An extension of collective risk model for stochastic claim reserving

Alessandro Ricotta[1] ***Gian Paolo Clemente***[2]

Abstract

The evaluation of outstanding claims uncertainty plays a fundamental role in managing insurance companies. This topic has gained an increasing interest over last years because of the development of a new capital requirement framework under the Solvency II project. In particular, as results of main Quantitative Impact Studies showed, reserve risk is an essential part of underwriting risks and it has a prominent weight on the capital requirement for non-life insurance companies. To this end, we provide here a stochastic methodology in order to evaluate the distribution of claims reserve and to quantify the capital requirement for reserve risk of a single line of business. This proposal extends some existing approaches (see [12], [13], [17] and [19]) and it could represent a viable alternative to well-known methodologies in literature. Finally, a detailed numerical analysis shows a comparison between the proposed methodology and the widely used bootstrapping Over Dispersed Poisson model.

Keywords: stochastic models for claims reserve, capital requirement for reserve risk, collective risk model, average cost methods, Solvency II.

---

[1] Catholic University of Milan
ricotta.alessandro@gmail.com
[2] Corresponding author
Catholic University of Milan, Department of Mathematics, Finance and Econometrics.
Largo Gemelli 1, 20123 Milan
gianpaolo.clemente@unicatt.it

Introduction

New international accounting principles and changes in the regulation frameworks (e.g. Solvency II for European Union member countries (see [4], [9] and [11])) produced a wide development of stochastic methods to evaluate the uncertainty of claims reserve, with the aim to measure the reserve risk. As well known, deterministic methods quantify only the expected value of claims reserve whereas stochastic models provide also the standard deviation or the probability distribution, necessary to assess the capital requirement.

In this regard, there is a variety of methodologies that may be used alone or in combination to derive the best estimate. The appropriateness of one method versus another will depend upon a number of factors including the volume of business, the characteristics of settlement process, the amount of historical data available and the actuary's interpretation of the data.

Focusing instead on stochastic models, a first approach to measure loss reserve uncertainty was proposed by Mack (see [14], [15], [16]) in order to evaluate the prediction error of Chain-Ladder estimate. Prediction Variance is here derived as the sum of purely random fluctuations (Process Variance) and the variability produced by the parameters estimation (Estimation Variance). Furthermore, other approaches (e.g. Bootstrapping ([5]), Generalized Linear Models ([6], [7]) or Bayesian methods [8]) lead to the claims reserve distribution. In this framework, Savelli and Clemente ([20]), extending International Actuarial Association ([13]) proposal, assumed a Collective Risk Model (CRM) to analyse outstanding claims reserve with the target to assess the capital requirement for reserve risk. Incremental payments of each cell are described by a compound Poisson process, either pure or mixed. Exact characteristics (expected value, variance and skewness) of the reserve distribution are proved under the independence between different cells. This strict assumption, that is unlikely to be met in practice, is overcome in [21] by considering correlation between incremental payments and providing mean and variance of claim reserve also in this case

Our goal is to extend this approach by assuming that incremental payments are a compound mixed Poisson process where the uncertainty on claim size is measured via a multiplicative structure variable. Two structure variables, on claim count and average cost, are here considered in order to describe parameter uncertainty on both random variables. Furthermore linear dependency between different development and accident years is also addressed.

Main advantage of this proposal is to directly consider the parameter uncertainty on claim size estimation neglected by previous models.
Under this framework, we obtain the exact characteristics of the claim reserve distribution. Moreover, Monte Carlo method is used to simulate outstanding claims distributions for each accident year, for the total reserve and for the next calendar year (in case of a one-year time horizon evaluation useful for reserve risk evaluation). Model's parameters are calibrated by using data-set of individual claims and an average cost method. The deterministic Frequency-Severity method is here used to estimate separately the number of claims and the average costs for each cell of the bottom part of the run-off triangle. It is also proposed an approach, based on the Mack's formula, to quantify the variance of the structure variables.
Furthermore, we analyse the one-year reserve risk as prescribed in Solvency II. By adapting the "re-reserving" method (see [3] and [18]), we estimate both the variability of claims development result and the extreme quantiles of its simulated probability distribution with the aim to assess the reserve risk capital requirement.
In Section 1, the methodological framework of the proposed model is defined. Main results according to exact moments are also reported. Section 2 describes how parameters can be calibrated. CRM is applied in Section 3 to two non-life insurers and it is compared also with Bootstrap methodology in Section 4 in order to analyse the effects on capital requirement.

1. Collective Risk Model

The aim of this model, based on concepts of the Collective Risk Theory, is to achieve the claims reserve distribution.
As usual in actuarial field, data are reported in a structure with a rectangular shape of dimension $\mathrm{N} \times \mathrm{N}^+$ where rows ($\mathrm{i} = 1, \ldots, \mathrm{N}$) represent the claims accident years (AY) and columns (with $\mathrm{j} = 1, \ldots, \mathrm{N}^+$) are the development years (DY) for the number or the amount of claims. Frequently columns are not equal to rows because of a payments tail. In this case all claims are not completely closed at DY N (i.e. $N^+ > N$, otherwise $N^+ = N$). These structures represent the so called Run-Off triangles (see Appendix A.2 for an example) where observations are available only in the upper triangle $D = \{X_{i,j};\ i + j \leq N + 1\}$ with the cell $(1, N^+)$ also known in case of triangle with tail and 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 (i.e. in the financial year $i+j-1$).
In a similar way, we can define the set $D^n = \{n_{i,j};\ i+j \leq N+1\}$ regarding observed number of paid claims $n_{i,j}$ in the upper triangle.

