sum_{h=1}^{\widetilde{K}_{i,j}} \widetilde{p}\widetilde{Z}_{i,j,h}$$

and the r.v. claims reserve, denoted by $\widetilde{R}$, is equal to the sum of the cells of lower run-off triangle:

$$\widetilde{R} = \sum_{i=1}^{N} \sum_{j=N-i+2}^{N} \widetilde{X}_{i,j},$$

where:

- $\widetilde{K}_{i,j}$ represents the r.v. number of claims related to the accident year i and paid after $j-1$ years. This r.v. is assumed to be a mixed Poisson process; parameter uncertainty is addressed through a multiplicative structure variable $\widetilde{q}$ with mean 1 and standard deviation $\sigma_{\widetilde{q}}$. Therefore, the r.v. claims number is parametrized as follows: $\widetilde{K}_{i,j} \sim Po(\widetilde{q}n_{i,j})$.
- $\widetilde{Z}_{i,j,h}$ is the random variable describing the amount of the h^{th} claim occurred in the accident year i and paid after $j - 1$ years.
- $\widetilde{p}$ denotes the parameter uncertainty related to claim size. This structure variable has mean and standard deviation equal to 1 and $\sigma_{\widetilde{p}}$ respectively.

The two structure variables enable the introduction of parameter uncertainty without affecting the expected value of claim number and amount. Furthermore, in the bottom part of the run-off triangle only one r.v. affects the claim number and the claim size respectively, allowing for the dependence between these random variables of different AY and DY, given by the settlement process. In [27] the exact moments of claims reserve are obtained under the following assumptions:

- claim number ($\widetilde{K}_{i,j}$), claim cost ($\widetilde{Z}_{i,j,h}$), and the structure variable $\widetilde{p}$ are mutually independent in each cell (i, j) of the lower run-off triangle;
- claim size values in different cells of the lower run-off triangle are independent and in the same cell are independent and identically distributed (i.i.d.);
- structure variable $\widetilde{q}$ is independent of the claim costs in each cell;

[1] A tilde superscript will henceforth denote random variables.

- $\tilde{q}$ and $\tilde{p}$ are mutually independent.

The expected value of the claims reserve is equal to:

$$E\left(\tilde{R} \mid D; D^n\right) = \sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} m_{i,j},$$

where $n_{i,j}$ and $m_{i,j}$ are the expected number and the average cost of paid claims respectively. As is shown later, these model parameters are estimated with the Frequency-Severity method; therefore, the expected value of the stochastic model corresponds to the claims reserve estimated by the underlying deterministic method.
The variance of reserve is:

$$\operatorname{Var}\left(\tilde{R} \mid D; D^n\right) = E\left(\tilde{p}^2\right) \sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} a_{2,\tilde{Z}_{i,j}} + \sigma^2_{\tilde{q}\tilde{p}} \left(\sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} m_{i,j} \right)^2, \tag{2.1}$$

where $E\left(\tilde{p}^2\right)$ is the quadratic mean of the r.v. $\tilde{p}$, $a_{k,\tilde{Z}_{i,j}} = E\left(\tilde{Z}^k_{i,j}\right)$ the raw moment of order k of the severity distribution and $\sigma^2_{\tilde{q}\tilde{p}}$ defines the variance[2] of the r.v. determined as the product between $\tilde{q}$ and $\tilde{p}$.
The ratio between the standard deviation of the reserve, $SD\left(\tilde{R} \mid D; D^n\right)$, and its expected value defines the coefficient of variation:

$$CV\left(\tilde{R} \mid D; D^n\right) = \sqrt{\sigma^2_{\tilde{q}\tilde{p}} + \frac{E\left(\tilde{p}^2\right) \sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} a_{2,\tilde{Z}_{i,j}}}{\left(\sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} m_{i,j} \right)^2}}. \tag{2.2}$$

It is to be highlighted that the relative variability of claims reserve decreases with respect to the number of reserved claims; nevertheless, the standard deviation of the r.v. defined as product between the two structure variables represents a systemic risk (i.e. non-pooling risk), which cannot be diversified by a larger portfolio.
Finally, the skewness of the claims reserve, defined as ratio between the third central moment and the cube of the standard deviation, is given by:

$$\begin{aligned} \gamma\left(\tilde{R} \mid D; D^n\right) = &\frac{\gamma_{\tilde{q}\tilde{p}} \sigma^2_{\tilde{q}\tilde{p}} \left(\sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} m_{i,j} \right)^3 + E\left(\tilde{p}^3\right) \sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} a_{3,\tilde{Z}_{i,j}}}{\left[E\left(\tilde{p}^2\right) \sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} a_{2,\tilde{Z}_{i,j}} + \sigma^2_{\tilde{q}\tilde{p}} \left(\sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} m_{i,j} \right)^2 \right]^{3/2}} \\ &+ \frac{\left(\sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} m_{i,j} \right) \left(\sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} a_{2,\tilde{Z}_{i,j}} \right) \left[E\left(\tilde{p}^3\right) E\left(\tilde{q}^2\right) - E\left(\tilde{p}^2\right) \right]}{\left[E\left(\tilde{p}^2\right) \sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} a_{2,\tilde{Z}_{i,j}} + \sigma^2_{\tilde{q}\tilde{p}} \left(\sum_{i=1}^{N} \sum_{j=N-i+2}^{N} n_{i,j} m_{i,j} \right)^2 \right]^{3/2}}. \end{aligned} \tag{2.3}$$

Each term of the formula (2.3) depends on the two structure variables; it is to be noted that the numerator displays the term $\gamma_{\tilde{q}\tilde{p}}$[3] that represents the skewness of the r.v. determined as product

[2] This term can be written as $\sigma^2_{\tilde{q}\tilde{p}} = \left(\sigma^2_{\tilde{p}} + 1\right) \sigma^2_{\tilde{q}} + \sigma^2_{\tilde{p}}$.

[3] $\gamma_{\tilde{q}\tilde{p}} = \frac{E(\tilde{p}^3) E(\tilde{q}^3) - 3\sigma^2_{\tilde{q}\tilde{p}} - 1}{\sigma^3_{\tilde{q}\tilde{p}}}$.

between $\tilde{q}$ and $\tilde{p}$. It is possible to show that the limit of the skewness of the claims reserve converges to the value $\gamma_{\tilde{q}\tilde{p}}$ when the number of claims increases.
Formulae (2.1), (2.2) and (2.3) show that the main characteristics of the reserve are affected by the structure variables with an impact that depends on the variability of $\tilde{q}$ and $\tilde{p}$.
It is to be pointed out that if the assumption of mutual independence between the two structure variables is relaxed, as in Section 4, the formulae (2.1), (2.2) and (2.3) do not hold and it is no longer possible to achieve closed-form expressions for the moments of the claims reserve.
To apply the CRM we need to estimate a set of parameters for each cell (i, j) of the lower triangle. The expected number of paid claims $(n_{i,j})$ and the expected claims cost $(m_{i,j})$ are obtained, conditionally to the set of information D (the run-off triangle of incremental payments) and D^n (the run-off triangle of incremental number of paid claims), with a deterministic average cost method. We use the Frequency-Severity method by applying the Chain-Ladder mechanics on the triangles of cumulative numbers and cumulative average costs. The other quantities necessary to implement the CRM are the cumulants of the severity. According to the claims data set, we estimate the variability coefficient of the claim size for each DY; later, adopting a distribution assumption, the moments of the r.v. $\tilde{Z}_{i,j}$ are obtained. Finally, to enforce the Collective Risk Model we need to evaluate the characteristics of the structure variables; the next section reports a Bayesian approach to assess their standard deviation, whereas Section 4 shows how to estimate Pearson correlation coefficient between $\tilde{q}$ and $\tilde{p}$.

3 Bayesian approach to estimate the standard deviation of structure variables

In classical statistics the parameters of a model are assumed to be fixed; Bayesian statistics contrasts with this approach and considers parameters to be random variables (an exhaustive dissertation of the topic can be found in [2], [17] and [13]). The aim of the Bayesian approach is to take parameters uncertainty into account; this variability is introduced through prior probability distributions that, together with observed data, allow the posterior probability distribution of the model parameters to be achieved. According to the Bayes theorem, the parameter posterior distribution, $f(\theta \mid x)$, can be computed as:

$$f(\theta \mid x) = \frac{f(x \mid \theta)\, f(\theta)}{f(x)}.$$

The term $f(x \mid \theta)$ is the sampling density of data under a chosen probability model; this element, viewed as function of θ for fixed x, is the likelihood function. The parameter prior distribution is $f(\theta)$, which refers to the parameter uncertainty, also interpretable as the prior opinion or knowledge related to parameter values. The denominator of Bayes' formula represents the marginal distribution of data and can be rewritten (considering a continuous sample space) as the integral of the sampling density multiplied by the prior over the sample space of θ, $f(x) = \int f(x \mid \theta)\, f(\theta)\, d\theta$. This quantity does not depend on θ and with fixed x turns out to be a constant quantity which acts as a normalizing factor that leads to a proper posterior distribution. Bayes theorem is often considered without the normalizing constant that has only the effect of rescaling the density:

$$f(\theta \mid x) \propto f(x \mid \theta)\, f(\theta).$$

Hence, the posterior distribution is proportional to the product of likelihood function and prior. Bayes' formula depends on data and prior distribution. The latter represents the opinion concerning the chances related to the different values that θ can assume. In general, as the sample size increases, the impact of prior becomes negligible on posterior distribution, namely with enough data the posterior distribution is independent of the prior information. When taking into consideration discrete parameter distribution, as $x \to +\infty$, the posterior shows an asymptotic behaviour that leads to a degenerate distribution on the true value of θ, whereas the posterior distribution of a continuous parameters θ, in general, converges to a normal distribution centered at its maximum likelihood estimate.

Bayes' approach also allows us to make inference on future observation through the posterior predictive distribution, where the adjective posterior refers to the consideration that the distribution is conditional to the observed data (x), and predictive because it is a prediction of new observable data (y). The posterior predictive distribution is an average of the probability distribution of y conditional on the unknown value of θ, weighted with the posterior distribution of θ:

$$f\left(y \mid x\right) = \int f\left(y \mid \theta\right) f\left(\theta \mid x\right) \, d\theta.$$

If the assumptions on the probability model are appropriate, the posterior predictive distribution $f\left(y \mid x\right)$ converges, when the number of observed data increase, to the real generating distribution of data, $f\left(y \mid \theta\right)$.

