

*Marek Karwański**, *Piotr Jałowiecki***, *Arkadiusz Orłowski****

THE PERSPECTIVES OF BAYESIAN METHODS FOR MODELING FINANCIAL RESERVES IN INSURANCE

Abstract. Building reserves for liabilities is an important issue in the financial statement of a general insurance company. The purpose of this paper is to present models for prediction IBNR (incurred but not reported) reserves. The modeling is based on data which describes the claims settlement of a car insurance portfolio – the data consists of about 60,000 claims, which incurred in 2001–2006. Several models of the claims prediction are proposed, from estimation in traditional deterministic Chain Ladder and the Poisson GLM model (Kramer 1998) – commonly used techniques in practices – to Markov Chain Monte Carlo. The models presented show the significant differences in variance of IBNR reserves. From that point of view the Bayesian approach has some characteristics that make it particularly attractive for their use in actuarial practice.

Key words: IBNR reserves, Chain Ladder, GLM, MCMC, Bayes approach.

I. INTRODUCTION

It was many efforts undertaken in European Union from the end of 90's to build principles of accounting system due to monitoring solvency and safety levels in insurance companies. The estimation of reserves is the basic function of such systems. Many types of reserves are subject to the legal discipline and should be supported in proposed models.

The technical reserves connected with payments for the claims are the most important. The core parts of these reserves are called IBNR: incurred but no reported. The estimating of the IBNR reserves should be done by prognoses the future payments and very often needs the help of stochastic approaches (Hoedemakers 2005).

Incorrect or even false in relation to the market, reserves estimation generates financial risks. The underestimating the reserves leads to activate the sources of funding that are little effective and in extreme cases it brings to loss

* M.Sc., Chair of Informatics, Warsaw University of Life Sciences.

** Ph.D., Chair of Informatics, Warsaw University of Life Sciences.

*** D.Sc., Chair of Informatics, Warsaw University of Life Sciences.

the financial liquidity. The overestimating, cause on the profitability of enterprise and can be also classify to unreliable financial game (for example: the taxes).

Despite large interest the models of reserves that are in practical use have the disadvantages in value of precision. The ultimate level of claims can never be known with certainty. As a result, the reserve estimates are subject to some degree of error. A close fit between actual and expected claims year by year indicates that the model used to establish the reserves was a good one. However, note that close correspondence to historic data might mean the model is inadequate for forecasting. This is particularly essential because of incubation period of claims for many insurance products is relatively long, and it set very large requirements regarding to the computer systems designed to collecting and processing data in insurance institutions.

The primary objective of this paper is to compare reserving methodologies in general insurance in order to show the benefits and lack of usage Bayesian models.

II. THE DATA DESCRIPTION

Reserve calculations require the data on any claims amounts that have already been paid, in the case that a claim has been reported. The data upon which projections of future claims are based are usually presented into a triangular "run-off" format. This format tabulates the claim data (payments, numbers, etc.) according to the year in which the claim arose and the year in which the payment was made. The difference between the payment date and the accident date is referred to as the development time. Different time periods can be used, particularly for short-tail classes. In this paper we use quarterly data which describes the claims settlement of a car insurance portfolio – the data consists of about 60,000 claims, which incurred in 2001–2006.

An example of claim payments in run-off format is set out below in Fig. 1. The totals in horizontal lines are the sums of claims paid and forecasted and have the name ultimate claims reserves. The aim of all techniques is to complete the empty triangle on the lower right hand side of the table. We use the ultimate reserves to comparisons the accuracy of any individual prognostic model.

		Development quarter						
		0	1	2	11	12	13
Accidence quarter	0	6 885 948	9 300 204	2 792 431		23 929	24 745	6 636
	1	5 467 407	14 424 164	3 186 565		4 465	6 408	
	2	7 107 011	11 637 456	3 158 137		12 204		
	3	7 762 889	14 134 719	3 093 034				
	4	6 268 682	13 473 372	2 813 021				
	5	6 152 633	9 821 930	2 998 257				
	6	8 623 089	12 220 004	2 824 476				
	7	6 083 174	12 128 078	3 564 129				
	8	4 961 384	10 506 169	3 534 131				
	9	4 151 186	9 378 533	2 400 385				
	10	5 756 542	14 445 841	3 816 426				
	11	4 495 405	11 598 023	3 370 382				
	12	6 791 702	11 419 797					
13	5 312 413							

Fig 1. Aggregate data table – the claims settlement of a car insurance portfolio

III. ANALYTICAL MODELS

In present paper three classes of models are compared (England, Verral 2002). Some of them are used as an industry standard the other are rarely and are just going to be more common. Because exhausted discussion about the reserving method is beyond of the scope of this paper the only one model from every class was chosen. The accuracies of the models are modeled by the average and the standard errors of forecasted part of ultimate claims reserves.