Future number or amount of payments must be estimated and assigned to the cells in the lower triangle. These cells include unknown values from a random variable whose characteristics must be identified.
We assume that the random variable (r.v.)[3] incremental claims of each cell $\tilde{X}_{i,j}$ will be equal to the aggregate claim amount:

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

and finally the r.v. claims reserve is equal to:

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

where:

- $\tilde{K}_{i,j}$ describes the r.v. number of claims concerning the accident year i and paid in the financial year $i+j-1$. This r.v. is described by a mixed Poisson process in order to consider the parameter uncertainty through a multiplicative structure variable $\tilde{q}$ ($\tilde{K}_{i,j} \sim Po(n_{i,j}\tilde{q})$). This variable is assumed having mean equal to one and standard deviation equal to $\sigma_{\tilde{q}}$.
  By using this mixed Poisson distribution, we catch the parameter uncertainty on number of claims without affecting the expected value of $\tilde{K}_{i,j}$.
  Furthermore, an only one r.v. $\tilde{q}$ affects the r.v. number of claims in the bottom part of the run-off triangle. This choice allows us to consider dependence between expected number of claims of different AY and DY given by the settlement process.

[3] Tilde will indicate a random variable

- $\tilde{Z}_{i,j,h}$ is the random variable that describes the amount of the h^th^ claim occurred in the accident year i and paid after $j-1$ years.
- $\tilde{p}$ describes analogously the parameter uncertainty on claim size. Also in this case, we assume a r.v. having mean equal to one and standard deviation equal to $\sigma_{\tilde{p}}$. With the r.v. $\tilde{p}$, we introduce dependence also between claim-sizes of different cells.

We obtain (see Appendix A.1 for proofs) the exact characteristics of claims reserve under the following assumptions:

- claim count, claim costs and the structure variable $\tilde{p}$ are mutually independent in each cell of the lower triangle;
- claim costs in different cells of the lower run-off triangle are reciprocally independent and in the same cell are i.i.d.;
- structure variable $\tilde{q}$ is independent of the claim costs in each cell
- $\tilde{q}$ and $\tilde{p}$ are independent.

The expected value is equal to:

$$E\left(\tilde{R} \mid D; D^n\right) = \sum_{i=1}^{N} \sum_{j=N-i+2}^{N^+} n_{i,j} m_{i,j} \tag{3}$$

where $n_{i,j}$ represents the expected number of paid claims and $m_{i,j}$ the average cost of paid claims. As described in the next Section, an average cost method can be used to estimate $n_{i,j}$ and $m_{i,j}$. Formula (3) assures that the mean of the stochastic model is equal to the claims reserve derived by the deterministic method.

The variance of the claims reserve is:

$$\sigma^2\left(\tilde{R} \mid D; D^n\right) = E\left(\tilde{p}^2\right) \sum_{i=1}^{N} \sum_{j=N-i+2}^{N^+} n_{i,j} a_{2,Z_{i,j}} + \sigma^2_{\tilde{q}\tilde{p}} \left(\sum_{i=1}^{N} \sum_{j=N-i+2}^{N^+} n_{i,j} m_{i,j} \right)^2 \tag{4}$$

where $a_{k,\tilde{Z}_{i,j}} = E(\tilde{Z}_{i,j}^k)$ is the simple moment of order k of the severity distribution (namely $m_{i,j} = a_{1,Z_{i,j}}$), while $\sigma^2_{\tilde{q}\tilde{p}}$ represents the variance of the r.v. derived as the product of $\tilde{q}$ and $\tilde{p}$ (i.e. $\sigma^2_{\tilde{q}\tilde{p}} = [\left(\sigma^2_{\tilde{p}} + 1\right)\sigma^2_{\tilde{q}} + \sigma^2_{\tilde{p}}]$) .

Variance derived in [21] is a specific case of formula (4) where only structure variable on claim count is considered (i.e. $\tilde{p} = 1$).

The first term is the variance of claims reserve in case of a pure compound Poisson process multiplied by the squared mean of the structure variable $\tilde{p}$. It is noteworthy how the second term is mainly based on the effect of the two structure variables and it takes into account of the positive correlation among incremental payments.
Therefore, structure variables affect variance of the claims reserve and parameters uncertainty appears as a systematic risk that could not be diversified by a larger number of claims. This result is clear when the variability coefficient (CV) is considered:

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

Let $n_{i,j} = T\delta_{i,j}$, then we have

(6) $$\lim_{T\to\infty} CV\left(\tilde{R} \mid D; D^n\right) = \sigma_{\tilde{q}\tilde{p}}$$

where T is the total number of reserved claims and $\delta_{i,j}$ the proportion of reserved claims in the cell (i,j) $\left(\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+} \delta_{i,j} = 1\right)$.
As expected, the relative variability of claims reserve decreases for a larger number of claims. The convergence of limit shows a non-pooling risk equal to the standard deviation of the r.v. defined as the product of the two structure variables considered in the model.
The exact skewness of the claims reserve is described by:

(7) $$\gamma\left(\tilde{R}\mid D;D^n\right)=$$

$$\frac{\gamma_{\tilde{q}\tilde{p}}\sigma^3_{\tilde{q}\tilde{p}}\left(\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}m_{i,j}\right)^3}{\left[\sigma^2_{\tilde{q}\tilde{p}}\left(\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}m_{i,j}\right)^2+E\left(\tilde{p}^2\right)\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}a_{2,Z_{i,j}}\right]^{3/2}}$$

$$+\frac{\left(\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}m_{i,j}\right)\left(\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}a_{2,Z_{i,j}}\right)\left[E\left(\tilde{p}^3\right)E\left(\tilde{q}^2\right)-E\left(\tilde{p}^2\right)\right]}{\left[\sigma^2_{\tilde{q}\tilde{p}}\left(\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}m_{i,j}\right)^2+E\left(\tilde{p}^2\right)\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}a_{2,Z_{i,j}}\right]^{3/2}}$$

$$+\frac{E\left(\tilde{p}^3\right)\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}a_{3,Z_{i,j}}}{\left[\sigma^2_{\tilde{q}\tilde{p}}\left(\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}m_{i,j}\right)^2+E\left(\tilde{p}^2\right)\sum_{i=1}^{N}\sum_{j=N-i+2}^{N^+}n_{i,j}a_{2,Z_{i,j}}\right]^{3/2}}$$