Hence, outcomes of the Bayesian analysis are the posterior predictive distribution, which provides information about new observations, and the posterior distribution, which contains information about the parameters underlying the model. With regards to the posterior distribution, it is possible to summarize this information by developing different types of inference analysis on this distribution (i.e. both point or region estimation and hypothesis testing). Concerning the point estimation, the choice of an estimator can be expressed as a decision problem where an appropriate loss function must be specified. When a quadratic loss function is selected, the Bayes estimator is the posterior mean. Other loss functions lead to the posterior median and posterior mode; it is to be noted that the Bayes estimator can theoretically assume any possible parameter value depending on the selected loss function.

The Bayesian framework here is used to calibrate the standard deviation of structure variables of CRM. These variables related to claim count and claim cost do not affect the expected value of the reserve but have an impact on the other characteristics (i.e. variance, skewness and so on).

As adopted in [27], we follow the usual assumption of Collective Risk Theory that structure variables are gamma distributed with identical parameters:

$$\widetilde{q} \sim \text{Gamma}\left(h; h\right), \quad \widetilde{p} \sim \text{Gamma}\left(k; k\right).$$

The variables $\widetilde{q}$ and $\widetilde{p}$ have mean equal to 1, given by the ratio of the parameters, and standard deviation $\sigma_{\widetilde{q}} = 1/\sqrt{h}$ and $\sigma_{\widetilde{p}} = 1/\sqrt{k}$. Therefore, the values of $\sigma_{\widetilde{q}}$ and $\sigma_{\widetilde{p}}$ determine the parameter of interest, h and k, which all the characteristics of the structure variable depend upon.

In [27] a deterministic approach based on the Estimation Variance derived via Mack is proposed to assess the parameters of structure variables. In Mack's formula, the Estimation Error measures the variability produced by the parameters estimation; because of this, it is ascribable to the structure variables that have the aim to introduce parameters uncertainty on quantities being considered (i.e. claim count and severity).

Here the standard deviations $\sigma_{\widetilde{q}}$ and $\sigma_{\widetilde{p}}$ are interpreted as random variables and consequently denoted by a tilde. Random variables and their parameters are denoted with the subscript $\widetilde{q}$ or $\widetilde{p}$ to indicate which r.v. is considered in the Bayes approach, whereas if general considerations are carried out, the subscript is omitted for a simpler notation. It is assumed that $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$, defined for positive values, follow a gamma distribution:

$$\widetilde{\sigma} \sim \text{Gamma}\left(\widetilde{A}; \widetilde{B}\right),$$

where the parameters $\widetilde{A}$ and $\widetilde{B}$ are random variables with regards to prior information is conveyed. Parameters of $\widetilde{A}$ and $\widetilde{B}$ are called hyperparameters of the model.

In this context, the evaluation of the standard deviation of structure variables is acquired through the Bayes' formula with the purpose to obtain a posterior distribution of the parameters which $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ depend on:

$$f(A, B \mid \widetilde{\sigma}) \propto f\left(\sigma \mid \widetilde{A}, \widetilde{B}\right) f(A) f(B). \tag{3.1}$$

With regards to the posterior distributions achieved via the Bayesian method, their expected values are used to calibrate the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$. Therefore, the posteriors means are adopted to estimate the parameters of the r.v. $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$:

$$\widetilde{\sigma} \sim \text{Gamma}\left(E\left(\widetilde{A} \mid \widetilde{\sigma}\right); E\left(\widetilde{B} \mid \widetilde{\sigma}\right)\right). \tag{3.2}$$

It may be noted that, as depicted in formula (3.1) above, we are assuming $\widetilde{A}$ and $\widetilde{B}$ prior probability distributions to be independent; this premise is however neither affecting nor restrictive on our model for two reasons. First and foremost, since only the posterior expected values of $\widetilde{A}$ and $\widetilde{B}$ enter formula (3.2), we are looking separately at the marginal posterior distributions of the parameters (when computing one parameter expectation, the other one is automatically marginalized out). Secondly, in the outlined framework, even if starting with independent priors, the Bayes theorem formula will generate a dependent posterior distribution, whose dependency is induced by the likelihood function.
Likelihood functions and prior distributions of formula (3.1) are implemented making use of Mack's formula and bootstrap methodology. The latter is carried out following the approach adopted in [9]. The bootstrap method, according to the Chain-Ladder technique, by resampling the upper triangle of model residuals, allows us to create different resampled data sets which can be used to calculate the quantity of interest and make inference on it. For our purposes, we applied the bootstrap approach to the run-off triangles of the cumulative claim count and cumulative average costs; for each iteration the relative variability related to the Estimation Error derived via Mack's formula is calculated, with the aim to measure the variability produced by the parameters estimation. In particular, regarding each resampled triangle via bootstrap, the Mack's formula is implemented on the resampled run-off triangles of the cumulative claim count and cumulative average costs respectively. The squared root of the Estimation Variance for both triangles is divided by the respective Chain-Ladder estimate (i.e. the mean of frequency and severity). These relative variabilities, concerning only the Estimation Error, are interpreted as the coefficient of variation of the structure variables $\widetilde{q}$ and $\widetilde{p}$; bearing in mind that their means are equal to 1, these values correspond to the standard deviations $\sigma_{\widetilde{q}}$ and $\sigma_{\widetilde{p}}$ and are interpreted as the uncertainty related to the parameters estimate.
This method, that we call the *Mack-Bootstrap approach*, is applied to the run-off triangles of both claim count and average cost of different dimensions; the aim is to take into consideration all available historical information. Starting from the run-off triangle related to the current accounting year, with dimension $N \times N$, the Mack-Bootstrap approach is performed on triangles obtained by gradually deleting the last diagonal available, one at a time. Therefore, the run-off triangles of different accounting years are considered up to the current triangle, where the first run-off triangle is chosen starting from the oldest information considered representative. Hence, in respect to the last l accounting years, the run-off triangles have dimensions $(N - l + 1 \times N - l + 1), \ldots, (N \times N)$ respectively and the Mack-Bootstrap approach lets us obtain for each historical triangle the sample distribution of random variables $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$.
The likelihood functions $f\left(\sigma_{\widetilde{q}} \mid \widetilde{A}_{\widetilde{q}}, \widetilde{B}_{\widetilde{q}}\right)$ and $f\left(\sigma_{\widetilde{p}} \mid \widetilde{A}_{\widetilde{p}}, \widetilde{B}_{\widetilde{p}}\right)$, based on the gamma model, are evaluated at the expected values of the distributions of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ related to the sequence of the l historical triangles. Thus, concerning a generic historical triangle, the sample mean of the distribution of the standard deviation, namely the average variability of parameter estimation that affects the triangle, is adopted as an estimate of the true unobservable value of the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$. Hence, the l values of

$E\left(\widetilde{\sigma}_{\widetilde{q}}\right)$ and $E\left(\widetilde{\sigma}_{\widetilde{p}}\right)$ are interpreted as data and used to compute likelihood functions; we assume this data to be independent and identically distributed. However, it is to be noted that the latter assumption, useful to calculate the likelihood, in practice does not fully hold since data is attained on the l historical triangles that share common cells, affecting the assumption of independence; moreover, the model is lacking in conditions apt to fulfill the identical distribution assumption of data.
Prior distributions, which describe the knowledge related to the positive parameters of the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$, are modeled via gamma and exponential distributions. We consider empirical priors distributions, namely distributions whose parameters are calibrated using available data, so as to prevent any sort of expert judgment. The choice of two distributions with different shapes allows us to evaluate the sensitivity of posterior distribution in respect to the priors. When a gamma distribution is adopted, priors are calibrated taking into consideration the distributions of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ achieved both on the most recent run-off triangle via the Mack-Bootstrap approach, and on the characteristics of the distributions related to run-off triangles of past accounting years. On the other hand, when priors follow an exponential distribution, they are calibrated only on the distributions of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ related to the run-off triangles of current year.
When priors are described by gamma distribution, $f(A)$ and $f(B)$ are parametrized as follows (formulae related to to the r.v. $\widetilde{\sigma}_{\widetilde{p}}$ are, *mutatis mutandis*, identical):

$$\widetilde{A}_{\widetilde{q}} \sim \text{Gamma}\left({}_1\alpha_{\widetilde{q}};{}_2\alpha_{\widetilde{q}}\right) \qquad \widetilde{B}_{\widetilde{q}} \sim \text{Gamma}\left({}_1\beta_{\widetilde{q}};{}_2\beta_{\widetilde{q}}\right),$$

where α and β are the hyperparameters and are calibrated as shown below.
With regards to the r.v. $\widetilde{\sigma}_{\widetilde{q}}$, bearing in mind that $\widetilde{\sigma}_{\widetilde{q}} \sim \text{Gamma}\left(\widetilde{A}_{\widetilde{q}};\widetilde{B}_{\widetilde{q}}\right)$, the expected value and variance are equal to:

$$E\left(\widetilde{\sigma}_{\widetilde{q}}\right) = \frac{\widetilde{A}_{\widetilde{q}}}{\widetilde{B}_{\widetilde{q}}} \qquad \text{Var}\left(\widetilde{\sigma}_{\widetilde{q}}\right) = \frac{\widetilde{A}_{\widetilde{q}}}{\widetilde{B}^2_{\widetilde{q}}}.$$

Concerning the distribution of $\widetilde{\sigma}_{\widetilde{q}}$ acquired on the most recent run-off triangle, we calculate $E\left(\widetilde{\sigma}_{\widetilde{q}}\right)$ and $\text{Var}\left(\widetilde{\sigma}_{\widetilde{q}}\right)$, and then define, through the method of moments, $\widetilde{A}_{\widetilde{q}}$ and $\widetilde{B}_{\widetilde{q}}$, which will be adopted as estimates of the expected values of prior distributions. On the other hand, the relative variability of priors, set identical for $\widetilde{A}_{\widetilde{q}}$ and $\widetilde{B}_{\widetilde{q}}$, is calculated as the mean of the variability coefficients of $E\left(\widetilde{\sigma}_{\widetilde{q}}\right)$ and $SD\left(\widetilde{\sigma}_{\widetilde{q}}\right)$ values computed on the distribution of $\widetilde{\sigma}_{\widetilde{q}}$ of the run-off triangles concerning the last l accounting years. Hence, priors coefficient of variability reflects the variability of sample mean and standard deviation stemming from Mack-Bootstrap distributions of historical run-off triangles.
Hyperparameters are therefore estimated according to the method of moments as follows:

$${}_1\alpha_{\widetilde{q}} = \frac{1}{CV\left(\widetilde{A}_{\widetilde{q}}\right)^2} \quad {}_2\alpha_{\widetilde{q}} = \frac{{}_1\alpha_{\widetilde{q}}}{E\left(\widetilde{A}_{\widetilde{q}}\right)}, \qquad {}_1\beta_{\widetilde{q}} = \frac{1}{CV\left(\widetilde{B}_{\widetilde{q}}\right)^2} \quad {}_2\beta_{\widetilde{q}} = \frac{{}_1\beta_{\widetilde{q}}}{E\left(\widetilde{B}_{\widetilde{q}}\right)}.$$

