Empirical Bayesian Approach in Micromodels of Reserve Risk

Claudia Fedorcakova
Junior Consultant, ADDACTIS® Worldwide
claudia.fedorcakova@addactis.com

ADDACTIS Worldwide
Risk & Insurance Solutions
+32 (0)2 526 13 10

Argentina - Brazil - Belgium - Chile - Colombia - Costa Rica - Ecuador - France - Germany - Guatemala - Italy - Luxembourg - Mexico
Morocco - The Netherlands - Panama - Peru - Poland - Senegal - Singapore - Spain - Switzerland

Contents

Following the recent legislative changes in insurance business, the insurers are forced to reconsider how their reserves are evaluated. In Europe, with the Solvency II regulation, reserve risk has been redefined from the ultimate claim horizon to a one-year risk horizon, meaning that also the distribution of claim estimates must be estimated on a one-year horizon. Therefore, it is desirable to develop a methodology to update the estimates based on new observations collected over one year, which is a typical task for Bayesian approach.

Whereas the traditional reserve estimation is based on the aggregated data, new trend is to utilize all the information available and analyse each claim separately. This way the application of claims' specific features, such as non-proportional reinsurance, policy limits or deductibles, is possible.
This study thus aims to construct individual claims reserving model and setup simulation method to calculate one-year reserve risk by applying Bayesian statistical methods.

(In cooperation with University of Economics in Prague)

1 INDIVIDUAL CLAIMS MODEL SPECIFICATION

The model is based on estimating two components of the risk for each loss modelled:

- Probability that the loss will be closed in particular development period,
- Change of severity of the loss until its closure.

The ultimate value of a single claim is defined as:

$$\hat{X} = X_0 \sum_{j=1}^{\omega} \hat{p}_j \hat{\mu}_j,$$

where X_0 represents the initial value, $(\hat{p}_j)$ stands for the estimator of probability of claim closure, $(\hat{\mu}_j)$ is the estimator of development factors and ω denotes the maximum development year. In other words, the ultimate value of a claim is equal to the initial value of the claim at reporting multiplied by weighted average of estimates of development factors.
The proposed model concerns the IBNER (Incurred But Not Enough Reported) claims representing the estimated amount of the future development on already reported open claims.

1- Initial value

The initial value X_0 is the current value of an open claim. This parameter is not estimated, since it is already known, but will serve as a starting point to simulate the ultimate position of a single claim and will be further adjusted by random number of random development factors.

2- Probability of claim closure

The estimator of probability of claim closure ($\hat{p}_j$) is a vector indicating in which development year was the claim closed. If a claim is being closed in given development period, the corresponding vector component will be equal to 1, otherwise 0. Since no reopenings are assumed in the model, a claim can be closed only once.

The probability of being closed p_j is defined as sum of claims being open in development year j-1 and closed in development year j, divided by sum of all claims open in j-1.

In order to simulate the probabilities of claim closure, and subsequently simulate, in which development year the claim will be closed, the Dirichlet distribution[1] will be applied.

Once the claim is projected as closed (the open/ closed indicator equals to 1), it will stay closed for all the subsequent development periods. The maximum development period ω is set to 10, meaning that all claims must be closed till then. This condition is satisfied by setting the open/ closed indicator in the last development year equal to 1.