The numerator is the sum of three terms, each of them affected by structure variables. In the first term the skewness of the r.v. $\tilde{q}\tilde{p}$ appears (equal to $\gamma_{\tilde{q}\tilde{p}}=\frac{E(\tilde{p}^3)E(\tilde{q}^3)-3\sigma^2_{\tilde{q}\tilde{p}}-1}{\sigma^3_{\tilde{q}\tilde{p}}}$).
Moreover, the limit of $\gamma(\tilde{R})$ converges to this value:

(8) $$\lim_{T\to\infty}\gamma\left(\tilde{R}\mid D;D^n\right)=\gamma_{\tilde{q}\tilde{p}}$$

If the common assumption of Gamma distribution is used for both structure variables, then $\gamma_{\tilde{q}\tilde{p}} = 2\sigma_{\tilde{q}\tilde{p}} + 2\left(\sigma_{\tilde{q}}^2\sigma_{\tilde{p}}^2\right)\left(\frac{1}{\sigma_{\tilde{q}\tilde{p}}} + \frac{1}{\sigma_{\tilde{q}\tilde{p}}^3}\right)$ leading to a positive skewed distribution of claims reserve.

2. Parameters estimation

To apply the Collective Risk Model, we need to estimate both the expected number of paid claims and the expected claim cost for each cell of the lower triangle conditionally to the set of information D and D^n. At this regard we here use the deterministic Frequency-Severity[4] methodology based on a separate application of the well-known Chain-Ladder method on the triangles of number and claims size respectively. This method allows us to easily estimate both information and to provide a stochastic version of this methodology.
For the sake of clarity, we briefly report the main steps of this method. According to the estimation of future number of paid claims (frequency), the first step is based on the evaluation of development factors (λ_j^n) for each DY as:

(9) $$\lambda_j^n = \frac{\sum_{i=1}^{N-j} n_{i,j+1}^c}{\sum_{i=1}^{N-j} n_{i,j}^c} \qquad with\ j = 1, \dots\dots\dots, N-1$$

where $n_{i,j}^c$ is the cumulative number of paid claims in the cell (i, j).
A tail factor λ_N^n could be included by using the information on the number of reserved claims of first AY at the valuation date or by applying extrapolation methods (see [10]).
Expected cumulative number claims are then obtained as:

[4] For details on this deterministic methodology, see, for instance, [10].

$$(10) \qquad \hat{n}^{c}_{i,j} = n^{c}_{i,N-i+1} \prod_{k=N-i+1}^{j-1} \lambda^{n}_{k} \qquad with \quad \begin{cases} i = 1,\dots\dots, N; \\ j = N-i+2,\dots\dots, N^{+} \end{cases}$$

Expected incremental number of claims $\hat{n}_{i,j}$ is then easily derived as difference of cumulative numbers. This value represents the average parameter of the r.v. $\widetilde{K}_{i,j}$ in the CRM.

The same development technique is also applied to the triangle of cumulative average costs, determined as the ratio between the cumulative amount of paid claims ($C_{i,j}$) and the cumulative number of paid claims in the same cell:

$$(11) \qquad CM^{C}_{i,j} = \frac{C_{i,j}}{n^{c}_{i,j}}$$

This information is easily obtained by the sets D and D^n respectively.
Lower triangle of cumulative average costs $\widehat{CM}^{C}_{i,j}$ is estimated by applying Chain-Ladder method.
Average cost of each cell $\widehat{m}_{i,j}$, that represents the mean of r.v. $\tilde{Z}_{i,j}$ in CRM model, is derived as the ratio between expected incremental payments

$$\hat{X}_{i,j} = \begin{cases} \hat{n}^{c}_{i,j}\widehat{CM}^{c}_{i,j} & if\ j = 1 \\ \hat{n}^{c}_{i,j}\widehat{CM}^{c}_{i,j} - \hat{n}^{c}_{i,j-1}\widehat{CM}^{c}_{i,j-1} & if\ j > 1 \end{cases}$$

and $\hat{n}_{i,j}$.

Parameter uncertainty is a key issue in claims reserve estimate. As shown in Equation (5), standard deviation of structure variables significantly affects the variability coefficient of the claims reserve distribution. We propose to evaluate the standard deviation of structure variables by using Mack's formula (see [14]), being the mean of frequency and severity distributions estimated by a Chain-Ladder technique. In particular the relative variability concerning only the Estimation Error derived via Mack formula allows us to calibrate the standard deviation of $\sigma_{\tilde{q}}$ and $\sigma_{\tilde{p}}$.
However, in the next case study, we preferred to use *a priori* values of $\sigma_{\tilde{q}}$ and $\sigma_{\tilde{p}}$, in order to provide a sensitivity analysis of the effects of these systematic components on cumulants of claims reserve distribution.

Finally, an accurate estimate of $c_{\tilde{Z}_j}$ is a key issue, since the standard deviation of incremental payments depends on it. In general, data from the claim database of the company for each development year are necessary.

3. A practical case study

The stochastic model has been applied to claim experience data of two Italian insurance companies working in the Motor Third Party Liability (MTPL) LoB and concerning accounting years from 1993 to 2004. Real data have been partially modified to save the confidentiality of the data-set. Main information concern number of paid and reserved claims, incremental payments and reserved amounts. For the sake of simplicity, in Appendix A.2 we have reported only the historical cost of incremental paid amounts for the two companies analysed. SIFA insurer is a small-medium company whereas AMASES insurer is a company roughly 10 times larger. The complete run-off period concerning the two insurers is longer than 12 development years and a tail must be considered in the run-off triangles. In the example the tails (i.e. cell $(1993, 12^+)$ of each triangle) are the statutory reserves fixed by the companies for the first accident year.