When an exponential distribution is assumed as prior, $f(A)$ and $f(B)$ are parametrized as:

$$\widetilde{A}_{\widetilde{q}} \sim \text{Exp}\left(\lambda_{\widetilde{q}}\right) \qquad \widetilde{B}_{\widetilde{q}} \sim \text{Exp}\left(\eta_{\widetilde{q}}\right).$$

Hyperparameters are estimated adopting only the distribution of the r.v. $\widetilde{\sigma}_{\widetilde{q}}$ related to the current accounting year. The expected values of priors, as in the previous case, are derived by the mean and variance of $\widetilde{\sigma}_{\widetilde{q}}$ and the parameters $\lambda_{\widetilde{q}}$ and $\eta_{\widetilde{q}}$ are determined as:

$$\lambda_{\widetilde{q}} = \frac{1}{E\left(\widetilde{A}_{\widetilde{q}}\right)} \qquad \eta_{\widetilde{q}} = \frac{1}{E\left(\widetilde{B}_{\widetilde{q}}\right)}.$$

Consequently, when priors are exponential distributions, their standard deviations and all the other characteristics are determined by the unique parameter upon which exponential distribution depends.

It is to be noted that prior distributions are calibrated with the aim to provide, on average, the mean and variance of the distributions of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ achieved via Mack-Bootstrap approach on the most recent run-off triangle.

The Bayesian procedure has been applied to the claim data sets of two non-life insurance companies working in the Motor Third Party Liability (MTPL) line of business and concerning accounting years from 1993 to 2004. DELTA insurer is a small-medium company, whereas OMEGA insurer is roughly 10 times larger. Appendix A reports the run-off triangles adopted to estimate, via the Frequency-Severity deterministic method, the claims reserve. Triangles related to cumulative claims count and cumulative average costs are used to enforce the Bayesian methodology detailed above.

The Mack-Bootstrap approach has been carried out regarding run-off triangles for 9 accounting years; therefore, the triangles adopted to acquire historical data and calibrate prior distributions have dimensions from 4×4 to 12×12. The number of iterations carried out in the bootstrap stage is equal to 10,000.

In respect to the insurer DELTA, the distributions of $\widetilde{\sigma}_{\widetilde{q}}$ acquired via the Mack-Bootstrap approach related to the 9 historical triangles of claim count show expected values included between 1.66% and 2.48%; for $\widetilde{\sigma}_{\widetilde{p}}$ the minimum value of the mean is 2.33% whereas the maximum is 4.29%. Table 1 details the expected values, 5% quantile and 95% quantile related to the distributions of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$.

Insurer: DELTA

Dimension		4×4	5×5	6×6	7×7	8×8	9×9	10×10	11×11	12×12
$\widetilde{\sigma}_{\widetilde{q}}$	**Exp. Value**	1.83%	1.66%	1.75%	2.30%	2.48%	2.40%	2.32%	2.42%	2.26%
	Quantile 5%	0.66%	0.82%	1.08%	1.32%	1.66%	1.66%	1.72%	1.84%	1.76%
	Quantile 95%	3.37%	2.82%	2.69%	3.58%	3.64%	3.45%	3.21%	3.28%	3.01%
$\widetilde{\sigma}_{\widetilde{p}}$	**Exp. Value**	4.15%	4.29%	2.81%	2.72%	2.93%	2.68%	2.66%	2.67%	2.33%
	Quantile 5%	1.11%	1.78%	1.36%	1.47%	1.84%	1.74%	1.81%	1.89%	1.72%
	Quantile 95%	9.77%	8.32%	5.09%	4.78%	4.65%	4.24%	4.04%	3.93%	3.23%

Table 1: Insurer DELTA - Expected value, 5% quantile and 95% quantile of the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ related to the historical triangles with dimensions from 4×4 to 12×12.

With respect to the larger insurer, OMEGA, the Mack-Bootstrap approach leads to expected values between 2.04% and 4.94% for $\widetilde{\sigma}_{\widetilde{q}}$, and between 2.34% and 3.16% for $\widetilde{\sigma}_{\widetilde{p}}$. Table 2 depicts the means and quantiles of order 5% and 95% regarding the Mack-Bootstrap distributions of the two random variables.

Insurer: OMEGA

Dimension		4×4	5×5	6×6	7×7	8×8	9×9	10×10	11×11	12×12
$\widetilde{\sigma}_{\widetilde{q}}$	**Exp. Value**	4.94%	3.29%	2.41%	2.25%	2.04%	2.04%	2.36%	2.80%	2.98%
	Quantile 5%	1.51%	1.32%	1.06%	1.24%	1.23%	1.33%	1.52%	1.88%	2.14%
	Quantile 95%	9.38%	5.92%	4.15%	3.57%	3.10%	2.98%	3.52%	4.20%	4.26%
$\widetilde{\sigma}_{\widetilde{p}}$	**Exp. Value**	2.38%	3.16%	2.93%	3.02%	2.62%	2.67%	2.38%	2.70%	2.34%
	Quantile 5%	0.62%	1.30%	1.46%	1.63%	1.48%	1.59%	1.49%	1.67%	1.61%
	Quantile 95%	5.42%	6.29%	5.28%	5.27%	4.57%	4.47%	3.84%	4.43%	3.48%

Table 2: Insurer OMEGA - Expected value, 5% quantile and 95% quantile of the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ related to the historical triangles with dimensions from 4×4 to 12×12.

As explained above, priors are calibrated according to the distributions of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$. In Appendix B the values of hyperparameters are reported concerning both gamma and exponential distributions.

Having calibrated the prior distributions, the next step is to calculate the posteriors. Since the posterior distributions being examined do not have a closed-form expression, we make use of the Metropolis-Hastings algorithm to draw samples from them. For generating a sample (commonly referred to as

chain) from the posterior distribution, this Markov Chain Monte-Carlo method requires only a function that is proportional to the real density, rather than exactly equal to it, avoiding the calculation of the normalization factor, which is extremely difficult in practice, especially when dealing with multi-dimensional distributions ([33]). In particular, a Random Walk Metropolis algorithm has been selected; this version of the Metropolis-Hastings design operates by proposing that the chain move to a candidate state obtained by disturbing the current one with a noise. Under mild conditions the chain converges to its stationary distribution and posterior quantities can be estimated from the simulation output[4]. A comprehensive dissertation of the topic can be found in [3].
Concerning the posterior distributions achieved via the above-mentioned algorithm, for our purposes, we use the expected values to assess the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$, whose parameters are set equal to the means of the posteriors, as shown in formula (3.2). It is to be noted that, for both insurers, gamma and exponential prior distributions lead to similar expected values of posteriors (see Appendix B). Therefore, prior distributions with very different shapes but the same mean do not significantly impact the expected values of posteriors. Indeed, in our case study both the characteristics of data adopted to calculate likelihood function and priors, calibrated to have same mean, determine a posterior expected value negligibly affected by the distribution family adopted for prior distributions. Hence, when the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ are calibrated, their characteristics are very close under the two assumptions related to prior distributions. Tables 3 and 4 indicate the expected values and coefficient of variations for both insurers:

Insurer: DELTA

Random variable	$\widetilde{\sigma}_{\widetilde{q}}$		$\widetilde{\sigma}_{\widetilde{p}}$	
Type of prior	**Gamma**	**Exponential**	**Gamma**	**Exponential**
Expected value	2.16%	2.16%	3.01%	3.02%
Coeff. of variation	17.30%	15.97%	19.70%	20.74%

Table 3: Insurer DELTA - Expected values and coefficient of variation of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$.

Insurer: OMEGA

Random variable	$\widetilde{\sigma}_{\widetilde{q}}$		$\widetilde{\sigma}_{\widetilde{p}}$	
Type of prior	**Gamma**	**Exponential**	**Gamma**	**Exponential**
Expected value	2.78%	2.78%	2.69%	2.70%
Coeff. of variation	25.08%	27.40%	23.45%	16.58%

Table 4: Insurer OMEGA - Expected values and coefficient of variation of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$.

The assessment of the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ allows us to determine the moments of the structure variables $\widetilde{q}$ and $\widetilde{p}$ adopted into the CRM. Taking into consideration the structural risk factors related to claim count (identical considerations hold for $\widetilde{p}$), we assume that $\widetilde{q} \sim \text{Gamma}\left(\frac{1}{\widetilde{\sigma}^2_{\widetilde{q}}};\frac{1}{\widetilde{\sigma}^2_{\widetilde{q}}}\right)$ and $\widetilde{\sigma}_{\widetilde{q}} \sim \text{Gamma}\left[E\left(\widetilde{A}^{post}_{\widetilde{q}}\right);E\left(\widetilde{B}^{post}_{\widetilde{q}}\right)\right]$, where $E\left(\widetilde{A}^{post}_{\widetilde{q}}\right)$ and $E\left(\widetilde{B}^{post}_{\widetilde{q}}\right)$ represent the expected values of the poster distributions , $E\left(\widetilde{A}_{\widetilde{q}} \mid \widetilde{\sigma}_{\widetilde{q}}\right)$ and $E\left(\widetilde{B}_{\widetilde{q}} \mid \widetilde{\sigma}_{\widetilde{q}}\right)$.
Moments of the structure variable depend on the parameters of the mixing variable $\widetilde{\sigma}_{\widetilde{q}}$, which however does not affect the mean of $\widetilde{q}$ that remains equal to 1 (see Appendix C for details). The variance is described by the following formula:

$$\text{Var}\left(\widetilde{q}\right) = \frac{E\left(\widetilde{A}^{post}_{\widetilde{q}}\right)\left[E\left(\widetilde{A}^{post}_{\widetilde{q}}\right)+1\right]}{E\left(\widetilde{B}^{post}_{\widetilde{q}}\right)^2}.$$

[4] **R** programming language for statistical computation (version 3.4.3) has been adopted for the implementation of the Markov Chain Monte Carlo algorithm. A useful package for the topic is [6].