Calculations were conducted in the following models:

- Deterministic Chain Ladder model – with the nonparametric standard error estimator proposed by T. Mack.
- Generalized linear model GLM with overdispersed-Poisson distribution and canonical $\log()$ linking function, to calculate the standard error two approach was applied: delta method and bootstrap technique
- Bayes model – with the Gibbs Sampling methods to simulate direct draws, with a priori Jeffreys distribution and the normal likelihood function

Mack Model Description

It is very simple model (England, Verral 2002) that is based on the assumption that there is a consistent delay pattern in the payment of claims. The deterministic model underlying the basic chain-ladder method can be expressed as:

- for cumulated claims $Z_{i,j}$ there exist the deterministic factors of "development" $f_0, \dots, f_{j-1} > 0$ such, that for all i and j :

$$E[Z_{i,j+1} | Z_{i,j}, \dots, Z_{i,j}] = E[Z_{i,j} | Z_{i,j-1}] = f_{j-1} Z_{i,j-1} \quad (1)$$

- The claim $Z_{i,j}$ for different accidence quarters are independent
- That variance fulfils condition:

$$Var[Z_{i,j+1} | Z_{i,j}, \dots, Z_{i,j}] = Z_{i,j-1} \sigma_j^2 \quad (2)$$

Method GLM Description

The class of generalized linear models (GLM) (Dobson 2002), (Halekoh 2004) can be defined as:

- The incremental claims $C_{i,j}$ are random variables distributed according to the overdispersed Poisson and there exist such positive parameters $\gamma_0 \dots \gamma_j, \mu_0 \dots, \mu_j$ and $\phi > 0$

$$C_{ij} \sim Poisson(m_{ij}) \quad Var[C_{ij}] = \phi m_{ij} \quad E[C_{ij}] = m_{ij} = \mu_i \gamma_j \quad (3)$$

- $\hat{\mu}_i$ are independent random variables such that $\mu_i = E(C_{i,j})$ for accidence time period i .
- $C_{i,j}$ and $\hat{\mu}_k$ are independent.

When trying to estimate the prediction error of future payments and reserve estimates using statistical methods, the problem reduces to estimating the two components: the process variance and the estimation variance. The standard errors of ultimate claims reserves estimators, which are the totals of paid and predicted values of claims error is the combination of sampling errors with suitable correlations and errors obtained from model. In practice however two methods are used more common: that are analytical delta approach and bootstrap technique applied to residual values.

In the first approach the prediction errors of reserve estimates for defined accidence and development period and the total reserve estimates can be calculated according to:

$$MSE = \phi \mu_{ij} + \mu_{ij}^2 Var(\eta) \\ MSE = \sum \phi \mu_{ij} + \sum \mu_{ij}^2 Var(\eta_{ij}) + 2 \sum Cov(\eta_{ij}, \eta_{ik}) \mu_{ij} \mu_{ik} \quad (4)$$

In those cases, the variance of the sum of predicted values is considered, taking account of any covariances between predicted values.

The second approach is more robust against deviations from the hypothesis of the model. We treat the obtained data as if they are an accurate reflection of the population, and then draw many bootstrapped samples by sampling, with replacement, from a pseudo-population consisting of the obtained data. Such technique is called bootstrapping [Pinheiro et al 2003]. A new bootstrap statistic is defined as a function of the bootstrap estimate and a bootstrap simulation of the residual between future reality and model prediction. This statistic is called the prediction error. For each bootstrap loop the prediction error is then kept in a vector and after adding process error the percentile method is used to obtain the desired results.