Expected number of claims ($\hat{n}_{i,j}$) and average cost ($\hat{m}_{i,j}$) are estimated by the Frequency-Severity method as described in Section 2. However, the standard deviation of both the structure variables is assumed to be equal to a fixed prior. The random variables $\tilde{q}$ and $\tilde{p}$, for both companies, are Gamma distributed with mean equal to 1 and standard deviation equal to 3%. The severity of each cell of the triangle is Gamma distributed with mean equal to the average cost $\hat{m}_{i,j}$. In order to estimate cumulants of the severity distribution and consequently the characteristics of the claims reserve we consider the variability coefficient of claim cost, $c_{\tilde{Z}_j}$ (obtained by the company claim database), different for each development year (see Table 1). It should be pointed out that this variability is obviously depending by the LoB, the characteristics of portfolio and the settlement speed of the insurer. For the sake of simplicity, we are assuming the same values for both insurers.

Development Years	$c_{\tilde{Z}_j}$
2	5.75

3	5.70
4	5.85
5	5.05
6	4.65
7	3.35
8	4.70
9	3.50
10	2.45
11	3.60
12	2.45
12+	3.22

Table 1 – Variability coefficients of claim cost for each DY for both companies

Next table shows the simulated characteristics (based on 100,000 simulations) of the claim reserve distribution for SIFA and AMASES (Table 2). The results of 100,000 iterations lead the values of the simulated mean and standard deviation very close to the exact values. The simulated values of the skewness are also not far away from the exact values equal to 0.142 and to 0.110 for the small and the big insurer respectively. We can conclude that this number of simulations provide consistent results.

	Mean*	CV	Skewness
SIFA	229,408	6.08%	0.144
AMASES	2,827,494	4.47%	0.105

*Mean expressed in Thousands of Euro

Table 2 –Main characteristics of simulated claims reserve distribution (100,000 simulations) for SIFA and AMASES

The CRM model provides for SIFA and AMASES a best estimate of approximately 230 and 2,827 millions of Euro. These values match to the claims reserve estimated by the Frequency-Severity deterministic method.
The variability coefficient is lower for AMASES (4.47%) than for SIFA (6.08%) due to a bigger number of reserved claims. In this case, the high number of outstanding claims leads to a relative variability of claims reserve close to the asymptotic value of the variability coefficient (equal to $\sigma_{\tilde{q}\tilde{p}} = 4.24\%$). Moreover, the value of the linear correlation coefficient ρ (calculated assuming equal correlation between the incremental payments) shows a greater dependence for AMASES ($\rho = 0.10$) than for SIFA ($\rho = 0.02$), due to

the greater impact of the structure variables on bigger portfolios. Skewness is quite low for both insurers. Also in this case it is noteworthy the diversification effect with a lower value of $\gamma(\tilde{R})$ for AMASES almost equal to the asymptotic value $\gamma_{\tilde{q}\tilde{p}}$.

Parameter uncertainty has a relevant importance on claims reserve distribution. To this end, we report a sensitivity analysis to evaluate the effect of structure variables on the variability coefficient and the skewness of the claims reserve for both companies. In particular, varying both $\sigma_{\tilde{q}}$ and $\sigma_{\tilde{p}}$ from 1% to 10%, we observe in Figure 1 a convex behaviour of the CV. Function is close-to-linearity when the standard deviations are greater than 10%. The effect of both structure variables ($\tilde{q}$ and $\tilde{p}$) is similar on the CV.

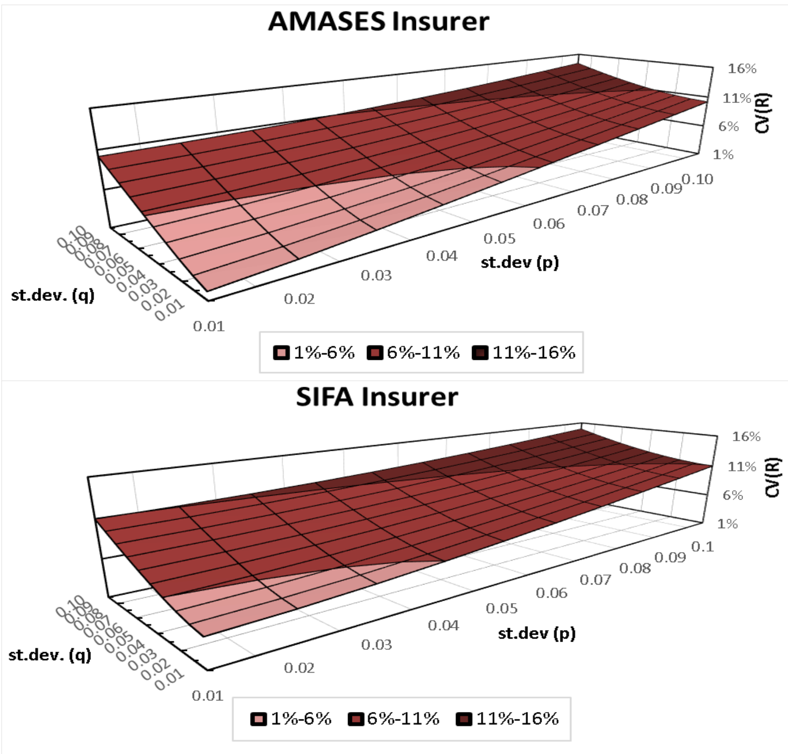

Figure 1 - Variability coefficient of the overall claims reserve for both insurers, depending on different standard deviations of the structure variables $\tilde{q}$ and $\tilde{p}$

A similar behaviour is observed also for skewness (see Figure 2). Parameter uncertainty on claim size tends to affect the skewness of severity distribution more than the r.v. $\tilde{q}$.