The coefficient of variation, equal to the squared root of the variance, is:

$$CV(\widetilde{q}) = \frac{\sqrt{E\left(\widetilde{A}_{\widetilde{q}}^{post}\right)\left[E\left(\widetilde{A}_{\widetilde{q}}^{post}\right)+1\right]}}{E\left(\widetilde{B}_{\widetilde{q}}^{post}\right)}.$$

Finally, the skewness of the structure variable is given by:

$$\gamma(\widetilde{q}) = \frac{2\left[E\left(\widetilde{A}_{\widetilde{q}}^{post}\right)+2\right]\left[E\left(\widetilde{A}_{\widetilde{q}}^{post}\right)+3\right]}{E\left(\widetilde{B}_{\widetilde{q}}^{post}\right)\sqrt{E\left(\widetilde{B}_{\widetilde{q}}^{post}\right)\left[E\left(\widetilde{B}_{\widetilde{q}}^{post}\right)+1\right]}}.$$

Tables 5 and 6 report the exact characteristics for the structure variable $\widetilde{q}$ and $\widetilde{p}$ for both the insurers:

Insurer: DELTA

Type of prior	Gamma		Exponential	
Structure variable	$\widetilde{q}$	$\widetilde{p}$	$\widetilde{q}$	$\widetilde{p}$
Expected value	1	1	1	1
Coeff. of variation	2.20%	3.07%	2.18%	3.09%
Skewness	0.049	0.071	0.048	0.073

Table 5: Insurer DELTA - Expected value, coefficient of variation and skewness related to structure variables $\widetilde{q}$ and $\widetilde{p}$.

Insurer: OMEGA

Type of prior	Gamma		Exponential	
Structure variable	$\widetilde{q}$	$\widetilde{p}$	$\widetilde{q}$	$\widetilde{p}$
Expected value	1	1	1	1
Coeff. of variation	2.87%	2.76%	2.88%	2.73%
Skewness	0.072	0.068	0.076	0.061

Table 6: Insurer OMEGA - Expected value, coefficient of variation and skewness related to structure variables $\widetilde{q}$ and $\widetilde{p}$.

4 Bayesian estimation of Pearson correlation coefficient between structure variables

The Collective Risk Model assumes that claim count and claim size are mutually independent in each cell (i, j) of the lower run-off triangle. However, this theoretical assumption does not hold in practice due to the dependence introduced on model parameters by the average cost method (i.e. Frequency-Severity). The aim of this section is to evaluate, using a Bayesian procedure, the Pearson correlation coefficient between claim count and claim cost, estimated on structure variables $\widetilde{q}$ and $\widetilde{p}$. Similarly to Section 3, we adopt a method based on bootstrap resampling and Mack's formula, in which, however, the former considers the dependency between the run-off triangles of claim count and average claim cost, by resampling pairs of data which fill the same position in the respective triangles. The scope is to build up the distributions of the r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ by implicitly allowing for the dependence between the two data sets of claim count and claim cost. Hence, the estimated Pearson correlation coefficient is used to calibrate a Gaussian copula with the purpose to set up a two-dimensional random variable where the marginals are the two r.v.s $\widetilde{q}$ and $\widetilde{p}$ calibrated in the previous section.

In the Bayesian framework, the Pearson correlation coefficient is interpreted as a random variable following a beta distribution:

$$\widetilde{\rho} \sim \text{Beta}\left(\widetilde{C};\widetilde{D}\right).$$

We analysed the dependence between claim count and claim cost on the interval $[0,1]$; therefore, we assume parameter variabilities to be positively correlated. As usual, the r.v.s $\widetilde{C}$ and $\widetilde{D}$ identify prior distributions. According to Bayes' rule we obtain a posterior distribution of the parameters which $\widetilde{\rho}$ depends on:

$$f\left(C,D \mid \widetilde{\rho}\right) \propto f\left(\rho \mid \widetilde{C},\widetilde{D}\right) f(C) f(D). \tag{4.1}$$

The expected value of the posterior is used to calibrate the r.v. $\widetilde{\rho}$:

$$\widetilde{\rho} \sim \text{Beta}\left(E\left(\widetilde{C} \mid \widetilde{\rho}\right); E\left(\widetilde{D} \mid \widetilde{\rho}\right)\right). \tag{4.2}$$

Finally, the mean of $\widetilde{\rho}$, calculated with the posterior expected value, is adopted to assess the Gaussian copula used to join the marginals $\widetilde{q}$ and $\widetilde{p}$.
Terms of formula (4.1) are performed making use of Mack's formula and the dependent bootstrap approach.
The likelihood function based on the beta model is executed using the Pearson correlation coefficient calculated between the distribution of $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ related the sequence of the l historical triangles.
Priors are described via gamma and exponential random variables. In the first case, the distributions are calibrated according to the linear correlation related to the historical series of run-off triangles in conjunction with an *a priori* assumption on variability. On the other hand, when we adopt an exponential distribution, priors are calibrated only with the historical values of Pearson correlation coefficient.
When $f(C)$ and $f(D)$ are gamma distributed, we have:

$$\widetilde{C} \sim \text{Gamma}\left({}_1\varepsilon; {}_2\varepsilon\right) \qquad \widetilde{D} \sim \text{Gamma}\left({}_1\psi; {}_2\psi\right),$$

where ε and ψ are the hyperparameters calibrated as shown below.
Concerning the r.v. $\widetilde{\rho}$, modeled as a $\text{Beta}\left(\widetilde{C};\widetilde{D}\right)$, the expected value and variance are:

$$E\left(\widetilde{\rho}\right) = \frac{\widetilde{C}}{\widetilde{C}+\widetilde{D}} \qquad \text{Var}\left(\widetilde{\rho}\right) = \frac{\widetilde{C}\widetilde{D}}{\left(\widetilde{C}+\widetilde{D}\right)^2\left(\widetilde{C}+\widetilde{D}+1\right)}.$$

We set r.v. $\widetilde{\rho}$ to have as mean the Pearson correlation coefficient observed on the most recent run-off triangles and relative variability equal to the coefficient of variation calculated on the vector of the Pearson correlation coefficients related to the l historical triangles.
Therefore, by calculating the values of $E\left(\widetilde{\rho}\right)$ and $\text{Var}\left(\widetilde{\rho}\right)$, the method of moments allows us to define the values of $\widetilde{C}$ and $\widetilde{D}$, that we adopt as estimates of the expected values of prior distributions. On the other hand, for the relative variability, set identical for both the priors, we use an *ad hoc* value equals to 10%. Hence, hyperparameters are estimated as,

$${}_1\varepsilon = \frac{1}{CV\left(\widetilde{C}\right)^2} \quad {}_2\varepsilon = \frac{{}_1\varepsilon}{E\left(\widetilde{C}\right)}, \qquad {}_1\psi = \frac{1}{CV\left(\widetilde{D}\right)^2} \quad {}_2\psi = \frac{{}_1\psi}{E\left(\widetilde{D}\right)}.$$

When exponential distributions are adopted as priors, $f(C)$ and $f(D)$ turn out to be:

$$\widetilde{C} \sim \text{Exp}\left(\varphi\right) \qquad \widetilde{D} \sim \text{Exp}\left(\omega\right).$$

Hyperparameters, φ and ω, are derived, as in the previous case, according to the method of moments by the values of $E\left(\widetilde{\rho}\right)$ and $\operatorname{Var}\left(\widetilde{\rho}\right)$:

$$\varphi = \frac{1}{E\left(\widetilde{C}\right)} \qquad \omega = \frac{1}{E\left(\widetilde{D}\right)}.$$

It is to be noted that *ad hoc* assumptions related to the variability are not needed to estimate prior distributions under the exponential distribution.
Below are the results of the previous Bayesian approach adopted to estimate correlation between structure variables, concerning the two insurers introduced in Section 3. As usual, the analyses are based on 10,000 iterations carried out with the dependent bootstrap technique.
Table 7 exhibits values of Pearson correlation coefficient computed on the 9 historical triangles via the Mack's formula and dependent bootstrap approach. The linear correlation of the small insurer, DELTA, is included between 0.014 and 0.333, whereas OMEGA shows values between 0.075 and 0.415.

Pearson Correlation Coefficient

Dimension	4×4	5×5	6×6	7×7	8×8	9×9	10×10	11×11	12×12
DELTA	0.231	0.333	0.127	0.014	0.098	0.065	0.154	0.194	0.087
OMEGA	0.289	0.174	0.220	0.415	0.289	0.324	0.175	0.106	0.075

Table 7: Pearson correlation coefficient for both insurers, DELTA and OMEGA, between r.v.s $\widetilde{\sigma}_{\widetilde{q}}$ and $\widetilde{\sigma}_{\widetilde{p}}$ related to the historical triangles with dimension from 4×4 to 12×12.

In Appendix D we report the values of hyperparameters of both gamma and exponential distributions. The posterior is achieved via Monte Carlo method through Metropolis-Hasting algorithm; the mean of the posterior (see Appendix D) is used to assess parameters of the r.v. $\widetilde{\rho}$ as shown in formula (4.2). Finally, we adopt the expected value of $\widetilde{\rho}$ as an estimate of the Pearson correlation coefficient between structure variables $\widetilde{q}$ and $\widetilde{p}$. Table 8 reports the correlation between structural risk factors estimated under the two distribution assumptions adopted for priors; for OMEGA insurer the posterior expected value of $\widetilde{\rho}$ results to be almost identical under the two type of priors.

Estimated expected values of $\widetilde{\rho}$

Insurer	Type of prior	Gamma	Exponential
DELTA	$E\left(\widetilde{\rho}\right)$	0.101	0.141
OMEGA	$E\left(\widetilde{\rho}\right)$	0.138	0.138

Table 8: Estimated Pearson correlation coefficient between the structure variables for both insurers, DELTA and OMEGA.

Concerning the Collective Risk Model, the structure variables are modeled with a two-dimensional meta-Gaussian distribution, where a Gaussian copula, with parameter the Pearson correlation coefficient estimated as shown above, joins the two marginals of $\widetilde{q}$ and $\widetilde{p}$ calibrated as explained in the previous section.