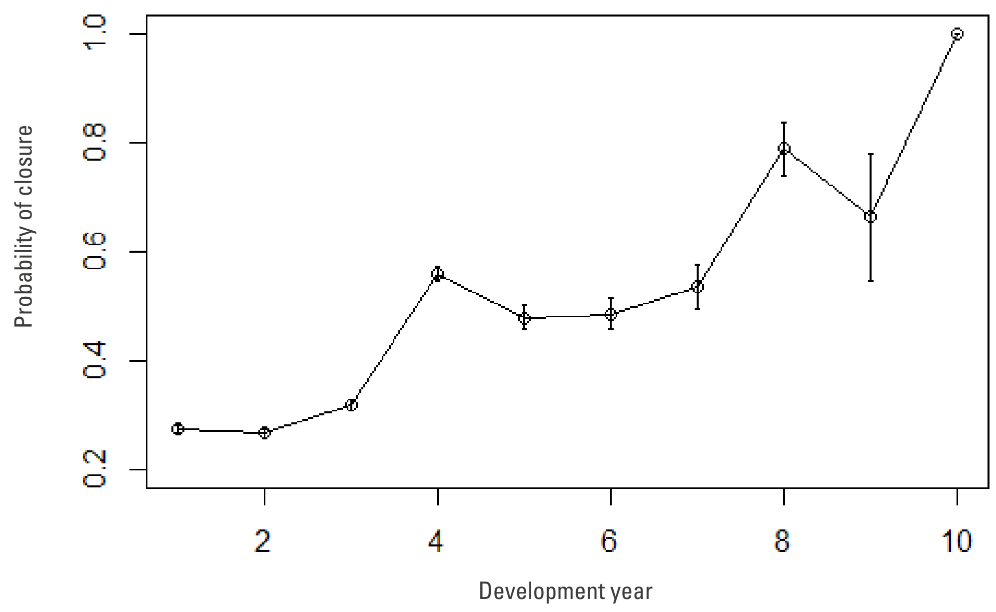

Fig. 1: Simulated probabilities of closure

3- Change in claim severity

The estimator of development factors[2] (μ_j) is a vector of simulated development factors for each claim. Development factors are simulated only when the claim is open, i.e. open/ closed indicator is 0.

The convenient way of fitting the development factors is to use the Gamma distribution, due to the fact that it is positively skewed. Another suitable feature of Gamma distribution is that it models non-negative values. The most practical approach of generating the development factors is by using the Generalized Linear Model (GLM). To allow updating the parameters after obtaining new data, the focus will be set on its Bayesian equivalent – Bayesian Generalized Linear Model (BGLM).

Firstly, we need to define the prior distribution – prior mean and scale. We have chosen to apply the noninformative prior by setting the prior mean value of regression parameters to 0 and the prior scale to a large value (e.g. 10 000). Secondly, the family distribution and link function need to be specified. The family distribution used in the study is Gamma distribution with log link function. Finally, symbolic description of the model to be fit needs to be defined. In our model we assume the presence of dependency of development factor on development year:

dev.factor ~ dev.yr.

The fitted values for each development factor are illustrated in the Figure below. Note that we are working with the incurred data, hence some of the development factors might be lower than 1.

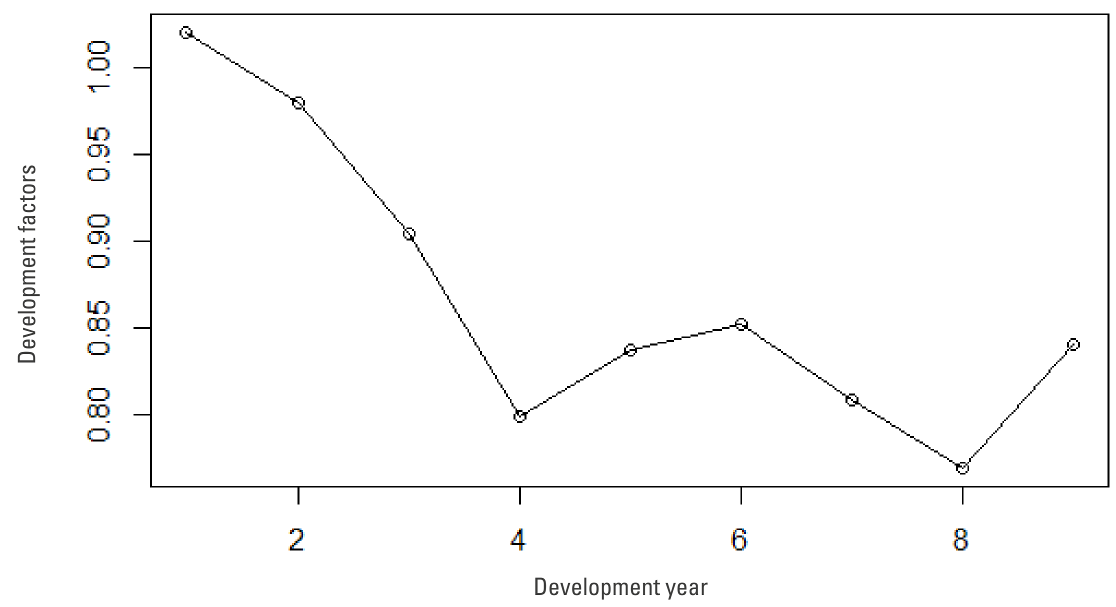

Fig. 2: Simulated values of development factors

4- Simulation of the ultimate

After estimation of the two model components, the ultimate claim value for each simulation might be calculated. For each development period greater than the current development period of a claim and smaller than the maximum development period, firstly the value of open/ closed indicator is simulated by randomly generating from multinomial distribution.

If the claim is simulated as closed, no development factor is generated. If the claim is still open, the model samples corresponding development factor from draws produced in the previous section.

The ultimate value of a single claim is computed by multiplying the present state of the accident with cumulated intra-year development factors. The total ultimate value is then obtained by aggregating results for all the open claims.

[1]The Dirichlet distribution is a multivariate generalization of the beta distribution, which very often serves as a prior distribution in Bayesian statistics, most of all for the multinomial distribution.
[2]Development factor is defined as the claim value in development period $j+1$ divided by the claim value in development period j.