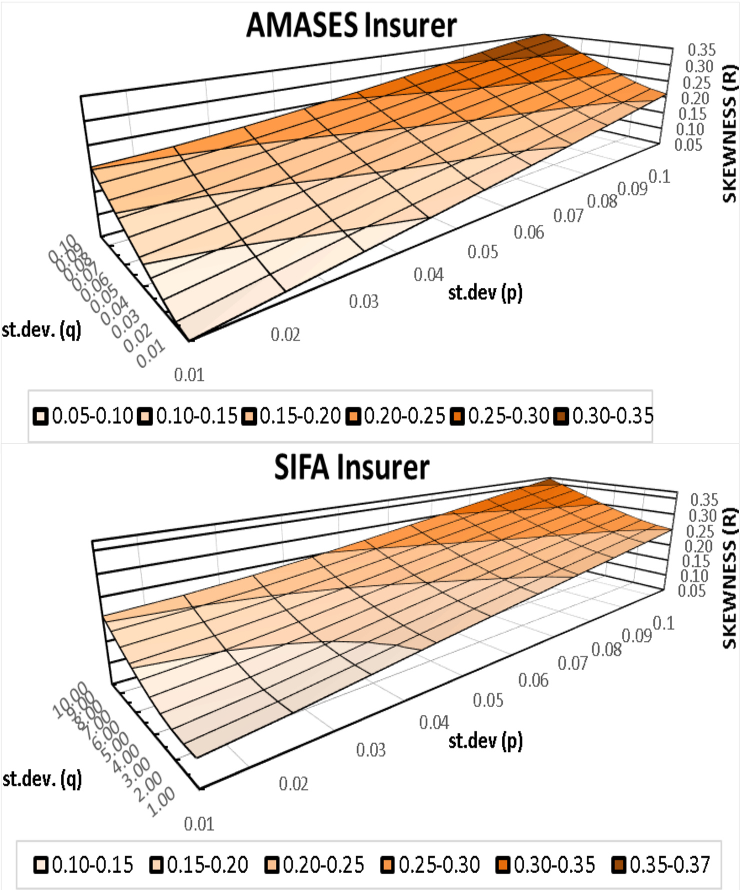

Figure 2- Skewness of the overall claims reserve for both the insurers, depending on different standard deviations of the structure variables $\tilde{q}$ and $\tilde{p}$.

Considering both companies, it is noticeable the greater effect of structure variables on AMASES (see Figure 3). The impact is slightly higher on skewness because of the r.v. $\tilde{p}$ (as shown also in Figure 2). When very high values of $\sigma_{\tilde{q}}$ and $\sigma_{\tilde{p}}$ are considered, CV of claims reserve tend to increase of a value equal to $\sigma_{\tilde{q}\tilde{p}}$. A similar behaviour is also observed for the skewness, where the increase is equal to $\gamma_{\tilde{q}\tilde{p}}$.

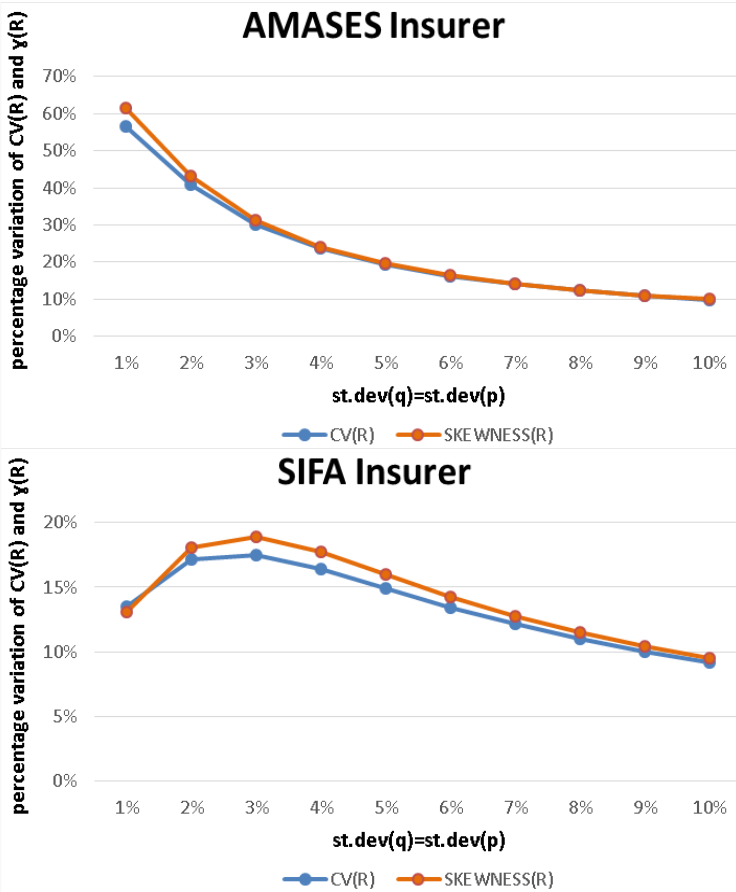

Figure 3 - Variation of CV and skewness of the overall claims reserve for both insurers, depending on different standard deviations of the structure variables $\tilde{q}$ and $\tilde{p}$ assuming $\sigma_{\tilde{q}} = \sigma_{\tilde{p}}$.

The estimate of structure variables based on the Mack's Estimation Error leads to a value of $\sigma_{\tilde{q}}$ and $\sigma_{\tilde{p}}$ equal to roughly 1.96% for SIFA whereas for AMASES the values are equal to 1.62% and 1.53% respectively. It is to be emphasized that estimation based on Mack's approach supplies a higher relative variability for the small insurer. Using these values we obtain the characteristics of claims reserve reported in Table 3.

	CV	Skewness
SIFA	5.13%	0.119
AMASES	2.66%	0.063

Table 3 – CV and skewness of simulated claims reserve distribution (100,000 iterations).

4. One-Year Approach

In this Section, we analyse the reserve risk on a one year time horizon as prescribed by Solvency II. To this end, we adapt the "re-reserving" approach (see [3] and [18]) to our context in order to obtain the "One-Year" reserve distribution of insurer obligations. In particular, we estimate the Solvency Capital Requirement (SCR) for the reserve risk as difference between the quantile at the 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 time zero. Both CRM and the well-known Bootstrap Over Dispersed Poisson (ODP) method (see [6]) are used. It should be highlighted that the two stochastic models lead to a different mean due to the different underlying deterministic method.