5 Case study

The estimates related to structure variables acquired in Sections 3 and 4 are deployed here into the Collective Risk Model in order to evaluate the claims reserve distribution concerning both a total run-off and a one-year time horizon. By adapting the re-reserving method, we obtain the "one-year" reserve distribution of insurer obligations. Reserve risk is assessed by calculating the Solvency Capital Requirement (SCR) as the difference between the quantile at 99.5% confidence level of the

distribution of the insurer obligations at the end of the next accounting year, opportunely discounted at time zero, and the best estimate at present time.
As explained in Section 1, model parameters related to claim size and claim count are estimated through the deterministic Frequency-Severity method; moreover, to calibrate cumulants of severity we consider the variability coefficient of claim cost for each development year and we assume that $\widetilde{Z}_{i,j}$ follows a gamma distribution in each cell of the triangle.
The deterministic method leads DELTA and OMEGA to a claims reserve of approximately 228 and 2,807 million Euro; these values match the expected values (best estimates) attained with the CRM.
The analyses shown below are based on 100,000 simulations; moreover, model parameters acquired via Bayesian approaches are based only on gamma priors. Under the assumption of uncorrelated structure variables, we verify that simulated moments of the claims reserve are close to the exact ones, proving that the number of simulations is adequate.
Table 9 reports the coefficient of variation and skewness of the claims reserve evaluated both under a total run-off and a one-year time horizon, assuming $\rho\left(\widetilde{q},\widetilde{p}\right)=0$.

Insurer	Time horizon	Coeff. of Var.	Skewness
DELTA	**Tot. run-off**	5.73%	0.131
	One-year	5.13%	0.222
OMEGA	**Tot. run-off**	4.23%	0.102
	One-year	3.05%	0.137

Table 9: Coefficient of variation and skewness of claims reserve assessed under total run-off and one-year time horizon for both insurers, DELTA and OMEGA, under the assumption of no correlation between $\widetilde{q}$ and $\widetilde{p}$.

The relative variability of the reserve assessed under a total run-off time horizon is higher compared to the one-year time horizon for both insurers; on the other hand, the claims reserve is more skewed under a one-year evaluation. It is to be noted that the coefficient of variation of the one-year reserve, compared to the total run-off value, is approximately around the 90% and 70% for DELTA and OMEGA respectively.
Comparing the two insurers, the coefficient of variation is lower for OMEGA than for DELTA, due to a bigger number of reserved claims that leads to a higher diversification of pooling risk, namely the variability not ascribable to structure variables. Similarly, the reserve of the bigger insurer is less skewed in respect of the obligations distribution of DELTA.
Table 10 refers to the one-year claims reserve and gives the quantile and Tail VaR at level 99.5% and 99% respectively; moreover, the Solvency Capital Requirement and its ratio respect to the best estimate (the so-called SCR ratio) are reported.

Insurer	$q_{99.5\%}$ *	$\text{TVaR}_{99\%}$ *	SCR *	SCR ratio
DELTA	261,902	263,068	33,340	14.59%
OMEGA	3,042,979	3,052,341	237,522	8.47%

* Amounts expressed in thousands of Euro.

Table 10: One-year reserve: quantile and TVaR at level 99.5% and 99% respectively, SCR and SCR ratio under the assumption of no correlation between $\widetilde{q}$ and $\widetilde{p}$.

OMEGA shows a smaller SCR ratio than DELTA, due to lower values of both relative variability and skewness.
In what follows, we investigate the impact that dependence between structure variables has on claims reserve; in addition to the correlation value estimated via the Bayesian approach, we impose perfect negative and positive linear correlations between $\widetilde{q}$ and $\widetilde{p}$.
When considering the claims reserve evaluated according to a total run-off time horizon, under the

assumption of no correlation between the structural risk factors, it is possible to calculate the coefficient of variation of the reserve in respect of the average Pearson correlation coefficient ($\bar{\rho}$) affecting the cells of the lower triangle,

$$\mathrm{Var}\left(\widetilde{R}\right) = \left[\sum_{i,j\in B} \mathrm{Var}\left(\widetilde{X}_{i,j}\right)\right]\left(1-\bar{\rho}\right) + \bar{\rho}\left[\sum_{i,j\in B} SD\left(\widetilde{X}_{i,j}\right)\right]^2,$$

where, to simplify the notation, $B = \left\{\widetilde{X}_{i,j}; i + j > N+1\right\}$ identifies cells of the lower run-off triangle. It is worth noting that the average Pearson correlation coefficient affecting $n \geq 3$ random variables has the lower bound (see [35]):

$$\bar{\rho}_{min} \geq -\frac{1}{n-1}.$$

In the triangle of dimension 12×12 the number of lower cells is 66: this leads to a theoretical value of $\bar{\rho}_{min}$ equal to -0.015.

Figures 1 and 2 exhibit the shape of the relative variability of the reserve as function of the average Pearson correlation coefficient acting on the lower triangle. It is worth emphasizing that insurer DELTA shows higher values of relative variability being equal values of $\bar{\rho}$.

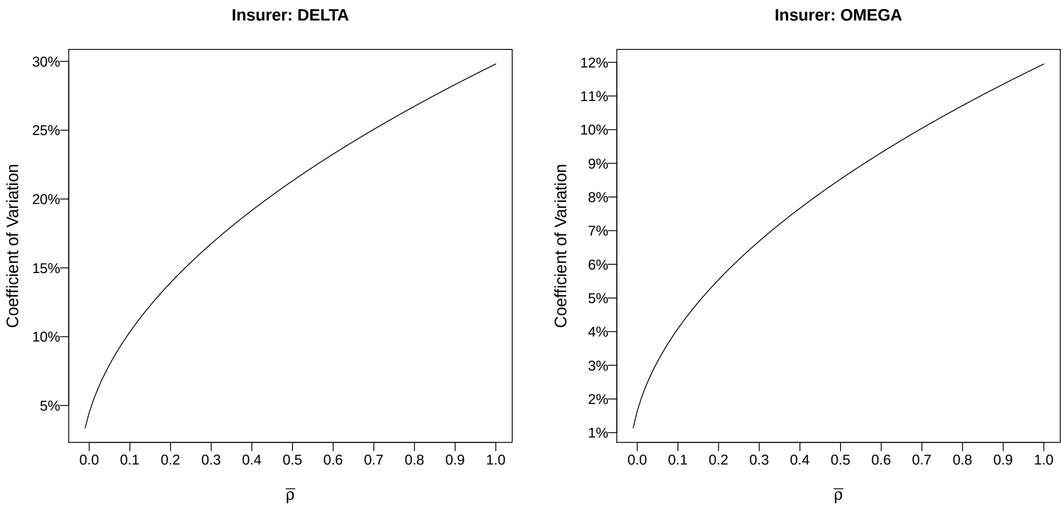

Figure 1: Coefficient of variation of claims reserve in respect of the average Pearson correlation coefficient for both insurers, DELTA (left) and OMEGA (right).

Later, through simulation, the coefficient of variation of the reserve is calculated using the correlation between structure variables estimated via the Bayesian approach, and values ± 1. This allows us to indirectly quantify the equivalent average Pearson correlation coefficient induced in the cells of the triangle under the assumption of no correlation between the r.v.s $\widetilde{q}$ and $\widetilde{p}$.

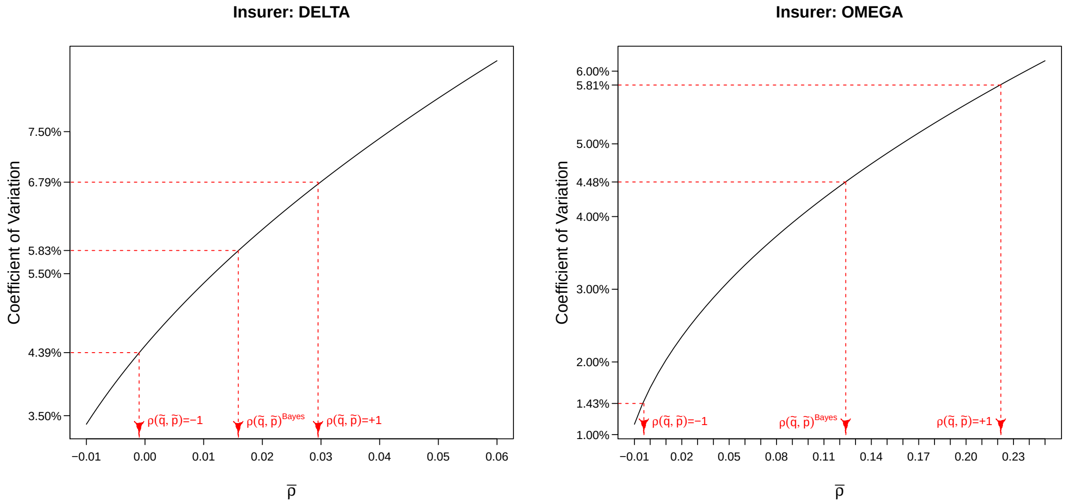

Figure 2: Average Pearson correlation coefficient induced by the dependence between structure variables, DELTA (left) and OMEGA (right).

The impact that the dependence between structure variables has on $\bar{\rho}$ is higher for the larger insurer: indeed the claims reserve distribution of OMEGA is affected mainly by the structure variables, due to its higher number of reserved claims which allows the pooling risk to be almost entirely diversified. According to the previously considered values of Pearson correlation between $\tilde{q}$ and $\tilde{p}$, tables 11 and 12 compare the characteristics of the total run-off and a one-year reserve.

Insurer: DELTA

Pearson correlation coeff.	Time horizon	Coeff. of Var.	Skewness
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = -1}$	**Tot. run-off**	4.39%	0.102
	One-year	4.43%	0.218
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = 0.101}$	**Tot. run-off**	5.83%	0.128
	One-year	5.19%	0.209
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = +1}$	**Tot. run-off**	6.79%	0.144
	One-year	5.76%	0.227

Table 11: Insurer DELTA - Coefficient of variation and skewness for both total run-off and one-year claims reserve for different levels of dependence between $\tilde{q}$ and $\tilde{p}$.

Insurer: OMEGA

Pearson correlation coeff.	Time horizon	Coeff. of Var.	Skewness
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = -1}$	**Tot. run-off**	1.43%	0.032
	One-year	1.62%	0.089
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = 0.138}$	**Tot. run-off**	4.48%	0.112
	One-year	3.21%	0.134
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = +1}$	**Tot. run-off**	5.81%	0.172
	One-year	4.00%	0.194

Table 12: Insurer OMEGA - Coefficient of variation and skewness for both total run-off and one-year claims reserve for different levels of dependence between $\tilde{q}$ and $\tilde{p}$.