RESULTS AND COMPARISON WITH AGGREGATED METHOD

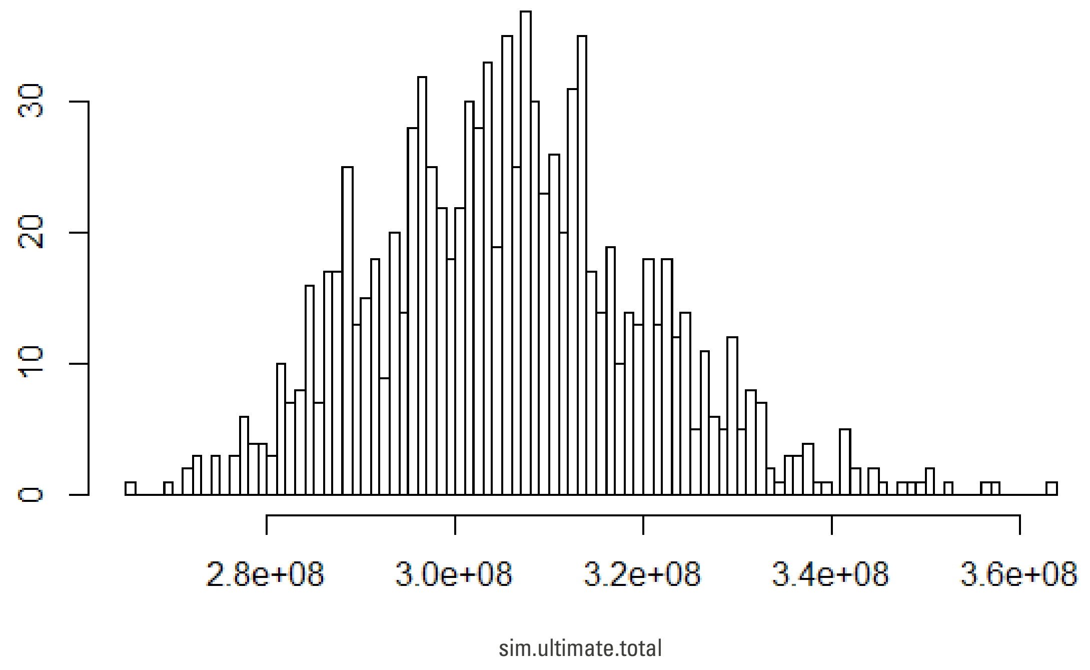

Fig. 3: Distribution of simulated total ultimate value on the open claims

Based on the previously defined model specifications, 1 000 simulations of the ultimate are performed for each open claim. After aggregation, the distribution of total ultimate value on all open claims is attained.

The results show that the model based on individual claims results in significantly lower variability compared to the method based on the aggregated data - Bootstrap (see Figure 4). The standard deviation decreased by more than 50% when utilizing detailed claim information and estimating the ultimate claim value on a claim by claim basis.

	Mean	Median	St deviation	VaR (99,5 %)
Individual claims model	305 810 973	305 592 292	14 676 489	350 561 918
Bootstrap	416 010 659	414 841 626	31 987 216	509 592 278

Fig. 4: The comparison of results obtained by Individual claims model and Bootstrap

3 MODEL FOR RESERVE RISK ON 1-YEAR RISK HORIZON

As was stressed in the beginning, in Europe, the new Solvency II regulatory regime requires the insurers to measure the reserve risk on a 1-year risk horizon. This requires estimating the ultimate at time *t*+1. First step is to simulate development of each open claim for the subsequent development year. Based on the newly generated data, new posterior distribution for both the components of ultimate claim value estimator is defined. The ultimate at time *t*+1 is then simulated the same way as described in the last section.

The following table compares the statistical properties of the two ultimates estimated at time *t*+1 and time *t*, respectively. The third row represents the mean value of 1-year reserve risk.

	Mean	Median	St deviation	VaR (99,5 %)
Individual claims model (time t+1)	179 889 140	179 029 305	12 870 925	220 568 872
Individual claims model (time t)	305 810 973	305 592 292	14 676 489	350 561 918
1-year reserve risk	125 921 833			

Fig. 5: Ultimate values estimated at time t+1 and time t, and the value of 1-year reserve risk calculated as their difference

CONCLUSION

For the example presented the ultimate value estimator based on the detailed data generates 2-times smaller deviation (which serves as a model efficiency indicator) than the model based on aggregated claims. For the one-year horizon, the variance is even lower. This leads to significantly lower Value at Risk value and, consequently, to a lower capital requirement. In conclusion, even though the models based on non-aggregated data require much more calculations and computing time, the significant increase in efficiency of estimation serves as a great motivation.

(In cooperation with University of Economics in Prague)