Table 4 compares the variability coefficient and the skewness of the "One-Year" reserve distribution given by the CRM and Bootstrap model. In the One-Year approach, both stochastic models provide higher values of relative variability and skewness for SIFA because of a greater pooling risk. In general, CRM leads to a greater CV for both companies than Bootstrap. On the other hand, skewness obtained by the sampling with replacement approach is lower than CRM for SIFA and higher for AMASES.

	CV		**SKEWNESS**	
	CRM(FS)	Bootstrap (CHL)	CRM(FS)	Bootstrap (CHL)
SIFA	5.33%	3.65%	0.217	0.176
AMASES	3.21%	2.86%	0.133	0.143

Table 4 - CV and skewness (One-Year approach) obtained by CRM and Bootstrap ODP for both insurers (100.000 simulations).

Table 5 shows the SCR ratio, evaluated as SCR divided by Best Estimate, obtained by both models. As expected, SIFA has a higher SCR ratio caused by greater CV and skewness. It is to be emphasized that CRM approach is more sensitive to the insurer size providing a higher difference between the SCR ratios. It is interesting to note that in this case study Bootstrap methodology allows to save for both insurers some capital requirement compared to the proposed CRM model. Nevertheless, it should be pointed out that the results of CRM method are widely influenced by the structure variables estimate (i.e.$\sigma_{\tilde{q}}$, $\sigma_{\tilde{p}}$).

	SCR ratio	
	CRM(FS)	Bootstrap (CHL)
SIFA	14.89%	10.13%
AMASES	8.70%	7.75%

Table 5 - SCR ratio obtained by CRM and Bootstrap ODP for both insurers (100.000 simulations).

Finally, it is to be pointed out that the variability coefficient ($c_{\tilde{Z}_j}$) of average cost also plays a key role. The sensitivity analysis, here reported, shows the effects of this variability on the "One-Year" reserve distribution and on the SCR ratio (see Table 6). We assume that $c_{\tilde{Z}_j}$ increases of 50% and 100% for AMASES and SIFA respectively. Higher variability coefficient of the severity leads, obviously, to a high variability and skewness of the One-Year distribution. However, a greater effect is observed for the small-medium insurer caused by a significant pooling risk. Consequently, the capital requirement of SIFA insurer is subjected to a higher increase.

		SIFA	
	Cv	Skewness	SCR ratio
$c_{\tilde{Z}}$	5.33%	0.217	14.88%
$2c_{\tilde{z}}$	9.09%	0.409	27.14%
		AMASES	
	Cv	Skewness	SCR ratio
$c_{\tilde{Z}}$	3.21%	0.133	8.70%
$1.5c_{\tilde{z}}$	3.65%	0.159	10.00%

Table 6 - Variability coefficient, skewness and SCR ratio of both insurers, according to an increase of 100% and 50% of the variability coefficient of the severity (100.000 simulations) for SIFA and AMASES respectively.

Conclusions

We proposed a stochastic model for claim reserving based on Collective Risk Theory approach. According to us, the CRM represents a useful and quite polished stochastic method to evaluate outstanding claims.
We have extended the existing CRM models introducing, by multiplicative way, a structure variable on the claim size. This extension allows us to also consider the parameter uncertainty on claim size, neglected by existing models.

Furthermore, parameters of the model are estimated using claims database and the deterministic model "Frequency-Severity" (based on the Chain-Ladder method) that allows to obtain the number of claims to be paid and the future average costs. We regard estimation of the structure variables as a key issue. The sensitivity analyses underline the strict connection between parameter uncertainty, variability coefficient and skewness of the overall claims reserve.

Moreover, the proposed method is also adapted to quantify the capital requirement as prescribed in Solvency II framework, turning out to be a potential Partial Internal Model for the reserve risk. The case study shows that CRM model supplies results more sensitive to the portfolio size than Bootstrap method. Finally, the sensitivity analysis, here reported, exhibits that the variability coefficient of average costs plays a crucial role on the SCR level.

References

1. Beard R.E., Pentikainen T., Pesonen M. [1984], *Risk Theory – the stochastic basis of insurance*, Third Edition, USA, Chapman & Hall.
2. Daykin C.D., Pentikainen T., Pesonen M. [1994], *Practical Risk Theory for Actuaries*, Monographs on Statistics and Applied Probability, 53, London, Chapman & Hall.
3. Diers D. [2009], *Stochastic re-reserving in multi-year internal models – An approach based on simulations*, Ulm University, Germany.
4. Directive 2009/138 EC.
5. Efron B., Tibshirani R.J. [1993], *An introduction to the bootstrap*, USA, Chapman & Hall.

6. England P., Verrall R. [2002], *Stochastic Claims reserving in general insurance*, British Actuarial Journal 8, pp.443-544.
7. England P., Verrall R. [2006], *Predictive distribution of outstanding liabilities in general insurance*, Annals of Actuarial Science, vol.1,n.2, pp.221-270.
8. England P., Verrall R., Wüthrich M.V. [2010], *Bayesian Overdispersed Poisson Model and the Bornhuetter-Ferguson Claims Reserving Method*, preprint.
9. European Commission [2014], Supplementing directive 2009/138/EU of the European Parliament and the Council on the taking-up and pursuit of the business of Insurance and Reinsurance, Delegated Regulation, Brussels.
10. Friedland J. [2009], *Estimating unpaid claims using basic techniques*, Casualty Actuarial Society.
11. Gisler A. [2010], *The insurance risk in the SST and in Solvency II: modelling and parameter estimation,* ETH Zurich.
12. Klugman S.A., Panjer H.H., Willmot G.E. [2004], *Loss Models*, USA, Wiley, second edition.
13. International Actuarial Association [2004], *A Global Framework for Insurer Solvency Assessment*, Research Report of The Insurer Solvency Assessment Working Party.
14. Mack T. [1993], *Distribution-free calculation of the standard error of Chain-Ladder reserve estimates*, ASTIN Bulletin, 23.
15. Mack T. [1994], *Measuring the variability of Chain-Ladder reserve estimates,* Casualty Actuarial Society.
16. Mack T. [1999], *The standard error of Chain-Ladder reserve estimates: recursive calculation and inclusion of a tail factor*, Austin Bulletin vol.29, pp. 361-366.
17. Meyers G.G., Klinker F.L., Lalonde D.A. [2003], *The aggregation and correlation of insurance exposure,* Casualty Actuarial Society.
18. Ohlsonn E., Lauzenings [2008], *The one-year non-life insurance risk*, Astin Colloquium, Manchester.
19. Rantala J. [1982], *Solvency of insurers and equalization reserves: Vol. II – Risk theoretical model*, Helsinki, Insurance Publishing Company Ltd.
20. Savelli N., Clemente G.P. [2009], *A collective risk model for claims reserve distribution*, Proceedings of "Convegno di Teoria del Rischio", Campobasso, pp.59-88.