Moreover, taking into consideration the distribution of insurer obligations at the end of the next accounting year, table 13 and 14 report some risk measures (i.e. VaR and Tail VaR at 99.5% and 99%

confidence level respectively), the Solvency Capital Requirement, the SCR ratio and the plots of the claims development result distribution. Figure 3 depicts the distribution of the Claims Development Result for both insurers.

Insurer: DELTA

Pearson correlation coeff.	$\mathbf{q_{99.5\%}}$ *	$\mathbf{TVaR_{99\%}}$ *	SCR *	SCR ratio
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = -1}$	256,731	257,854	28,309	12.39%
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = 0.101}$	261,643	262,943	33,052	14.46%
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = +1}$	265,522	267,400	36,861	16.12%

* Amounts expressed in thousands of Euro.

Table 13: Insurer DELTA - One-year reserve: quantile and TVaR at level 99.5% and 99% respectively, SCR and SCR ratio for different levels of dependence between structure variables.

Insurer: OMEGA

Pearson correlation coeff.	$\mathbf{q_{99.5\%}}$ *	$\mathbf{TVaR_{99\%}}$ *	SCR *	SCR ratio
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = -1}$	2,925,433	2,929,865	121,780	4.34%
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = 0.138}$	3,055,208	3,066,441	249,328	8.89%
$\boldsymbol{\rho(\tilde{q},\tilde{p}) = +1}$	3,133,079	3,150,322	325,889	11.61%

* Amounts expressed in thousands of Euro.

Table 14: Insurer OMEGA - One-year reserve: quantile and TVaR at level 99.5% and 99% respectively, SCR and SCR ratio for different levels of dependence between structure variables.

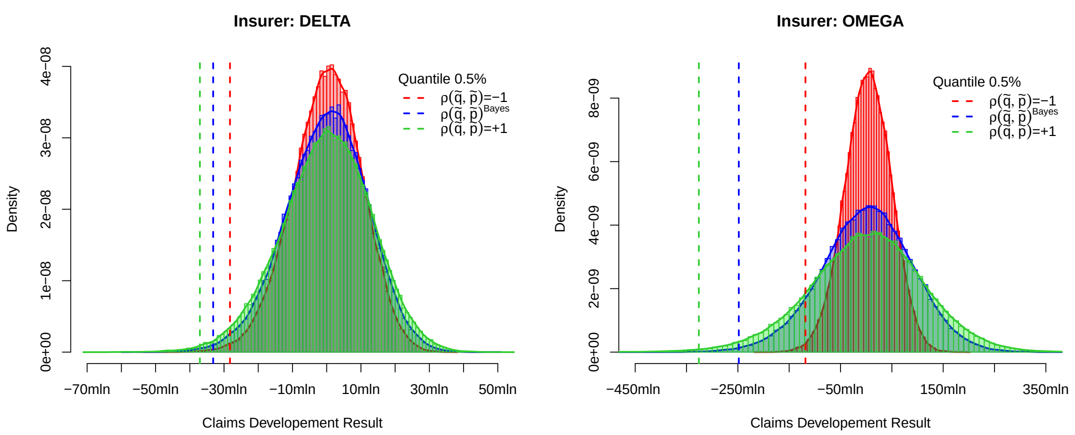

Figure 3: One-year time horizon: distribution of the Claims Development Result for both insurers, DELTA (left) and OMEGA (right).

It is to be pointed out that for both insurers a value of linear correlation between $\tilde{q}$ and $\tilde{p}$ equal to -1 leads to a particular situation where the coefficient of variation of the one-year reserve is higher than the relative variability of the total run-off distribution. The ratio between the coefficients of variation of the one-year reserve and the total run-off reserve are 100.91% and 113.31% for DELTA and OMEGA respectively.

To investigate this unique circumstance we focus on the variance of the reserve evaluated under the two time horizons (since the one-year and the total run-off distributions have the same mean) expressed in terms of the average Pearson correlation coefficient:

$$
\begin{aligned}
&\sum_{i,j\in B}\mathrm{Var}\left(\widetilde{X}^{OY}_{i,j}\right)+\bar{\rho}^{OY}\sum_{i,j\in B}\sum_{\substack{h,k\in B\\(h\neq i\vee k\neq j)}}SD\left(\widetilde{X}^{OY}_{i,j}\right)SD\left(\widetilde{X}^{OY}_{h,k}\right)\\
>&\sum_{i,j\in B}\mathrm{Var}\left(\widetilde{X}^{Tot}_{i,j}\right)+\bar{\rho}^{Tot}\sum_{i,j\in B}\sum_{\substack{h,k\in B\\(h\neq i\vee k\neq j)}}SD\left(\widetilde{X}^{Tot}_{i,j}\right)SD\left(\widetilde{X}^{Tot}_{h,k}\right)
\end{aligned}
\tag{5.1}
$$

The left side of the previous equation refers to the variance of the one-year reserve, whereas the right side refers to the total run-off reserve (here superscripts "OY" and "Tot" indicate the one-year and the total run-off time horizon).
The terms $\sum_{i,j\in B}\mathrm{Var}\left(\widetilde{X}^{OY}_{i,j}\right)$ and $\sum_{i,j\in B}\mathrm{Var}\left(\widetilde{X}^{Tot}_{i,j}\right)$ include the variance of each single cell of the lower triangle; the variances relative to the first diagonal (the forthcoming accounting year) match by construction in the total run-off and in the one-year evaluations, thus it can be neglected. We show by simulation that the variability of each single cell of the one-year reserve is lower than the one of the total run-off reserve. Therefore, the inequality of formula (5.1) is imputable to the covariance terms; hence, the average Pearson correlation coefficient of the one-year reserve exceeds the one related to a total run-off time horizon.
The modeling explanation of this circumstance is ascribable to the use of two different approaches to assess the reserve under a total run-off and a one-year time horizon. When the whole lifetime of obligations is considered, we adopt the CRM, whereas when only the next 12 months are taken into account, the reserve is determined according to the re-reserving approach. The latter imposes the calculation, at each simulation step, of the first diagonal that it is adopted to estimate, in line with the underling deterministic model, the lower residual cells of the triangle. With regards to the Frequency-Severity method, at each iteration, according to the first simulated diagonal, the one-year approach calculates the Chain-Ladder development factors to estimate the remaining lower triangle. Therefore, by construction, the re-reserving approach induces a not negligible dependence between cells of the lower triangle; the average Pearson correlation coefficient affecting the triangle for the one-year reserve is higher than the one for the total run-off view, which, in our case study, is almost zero due to the perfect negative dependence between $\widetilde{q}$ and $\widetilde{p}$.
The logical explanation related to a higher variability under a one-year time horizon evaluation, in respect of the total run-off, lies of course in the different time horizon taken into consideration. When we consider the whole lifetime of obligations, the random variables related to the cells of the lower triangle tend to compensate each other, especially thanks to the negative dependence of the structure variables, thus reducing the reserve variability. On the other hand, when the reserve is evaluated taking only the next 12 months into account, the random variables are not able to offset one another as much as they do over their whole lifetime, leading to a higher variability in respect of the one related to the total run-off time horizon. Indeed, the one-year evaluation of the reserve disregards the stochastic claims process over the next 12 months, not allowing the matching among the r.v.s $\widetilde{X}_{i,j}$, driven by the negative dependence between $\widetilde{q}$ and $\widetilde{p}$, to get completely displayed.
Moreover, under the Frequency-Severity method based on Chain-Ladder mechanics, the re-reserving approach stresses this particular case when the triangle is small due to the higher impact that the first simulated diagonal has on the residual lower triangle. In table 15 we report the Pearson correlation coefficient estimated by simulation between the first diagonal of the lower triangle and the residual cells.

Dimension		4 × 4	5 × 5	6 × 6	7 × 7	8 × 8	9 × 9	10 × 10	11 × 11	12 × 12
DELTA	**Tot. run-off**	0.009	0.015	0.016	0.023	0.018	0.021	0.019	0.017	0.018
	One-year	0.854	0.819	0.758	0.776	0.647	0.643	0.688	0.740	0.587
OMEGA	**Tot. run-off**	-0.001	-0.001	0.002	0.004	0.003	0.000	0.008	0.004	0.008
	One-year	0.890	0.841	0.803	0.774	0.712	0.722	0.742	0.685	0.629

Table 15: Pearson correlation coefficient for both insurers, DELTA and OMEGA, between first diagonal and the remaining lower cells of triangle.

With regards to the different dimensions of the run-off triangle, table 16 shows the SCR ratio and the ratio between the coefficient of variation of the reserve evaluated under a one-year time horizon and the one related to the total run-off reserve.

Dimension		4 × 4	5 × 5	6 × 6	7 × 7	8 × 8	9 × 9	10 × 10	11 × 11	12 × 12
DELTA	CV_{OY}/CV_{Tot}	111.92%	108.90%	108.60%	103.17%	107.21%	103.96%	97.25%	92.61%	100.91%
	SCR ratio	16.02%	14.65%	14.39%	13.25%	15.30%	14.64%	13.24%	12.40%	12.39%
OMEGA	CV_{OY}/CV_{Tot}	113.90%	113.95%	111.93%	110.87%	117.38%	111.89%	104.43%	105.66%	113.31%
	SCR ratio	5.94%	5.78%	5.59%	5.30%	5.99%	5.30%	4.56%	4.49%	4.34%

Table 16: Coefficient of variation of the one-year reserve in respect of the one related to the total run-off reserve and SCR ratio for both insurers, DELTA and OMEGA.

As expected, the ratio between the coefficients of variation is higher for triangles with low dimension due to the higher correlation induced by the re-reserving approach. When the triangle dimension increases, the weight of the one-year coefficient of variation over the total run-off relative variability tends to decrease, but not necessarily in a monotonic way. Also the SCR ratio exhibits, in general, a decreasing trend in respect of the triangle dimension caused by the increasing number of reserved claims that highlights the diversification effect. If we compare the two insurers, the ratio between the coefficient of variation is higher for OMEGA, due to the greater impact that structure variables have on the total run-off reserve. On the other hand, the SCR ratio results to be lower for OMEGA because of the greater number of reserved claims that allows the insurer to diversify mainly the component of pooling risk.
In sum, the coefficient of variation of the one-year reserve in respect of the one assessed under a total run-off time horizon depends on three factors: the run-off triangle dimension, namely the impact that the future diagonal has on the remaining cells to be estimated, the level of dependence between structure variables and, in general, the characteristics of the data set. It is to be pointed out that from a mathematical point of view, it is not possible *a priori* to know the direction of the inequality between the coefficients of variation of the one-year and total run off reserve in respect of the value of correlation between $\widetilde{q}$ and $\widetilde{p}$.