Bayesian Model Description

Rather than performing a complicated maximization one can calculate and subsequently performs a series of simple simulations. Bayesian Markov Chain Monte Carlo methods allow us to estimate the parameters in the model [Walsh 2004]. The idea of Monte Carlo simulation is to draw a set of samples from a target density. These samples can be used to approximate the target density with the empirical point-mass function. However, we need to introduce sophisticated techniques to draw the samples based on for example Gibbs sampling algorithms [Congdon 2003]. So, a posteriori distribution is used not in analytical form but in tabular figure:

$$p^N(\theta, x_1, \dots, x_K) = \frac{1}{N} \sum_1^N \delta_{\theta(i)}(\theta, x_1, \dots, x_K)$$

$$E_{\theta, N}(f(x_1, \dots, x_K)) = \frac{1}{N} \sum_1^N f(x_1, \dots, x_K, \theta(i)) \xrightarrow{a.s.} E_{\theta}(f) = \int_X f(x) p(\theta, x_1, \dots, x_K) dx_1, \dots, x_K$$
(6)

The key word is probability a posteriori . The probability a posteriori is proportional to probability a priori multiplied by likelihood function.

$$p(\theta | x_1, \dots, x_K) \propto p(x_1, \dots, x_K | \theta) p(\theta) \Rightarrow \text{posterior} \propto \text{likelihood} \times \text{priori} \quad (5)$$

MCMC is a strategy for generating samples $x(i)$ while exploring the state space X using a Markov chain mechanism. The Gibbs sampler is a special case of MCMC sampling (Mandrekar et al 2005). The key to the Gibbs sampler is that one only considers univariate conditional distributions – the distribution when all of the random variables but one are assigned fixed values. Such conditional distributions are far easier to simulate than complex joint distributions and usually have simple forms (Kass 1998).

IV. RESULTS

The aim of this present paper was the comparison and the evaluation of different models for estimating IBNR reserves. Following models were used to comparisons:

- Chain Ladder model – with standard error estimator proposed by T. Mack.
- Generalized linear model GLM – with standard error estimator by delta method
- Generalized linear model GLM – with standard error estimator by bootstrap technique
- Bayes model – with the Gibbs Sampling methods to simulate direct draws

The estimators of average IBNR reserves as well as its standard error for ultimate reserves for every of above methods were calculated for every accident period. The real values were showed for the whole period in fig 2.

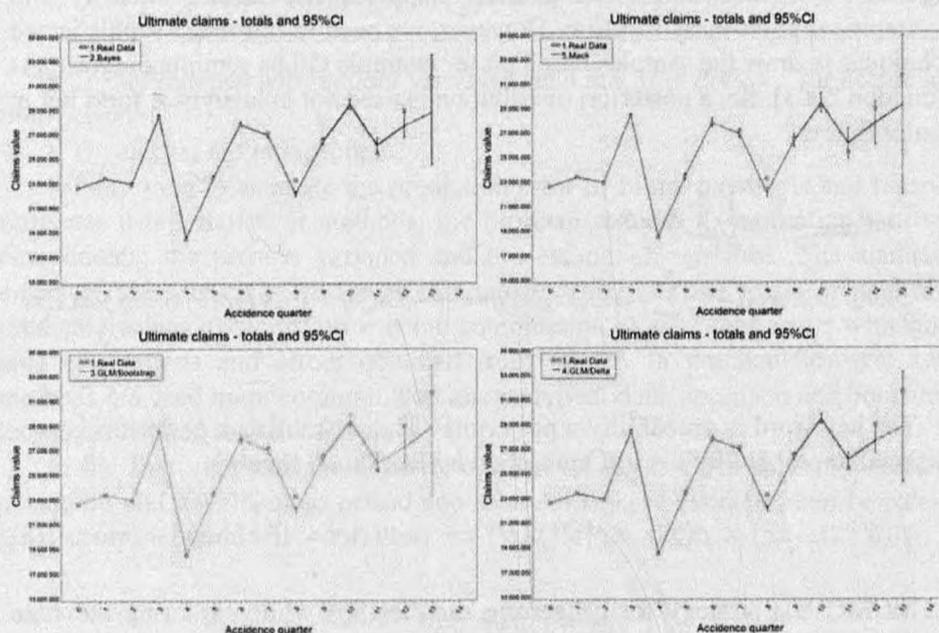


Fig 2. Ultimate claims – means and standard errors – for different accident period

In actuarial practice the reserves are built not on the best estimates but on 90-centil or sometimes on 70-centil of estimated ultimate claims. In the figure below the value of estimated reserves as a function of different significance levels is shown.