21. Savelli N., Clemente G.P. [2011], *Stochastic claim reserving based on CRM for Solvency II purposes*, ASTIN Colloquium 2011, Madrid 19-22 June 2011.
22. Wüthrich M.V., Merz M. [2008], *Modelling the claims development results for solvency purposes*, Cass E-Forum, pp. 542-568.

Appendix A.1
Variance of claims reserve – Proof of Formula (4)

We here compute the conditional variance of incremental payments $\tilde{X}_{i,j}$ and the conditional variance of claims reserve $\tilde{R}$ given the sets $D = \{X_{i,j};\ i + j \le N + 1\}$ and $D^n = \{n_{i,j};\ i + j \le N + 1\}$. For the sake of brevity we will omit the conditioning on D and D^n.

We firstly focus on the variance of a single cell ($\tilde{X}_{i,j}$):

$$\sigma^2\left(\tilde{X}_{i,j}\right) = E\left(\tilde{X}^2_{i,j}\right) - \left[E\left(\tilde{X}_{i,j}\right)\right]^2 = E\left[E\left(\tilde{X}^2_{i,j} \mid \tilde{p}\right)\right] - \left[E\left(\tilde{X}_{i,j}\right)\right]^2 =$$

$$E\left(\tilde{p}^2\right)\left[E\left(\tilde{q}^2\right)n^2_{i,j}m^2_{i,j} + n_{i,j}a_{2,\tilde{z}_{i,j}}\right] - n^2_{i,j}m^2_{i,j} =$$

$$\left[\left(\sigma^2_{\tilde{p}} + 1\right)\sigma^2_{\tilde{q}} + \sigma^2_{\tilde{p}}\right]n^2_{i,j}m^2_{i,j} + E\left(\tilde{p}^2\right)n_{i,j}a_{2,\tilde{z}_{i,j}} = \sigma^2_{\tilde{q}\tilde{p}}n^2_{i,j}m^2_{i,j} + E\left(\tilde{p}^2\right)n_{i,j}a_{2,\tilde{z}_{i,j}}$$

where

$$\sigma^2\left(\tilde{q}\tilde{p}\right) = E\left[\sigma^2\left(\tilde{q}\tilde{p} \mid \tilde{p}\right)\right] + \sigma^2\left[E\left(\tilde{q}\tilde{p} \mid \tilde{p}\right)\right] = \left(\sigma^2_{\tilde{p}} + 1\right)\sigma^2_{\tilde{q}} + \sigma^2_{\tilde{p}}.$$

Now it is possible to calculate the variance of $\tilde{R}$ as shown below:

$$\sigma^2\left(\tilde{R}\right) = \sum_{i,j\in B}\sigma^2\left(\tilde{X}_{i,j}\right) + \sum_{i,j\in B}\sum_{\substack{h,k\in B\\(h\neq i\,\vee\, k\neq j)}}\operatorname{cov}\left(\tilde{X}_{i,j};\tilde{X}_{h,k}\right)$$

The second term measures the covariances between the cells of the lower run-off triangle, here indicated with the notation $B = \{i + j > N + 1\}$, and it equals to:

$$\sum_{i,j\in B}\sum_{\substack{h,k\in B\\(h\neq i\,\vee\, k\neq j)}}\left\{E\left[\operatorname{cov}\left(\tilde{X}_{i,j};\tilde{X}_{h,k} \mid \tilde{p}\right)\right] + \operatorname{cov}\left[E\left(\tilde{X}_{i,j} \mid \tilde{p}\right);\mathrm{E}\left(\tilde{X}_{h,k} \mid \tilde{p}\right)\right]\right\} =$$

$$\sum_{i,j\in B}\sum_{\substack{h,k\in B\\(h\neq i\,\vee\, k\neq j)}}\left\{n_{i,j}m_{i,j}n_{h,k}m_{h,k}\left[\left(\sigma^2_{\tilde{p}} + 1\right)\sigma^2_{\tilde{q}} + \sigma^2_{\tilde{p}}\right]\right\}.$$

Therefore,

$$\sigma^2\left(\tilde{R}\right) = \underbrace{\left(\sigma^2_{\tilde{p}} + 1\right)}_{E\left(\tilde{p}^2\right)}\sum_{i,j\in B}n_{i,j}a_{2,Z_{i,j}} + \underbrace{\left[\left(\sigma^2_{\tilde{p}} + 1\right)\sigma^2_{\tilde{q}} + \sigma^2_{\tilde{p}}\right]}_{\sigma^2\left(\tilde{q}\tilde{p}\right)}\left(\sum_{i,j\in B}n_{i,j}m_{i,j}\right)^2 =$$

$$E\left(\tilde{p}^2\right)\sum_{i,j\in B}n_{i,j}a_{2,Z_{i,j}} + \sigma^2_{\tilde{q}\tilde{p}}\left(\sum_{i,j\in B}n_{i,j}m_{i,j}\right)^2.$$

Skewness of claims reserve – Proof of Formula (7)