6 Conclusions

As shown in [27] the estimation of structure variables embedded into the Collective Risk Model to stochastically evaluate the claims reserve is a key issue. In the present work, we developed a Bayesian approach to quantify the variability of structure variables implementing the Bayes' rule through empirical prior distributions and data obtained by applying a bootstrapping-based procedure to run-off triangles integrated via Mack's formula. In addition, the dependence between structural risk factors has been investigated in a Bayesian manner: we proposed a joined resampling scheme aimed at capturing the inherent dependency of data. Through a case study we showed the impact that the dependence between structure variables has on claims reserve distribution, evaluated with respect to both the entire liability settlement period, the so-called total run-off approach, and the one-year

time horizon, in order to assess the reserve risk capital requirement. When perfect negative linear dependence is addressed on structural risk factors, we come across a unique situation where the coefficient of variation of the one-year reserve exceeds the relative variability of the total run-off reserve. Starting from this circumstance, we analysed both the modeling connection and the logical link between the coefficient of variation of reserve appraised under the two time horizons.
In our opinion, the methodology developed in the present paper, making use of only historical data, allows us to estimate the magnitude and the dependence between structure variables both avoiding any sort of expert judgment and providing a coherent approach with the Collective Risk Model to assess structural risk factors. Nevertheless, the use of Bayes' rule based on a selected parametric model turns out to be a hard to prove assumption; further developments may consider likelihood-free frameworks, where the statistical model is defined in terms of a stochastic generating mechanism of data.

References

[1] F. Beens, L. Bui, S. Collings, and A. Gill. Stochastic reserving using Bayesian models - can it add value? *Presented to the Institute of Actuaries of Australia 17th General Insurance Seminar*, 2010.

[2] J. B. Bernardo. Bayesian Statistics. In R. Viertl, editor, *Probability and Statistics, Encyclopedia of Life Support Systems*. Eolss publisher, Oxford, UK, 2003.

[3] S. Brooks, A. Gelman, G. Jones, and X. Meng. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC press, New York, USA, 2011.

[4] B.P. Carlin and T.A. Louis. *Bayesian Methods for Data Analysis, Third Edition*. Chapman and Hall/CRC Texts in Statistical Science. CRC, 3 edition, 2009.

[5] G. Casella. An introduction to empirical bayes data analysis. *The American Statistician*, 39(2):83–87, 1985.

[6] C. Chivers. *MHadaptive: General Markov Chain Monte Carlo for Bayesian Inference using adaptive Metropolis-Hastings sampling*, 2012. R package version 1.1-8.

[7] D. Diers. Stochastic re-reserving in multi-year internal models - An approach based on simulations. *ASTIN Colloquium, Helsinki*, 2009.

[8] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. Chapman and Hall/CRC press, New York, USA, 6th edition, 1994.

[9] P. England and R. J. Verrall. Stochastic Claims reserving in General Insurance. *British actuarial Journal*, 8(3):443–518, 2002.

[10] P. England and R. J. Verrall. Predictive distribution of outstanding liabilities in general insurance. *Annals of Actuarial Science*, 1(2):221–270, 2006.

[11] P. England, R. J. Verrall, and M. V. Wüthrich. Bayesian Overdispersed Poisson Model and the Bornhuetter-Ferguson Claims Reserving Method. *Annals of Actuarial Science*, 6(2):258–283, 2012.

[12] European Commission. Directive 2009/138/EC "Solvency II", 2009. `https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32009L0138`.

[13] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC press, New York, USA, 3th edition, 2013.

[14] International Actuarial Association. A Global Framework for Insurer Solvency Assessment. *Research Report of The Insurer Solvency Assessment Working Party*, 2004.

[15] P. Jindrová and L. Kopecká. Empirical bayes credibility models for economic catastrophic losses by regions. *ITM Web Conf.*, 9:01006, 2017.

[16] S.A. Klugman. *Bayesian Statistics in Actuarial Science: With Emphasis on Credibility*, volume 15 of *Huebner International Series on Risk, Insurance and Economic Security*. Kluwer Academic Publishers, 1992.

[17] S. M. Lynch. *Introduction to Applied Bayesian Statistics and estimation for Social Scientists*. Springer-Verlag, New York, USA, 1st edition, 2007.

[18] T. Mack. Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin*, 23:213–225, 1993.

[19] T. Mack. Measuring the variability of Chain-Ladder reserve estimates. *Casualty Actuarial Society Forum*, 1994.

[20] T. Mack. The standard error of Chain-Ladder reserve estimates: recursive calculation and inclusion of a tail factor. *Astin Bulletin*, 29(2):361–366, 1999.

[21] G. Meyers. Stochastic loss reserving with the Collective Risk Model. *Casualty Actuarial Society*, 3(2):239–269, 2009.

[22] G. Meyers. Stochastic loss reserving using Bayesian MCMC models. *Casualty Actuarial Society*, Monograph series(1):1–64, 2015.

[23] R. Norberg. Empirical bayes credibility. *Scandinavian Actuarial Journal*, 1980(4):177–194, 1980.

[24] E. Ohlsonn and J. Lauzeningks. The one-year non-life insurance risk. *Insurance: Mathematics and Economics*, 45(2):203 – 208, 2009.

[25] G. W. Peters, M. V. Wüthrich, and P. V. Shevchenko. Chain Ladder Method: Bayesian Bootstrap versus Classical Bootstrap. *Insurance: Mathematics and Economics*, 47:36–51, 2010.

[26] S. Petrone, S. Rizzelli, J. Rousseau, and C. Scricciolo. Empirical bayes methods in classical and bayesian inference. *METRON*, 72(2):201–215, 2014.

[27] A. Ricotta and G. P. Clemente. An Extension of Collective Risk Model for Stochastic Claim Reserving. *Journal of Applied Finance & Banking*, 6(5):45–62, 2016.

[28] J. R. Sanchez, R. Verrall, and J. L. Vilar. Bayesian and Credibility Estimation for the Chain Ladder Reserving Method. *Istituto de Actuarios Españoles, Anales*, 17:51–74, 2011.

[29] N. Savelli and G. P. Clemente. A Collective Risk Model for Claims Reserve Distribution. In *Atti del XVI Convegno di Teoria del Rischio (Campobasso, 18-18 September 2009)*, pages 59–87. Aracne, Campobasso, 2010.

[30] N. Savelli and G. P. Clemente. Stochastic claim reserving based on CRM for Solvency II purposes. *ASTIN Colloquium, Madrid*, 2011.

[31] D. P. M. Scollnik. Actuarial Modeling with MCMC and BUGs. *North American Actuarial Journal*, 5(2):96–124, 2001.

[32] D. P. M. Scollnik. Bayesian Reserving Models Inspired by Chain Ladder Methods and Implemented Using WinBUGS. *Society of Actuaries*, 2004.

[33] L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22(4):1701–1762, 1994.

[34] R.J. Verrall. Bayes and empirical bayes estimation for the chain ladder model. *ASTIN Bulletin: The Journal of the International Actuarial Association*, 20(02):217–243, 1990.

[35] T. D. Wickens. *The Geometry of Multivariate Statistics*. Lawrence Erlbaum Associates, Publishers Hillsdale, New Jersey, USA, 1995.

[36] M. V. Wüthrich. Using a Bayesian approach for claims reserving. *Variance*, 1:292–301, 2007.

[37] M. V. Wüthrich and M. Merz. *Stochastic claims reserving methods in insurance*. Wiley finance series. John Wiley & Sons, New Jersey, USA, 2008.

[38] M. V. Wüthrich and M. Merz. Stochastic Claims Reserving Manual: Advances in Dynamic Modeling. *Swiss Finance Institute Research Paper No. 15-34*, 2015.

Appendix A

Insurer: DELTA

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12
1993	38,364	76,319	91,669	97,769	100,947	103,649	105,152	106,513	107,521	108,419	108,707	110,502
1994	41,475	85,940	101,878	108,718	112,018	114,748	115,757	116,909	117,675	118,143	118,599	
1995	46,520	94,099	109,195	116,103	119,495	120,885	122,223	123,409	124,331	124,890		
1996	47,925	99,792	117,390	123,696	126,570	128,694	130,927	132,135	133,008			
1997	51,420	103,505	120,796	126,816	129,536	132,573	133,893	135,017				
1998	57,586	111,736	131,345	138,876	142,986	145,766	148,032					
1999	55,930	110,872	131,819	142,318	148,182	151,495						
2000	51,005	104,197	126,016	134,380	139,095							
2001	51,693	103,266	121,933	130,767								
2002	54,954	106,565	125,169									
2003	59,763	113,506										
2004	60,361											

Table 17: Insurer DELTA - Cumulative paid amounts (thousands of Euro).

Insurer: DELTA

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12
1993	34,433	48,229	49,819	50,387	50,665	50,816	50,920	50,975	51,004	51,036	51,053	51,108
1994	35,475	49,193	50,693	51,241	51,451	51,584	51,635	51,679	51,706	51,726	51,742	
1995	37,004	50,824	52,351	52,788	52,981	53,053	53,099	53,128	53,145	53,162		
1996	37,038	50,669	52,131	52,631	52,795	52,875	52,938	52,977	53,003			
1997	36,849	50,265	51,828	52,251	52,432	52,539	52,620	52,662				
1998	39,171	51,772	53,364	53,924	54,197	54,372	54,527					
1999	37,492	49,774	51,831	52,571	52,961	53,248						
2000	34,188	46,433	48,371	49,132	49,427							
2001	31,308	42,051	43,958	44,597								
2002	30,357	40,474	42,085									
2003	30,717	41,799										
2004	30,590											

Table 18: Insurer DELTA - Cumulative number of paid claims.

Insurer: OMEGA

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12
1993	193,474	366,091	453,292	499,090	528,858	548,653	568,435	585,750	599,122	611,674	620,504	648,446
1994	199,854	368,820	449,363	490,019	519,072	540,193	560,158	574,406	585,126	598,810	604,818	
1995	225,578	412,343	505,692	553,301	584,272	610,562	628,183	646,594	661,255	668,846		
1996	256,398	493,076	598,692	649,864	687,202	711,287	732,041	744,122	758,260			
1997	282,956	546,152	666,535	730,224	767,444	796,684	819,804	835,313				
1998	292,428	576,829	718,229	774,619	814,814	842,770	872,756					
1999	312,350	597,857	729,544	804,796	851,345	890,076						
2000	327,673	635,665	797,182	875,147	927,842							
2001	339,899	666,179	852,090	953,363								
2002	371,275	757,122	950,128									
2003	388,025	778,762										
2004	398,686											

Table 19: Insurer OMEGA - Cumulative paid amounts (thousands of Euro).