The results of comparisons for original data are showed in graphical form on Fig. 3. The “effectiveness” of considered models looks very like for first quarters of accident. Differences begin really in last quarters. From practical point of view it is the most important because these are the recent reserves for insurance institution and business profitability depends on the level of these reserves.

It is worth to notice that reserving methods applied to real data provide a range of estimates. The prediction error is of more interest, representing, not only the standard deviation of the expected reserves, but the practical frame for decision making about the costs based on the actual level of agreement of risk. This is the reason why the precision of estimators is the crucial for business.

Based on a comparison of predicted IBNR reserves, the Bayesian model appears to produce quite good, unbiased predictions and reasonable confidence interval estimates. It seems clear that this procedure is general enough and provides very useful information of the characteristics such as probabilities that allow us to build financial scenarios and cross them with premiums. It provides to calculation of various types of financial risk measures. One must be careful, however, since the results depend slightly on the used prior distributions on an unknown quantity.

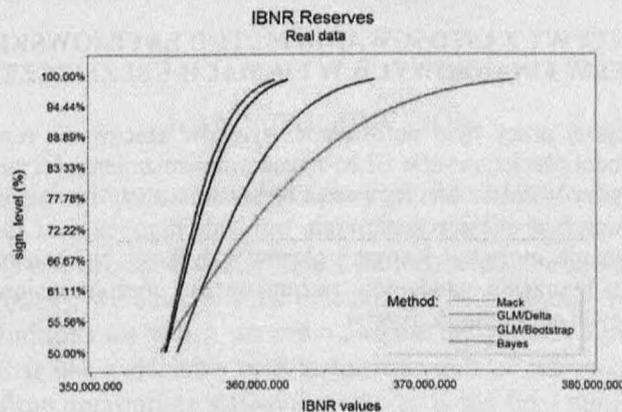


Fig 3. IBNR reserves as a function of significance level for different methods

Further work in this direction should focus on the use of this technique when there is instability in the proportion of ultimate claims paid in the early development period, so that the other technique yields unsatisfactory results.

REFERENCES

- Congdon P. (2003), *Applied Bayesian Modelling*, Willey Series in Prob.
- Dobson A. J. (2002), *An Introduction to Generalized Linear Models*, Hapman&Hall.
- England P. D., Verrall R. J., (2002), *Stochastic Claims Reserving In General Insurance*, Institute of Actuaries.
- Halekoh U. (2004), *ST112. Generalized Linear Models*, MoAS.
- Hoedemakers T. (2005), *Modern Reserving Techniques for the Insurance Business*, Leuven.
- Kass E. (1998), *Markov Chain Monte Carlo in Practice: A Round Table Discussion*, Amer.Statist.Associacion.
- Kramer E. (1982), *IBNR Claims and the Two Way Model Of ANOVA*, Scandinavian actuarial Journal.
- Mandrekar J., Sargent D., Novotny P., Solan J. (2005), *A General Gibbs Sampling Algorithm for Analysing Linear Models Using the SAS System*, Mayo Clinic.
- Pinheiro P., Silva A., Louders C. M. (2003), *Bootstrap Methodology in Claim Reserving*, Tech.University of Lisbon.
- Walsh B. (2004), *Markov Chain Monte Carlo and Gibbs Sampling*, Lecture Notes for EEB 581.

Marek Karwański, Piotr Jałowiecki, Arkadiusz Orłowski

PERSPEKTYWY ZASTOSOWANIA METOD BAYESOWSKICH DO BUDOWY REZERW FINANSOWYCH W FIRMACH UBEZPIECZENIOWYCH

Celem niniejszej pracy było porównanie wyników szacowania rezerw metodami tradycyjnymi: model Macka, modele GLM z szacowaniem zmienności metodą bootstrap oraz metody Bayesa MCMC. Analizy przeprowadzone zostały na danych jednej z dużych amerykańskich firm ubezpieczeniowych.

Takie porównanie, może być jednym z głosów w dyskusji, w kontekście trwającego aktualnie procesu tworzenia standardów rachunkowości ubezpieczeniowej oraz systemów monitorujących adekwatność rezerw.