In a similar way, we derive the skewness of claims reserve, defined as:

$$\gamma\left(\tilde{R}\right)=\frac{\mu_3\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)}{\sigma^3\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)}$$

where the third central moment can be rewritten as:

$$\mu_3\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)=E\left[\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)^3\right]-3E\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)\sigma^2\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)-\left[E\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)\right]^3$$

The key issue is to determine the first term. The cube of a polynomial is equal to:

$$E\left[\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)^3\right]=$$

$$E\left[\sum_{i,j\in B}\left(\tilde{X}_{i,j}^3\right)\right]+2E\left\{\sum_{i,j\in B}\left[\left(\tilde{X}_{i,j}^2\right)\left(\sum_{\substack{h,k\in B\\(h\neq i\vee k\neq j)}}\tilde{X}_{h,k}\right)\right]\right\}+E\left\{\sum_{i,j\in B}\left[\left(\tilde{X}_{i,j}\right)\left(\sum_{\substack{h,k\in B\\(h\neq i\vee k\neq j)}}\tilde{X}_{h,k}\right)^2\right]\right\}$$

By using conditional mean with respect to $\tilde{p}$ and $\tilde{q}$ respectively, we obtain:

$$E\left[\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)^3\right]=E\left(\tilde{p}^3\right)E\left(\tilde{q}^3\right)\left(\sum_{i,j\in B}n_{i,j}m_{i,j}\right)^3+E\left(\tilde{p}^3\right)\sum_{i,j\in B}n_{i,j}a_{3,\tilde{z}_{i,j}}$$

$$+3E\left(\tilde{p}^3\right)E\left(\tilde{q}^2\right)\left(\sum_{i,j\in B}n_{i,j}m_{i,j}\right)\left(\sum_{i,j\in B}n_{i,j}a_{2,\tilde{z}_{i,j}}\right)$$

where for a single cell the following relation holds:

$$E\left(\tilde{X}_{i,j}^{3}\right)=E\left[\left(\tilde{X}_{i,j}^{3}\mid\tilde{p}\right)\right]=E\left(\tilde{p}^{3}\right)\left[E\left(\tilde{q}^{3}\right)n_{i,j}^{3}m_{i,j}^{3}+3E\left(\tilde{q}^{2}\right)n_{i,j}^{2}m_{i,j}a_{2,\tilde{Z}_{i,j}}+n_{i,j}a_{3,\tilde{Z}_{i,j}}\right]$$

The second and third term of the skewness' numerator are equal respectively to:

$$3E\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)\sigma^{2}\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)=$$

$$3\left\{\sigma_{\tilde{q}\tilde{p}}^{2}\left(\sum_{i,j\in B}n_{i,j}m_{i,j}\right)^{3}+E\left(\tilde{p}^{2}\right)\left(\sum_{i,j\in B}n_{i,j}m_{i,j}\right)\left(\sum_{i,j\in B}n_{i,j}a_{2,\tilde{z}_{i,j}}\right)\right\}$$

and

$$E\left[\left(\sum_{i,j\in B}\tilde{X}_{i,j}\right)\right]^{3}=\left(\sum_{i,j\in B}n_{i,j}m_{i,j}\right)^{3}.$$

Summing up the three addends of the numerator, we have a term equal to the third central moment of the product of structure variables:

$$\sigma^{3}\left(\tilde{q}\tilde{p}\right)\gamma\left(\tilde{q}\tilde{p}\right)=\mu_{3}\left(\tilde{q}\tilde{p}\right)=$$

$$E\left[\left(\tilde{q}\tilde{p}\right)^{3}\right]-3E\left(\tilde{q}\tilde{p}\right)\sigma^{2}\left(\tilde{q}\tilde{p}\right)-E\left[\left(\tilde{q}\tilde{p}\right)\right]^{3}=E\left(\tilde{q}^{3}\right)E\left(\tilde{p}^{3}\right)-3\sigma^{2}\left(\tilde{q}\tilde{p}\right)-1$$

and finally it is easy to obtain Formula (7).

Appendix A.2

SIFA

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12	12+
1993	38,364	37,956	15,350	6,100	3,178	2,701	1,503	1,361	1,008	899	287	727	1,068
1994	41,475	44,466	15,938	6,840	3,300	2,730	1,009	1,152	767	467	456		
1995	46,520	47,579	15,095	6,909	3,392	1,390	1,338	1,186	922	559			
1996	47,925	51,866	17,599	6,305	2,875	2,124	2,233	1,208	873				
1997	51,420	52,085	17,290	6,021	2,719	3,037	1,320	1,124					
1998	57,586	54,150	19,610	7,530	4,110	2,780	2,267						
1999	55,930	54,941	20,947	10,499	5,864	3,313							
2000	51,005	53,191	21,819	8,365	4,714								
2001	51,693	51,572	18,668	8,833									
2002	54,954	51,611	18,604										
2003	59,763	53,743											
2004	60,361												

Figure A2.1 – Triangle SIFA (Incremental paid amounts, thousands of Euro)

AMASES

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12	12+
1993	193,474	172,618	87,200	45,798	29,768	19,795	19,782	17,315	13,372	12,552	8,831	8,053	19,889
1994	199,854	168,966	80,543	40,656	29,053	21,121	19,964	14,249	10,720	13,684	6,008		
1995	225,578	186,764	93,349	47,609	30,971	26,291	17,621	18,410	14,662	7,591			
1996	256,398	236,678	105,616	51,172	37,338	24,085	20,754	12,082	14,137				
1997	282,956	263,196	120,383	63,689	37,220	29,239	23,120	15,509					
1998	292,428	284,401	141,400	56,390	40,195	27,955	29,987						
1999	312,350	285,506	131,687	75,252	46,549	38,731							
2000	327,673	307,992	161,516	77,965	52,696								
2001	339,899	326,280	185,911	101,273									
2002	371,275	385,847	193,006										
2003	388,025	390,737											
2004	398,686												

Figure A2.2 – Triangle AMASES (Incremental paid amounts, thousands of Euro