Insurer: OMEGA

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12
1993	284,236	395,010	407,622	412,348	414,554	415,754	416,509	417,012	417,415	417,820	418,070	418,767
1994	274,524	367,933	379,865	383,968	385,970	387,096	387,811	388,278	388,747	389,065	389,285	
1995	284,017	374,422	386,345	390,312	392,114	393,119	393,780	394,370	394,770	395,011		
1996	299,605	400,289	412,734	416,404	417,964	418,849	419,473	419,855	420,157			
1997	308,092	411,618	426,850	431,376	433,420	434,608	435,384	435,926				
1998	295,813	391,103	405,954	410,705	413,070	414,337	415,277					
1999	288,418	381,906	399,078	405,364	408,304	410,063						
2000	285,940	387,422	407,558	415,654	419,682							
2001	290,023	386,767	409,717	418,962								
2002	280,008	374,305	396,095									
2003	277,412	369,192										
2004	252,239											

Table 20: Insurer OMEGA - Cumulative number of paid claims.

Appendix B

Insurer: DELTA

Type of r.v.	**Gamma**				**Exponential**	
Priors	$f(A)$		$f(B)$		$f(A)$	$f(B)$
Parameters	${}_1\alpha$	${}_2\alpha$	${}_1\beta$	${}_2\beta$	λ	η
$\tilde{\sigma}_{\tilde{q}}$	28.464	0.847	28.464	0.019	0.030	0.001
$\tilde{\sigma}_{\tilde{p}}$	5.948	0.243	5.948	0.006	0.041	0.001

Table 21: Insurer DELTA - Hyperparameters of gamma and exponential prior.

Insurer: OMEGA

Type of r.v.	**Gamma**				**Exponential**	
Priors	$f(A)$		$f(B)$		$f(A)$	$f(B)$
Parameters	${}_1\alpha$	${}_2\alpha$	${}_1\beta$	${}_2\beta$	λ	η
$\tilde{\sigma}_{\tilde{q}}$	4.796	0.246	4.796	0.007	0.051	0.002
$\tilde{\sigma}_{\tilde{p}}$	24.368	1.525	24.368	0.036	0.063	0.001

Table 22: Insurer OMEGA - Hyperparameters of gamma and exponential prior.

Insurer: DELTA

Type of prior	**Gamma**	**Exponential**
$E\left(\tilde{A}_{\tilde{q}} \mid \tilde{\sigma}_{\tilde{q}}\right)$	33.432	39.200
$E\left(\tilde{B}_{\tilde{q}} \mid \tilde{\sigma}_{\tilde{q}}\right)$	1,545.619	1,817.945
$E\left(\tilde{A}_{\tilde{p}} \mid \tilde{\sigma}_{\tilde{p}}\right)$	27.771	23.250
$E\left(\tilde{B}_{\tilde{p}} \mid \tilde{\sigma}_{\tilde{p}}\right)$	855.290	769.017

Table 23: Insurer DELTA - Expected values of posterior distributions.

Insurer: OMEGA

Type of prior	**Gamma**	**Exponential**
$\boldsymbol{E\left(\widetilde{A}_{\widetilde{q}} \mid \widetilde{\sigma}_{\widetilde{q}}\right)}$	15.898	13.320
$\boldsymbol{E\left(\widetilde{B}_{\widetilde{q}} \mid \widetilde{\sigma}_{\widetilde{q}}\right)}$	571.847	478.333
$\boldsymbol{E\left(\widetilde{A}_{\widetilde{p}} \mid \widetilde{\sigma}_{\widetilde{p}}\right)}$	18.183	36.388
$\boldsymbol{E\left(\widetilde{B}_{\widetilde{p}} \mid \widetilde{\sigma}_{\widetilde{p}}\right)}$	677.024	1,349.514

Table 24: Insurer OMEGA - Expected values of posterior distributions.

Appendix C

Let $\widetilde{X}$ be a generic random variable following a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$. The density function is:

$$f_{\widetilde{X}}(X) = \frac{\beta^{\alpha} x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}, \qquad x \in \mathbb{R}^{+},$$

where $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$ is the gamma function.
The j-th moment about zero is

$$E\left(\widetilde{X}^{j}\right) = \frac{\alpha(\alpha+1)\dots(\alpha+j-1)}{\beta^{j}},$$

whereas the j-th cumulant is

$$K_j\left(\widetilde{X}\right) = \frac{(j-1)!\alpha}{\beta^{j}}.$$

We here compute the characteristics of the structure variable $\widetilde{q}$ (similar considerations hold for $\widetilde{p}$). The structure variable related to the claim count follows a gamma distribution with same parameters:

$$\widetilde{q} \sim \text{Gamma}\left(\frac{1}{\widetilde{\sigma}^{2}_{\widetilde{q}}}; \frac{1}{\widetilde{\sigma}^{2}_{\widetilde{q}}}\right),$$

where $\widetilde{\sigma}^{2}_{\widetilde{q}}$ is itself a random variable. In general it is possible to compute the moments of the structure variable $\widetilde{q}$ without knowing the distributional form of the r.v. $\widetilde{\sigma}_{\widetilde{q}}$.
The expected value is given by

$$E\left(\widetilde{q}\right) = E_{\widetilde{\sigma}_{\widetilde{q}}}\left[E_{\widetilde{q}}\left(\widetilde{q} \mid \widetilde{\sigma}_{\widetilde{q}}\right)\right] = E_{\widetilde{\sigma}_{\widetilde{q}}}\left[\frac{\widetilde{\sigma}^{2}_{\widetilde{q}}}{\widetilde{\sigma}^{2}_{\widetilde{q}}}\right] = 1,$$

therefore, the mean of the structure variable remains equal to 1.
With regards to the variance, it can be compute as:

$$\begin{aligned}\text{Var}\left(\widetilde{q}\right) =& E_{\widetilde{\sigma}_{\widetilde{q}}}\left[\text{Var}_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{q} \mid \widetilde{\sigma}_{\widetilde{q}}\right)\right] + \text{Var}_{\widetilde{\sigma}_{\widetilde{q}}}\left[E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{q} \mid \widetilde{\sigma}_{\widetilde{q}}\right)\right]\\ & E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{\sigma}^{2}_{\widetilde{q}}\right) + \text{Var}_{\widetilde{\sigma}_{\widetilde{q}}}(1) = E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{\sigma}^{2}_{\widetilde{q}}\right),\end{aligned}$$

showing that the second raw moment of the r.v. $\widetilde{\sigma}_{\widetilde{q}}$ represents the variance of $\widetilde{q}$.
The relative variability is equal to the squared root of the variance:

$$CV\left(\widetilde{q}\right) = \sqrt{E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{\sigma}^{2}_{\widetilde{q}}\right)}.$$

Finally, the skewness equals:

$$\gamma(\widetilde{q}) = \frac{\mu_3(\widetilde{q})}{[\mathrm{Var}(\widetilde{q})]^{\frac{3}{2}}} = \frac{E(\widetilde{q}^3) - 3E(\widetilde{q})\mathrm{Var}(\widetilde{q}) - E(\widetilde{q})^3}{[\mathrm{Var}(\widetilde{q})]^{\frac{3}{2}}}$$
$$\frac{1 + 3E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{\sigma}^2_{\widetilde{q}}\right) + 2E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{\sigma}^4_{\widetilde{q}}\right) - 3E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{\sigma}^2_{\widetilde{q}}\right) - 1}{[\mathrm{Var}(\widetilde{q})]^{\frac{3}{2}}} = \frac{2E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{\sigma}^4_{\widetilde{q}}\right)}{[\mathrm{Var}(\widetilde{q})]^{\frac{3}{2}}},$$

where it holds that $E(\widetilde{q}^3) = E_{\widetilde{\sigma}_{\widetilde{q}}}\left[E_{\widetilde{\sigma}_{\widetilde{q}}}\left(\widetilde{q}^3 \mid \sigma_{\widetilde{q}}\right)\right] = E_{\widetilde{\sigma}_{\widetilde{q}}}\left(1 + 3\widetilde{\sigma}^2_{\widetilde{q}} + 2\widetilde{\sigma}^4_{\widetilde{q}}\right)$. It is to be noted that the third cumulant of $\widetilde{q}$, i.e. the numerator of skewness, depends only on the 4-th central moment of $\widetilde{\sigma}_{\widetilde{q}}$. Under the assumption that the r.v. $\widetilde{\sigma}_{\widetilde{q}}$ follows a gamma distribution of parameters α and β, $\widetilde{\sigma}_{\widetilde{q}} \sim \mathrm{Gamma}(\alpha;\beta)$, the characteristics of the structure variable can be rewritten as follow.

Variance:

$$\mathrm{Var}(\widetilde{q}) = \frac{\alpha(\alpha+1)}{\beta^2}.$$

Coefficient of variation:

$$CV(\widetilde{q}) = \frac{\sqrt{\alpha(\alpha+1)}}{\beta}.$$

Skewness:

$$\gamma(\widetilde{q}) = \frac{2(\alpha+2)(\alpha+3)}{\beta\sqrt{\alpha(\alpha+1)}}.$$

Appendix D

Hyperparameters related to $\widetilde{\rho}$

Type of r.v.	**Gamma**				**Exponential**	
Priors	$f(C)$		$f(D)$		$f(C)$	$f(D)$
Parameters	$_1\varepsilon$	$_2\varepsilon$	$_1\psi$	$_2\psi$	ϕ	ω
DELTA	100	45.031	100	4.285	0.450	0.043
OMEGA	100	24.531	100	1.984	0.245	0.020

Table 25: Hyperparameters for both insurers, DELTA and OMEGA, of gamma and exponential distributions.

Posterior expected values

Insurer	**Type of prior**	**Gamma**	**Exponential**
DELTA	$E(\widetilde{C} \mid \widetilde{\rho})$	2.355	1.846
	$E(\widetilde{D} \mid \widetilde{\rho})$	20.970	11.292
OMEGA	$E(\widetilde{C} \mid \widetilde{\rho})$	5.004	5.003
	$E(\widetilde{D} \mid \widetilde{\rho})$	31.160	31.151

Table 26: Expected values of posterior distributions for both insurers, DELTA and OMEGA.