Uniwersytet Ekonomiczny w Katowicach
Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 291 · 2016

Alicja Wolny-Dominiak

Uniwersytet Ekonomiczny w Katowicach
Wydział Ekonomii
Katedra Metod Statystyczno-Matematycznych w Ekonomii
alicja.wolny-dominiak@ue.katowice.pl

ON THE USE OF GROWTH CURVE IN LOSS RESERVING

Summary: This paper propose some modification of the method of prediction of incurred but not reported (IBNR) claim reserves in non-life insurance based on the growth curve modelling. Literature put forwards a wide variety of methods to predict the IBNR claim reserves, mostly using the chain-ladder technique. The method discussed herein is based on two-stage estimation of the expected amount of losses to emerge: the estimation of the ultimate loss by year and the estimation of the pattern of the loss emergence. In this procedure, a non-linear model of the growth curve is applied in which two restricting assumptions are made. Firstly, it is assumed that incremental losses follow an over-dispersed Poisson (ODP) distribution. Secondly, as the pattern of the loss to emerge, two-parametric log-logistic and Weibull growth curves are assumed. A different non-linear model is proposed in this paper to predict the IBNR claim reserves. In it, the three-parametric Gompertz growth curve is adopted. In order to estimate the model parameters, the non-linear (weighted) least squares (NLS) method is applied, in which incremental losses follow the normal distribution. Moreover, a non-parametric approach to the growth curve modelling based on spline fitting is proposed as an alternative. All calculations are carried out in **R** party using the package {ChainLadder}.

Keywords: loss reserving, Gompertz growth curve, non-linear least squares method, spline.

Introduction

Every non-life insurance company has to manage three basic business areas: *pricing*, *reserving* and *solvency*. In the pricing area the interest is: *How much should individual policyholders be charged? (premium price)*. Consequently: *How much of the aggregate premium income should be earmarked to meet fu-*

ture commitments, i.e. pay insurance compensations? (technical provisions). And finally: *What is the probability that the company will maintain solvency? (ruin probability).*

The largest item on an insurer's balance sheet are technical provisions. Any variations in their values have a great impact on the insurer's financial strength. A large part of the provisions is the reserve for incurred but not reported (IBNR) claims or – simply – the loss reserve, which is crucial to the insurer's solvency. The total loss reserve is generally determined by statistical methods based on both deterministic techniques and stochastic models and is a sum of outstanding loss liabilities. Two important changes are introduced under the Solvency II Directive, compared to Solvency I: the possible estimation methodology is based on "the best estimate" principle and the company total loss reserve is divided into separate loss reserves for individual lines of business (LOB). In this context, the best estimate can be defined as appropriate estimation of the expected volume of a certain value of the loss reserve excluding any margins – especially security margins – based on currently available information. This paper is focused on the total loss reserve for a single LOB.

A wide variety of the loss reserve calculation methods are presented in literature, see e.g. Mack [1999]; England & Verrall [2002]; Wüthrich & Merz [2008]; Pobłocka [2011]; Wolny-Dominiak [2014]. The method investigated herein is the one in which the loss reserves are predicted using a statistical model based on the growth curve modelling. This approach is considered in Clark [2003], where the classical log-likelihood is used, and extended by Zhang et al. [2012], where the Bayesian estimation is made.

The aim of this paper is to propose a non-linear model to predict IBNR claim reserves, in which the three-parametric Gompertz growth curve is adopted. In order to estimate the model parameters, the non-linear (weighted) least squares (NLS) method is applied, in which incremental losses follow the normal distribution. Moreover, a non-parametric approach to the growth curve modelling based on spline fitting is proposed as an alternative.

The introduction of the paper describes the specific type of data used in loss reserving and then the total loss reserve is defined. The first part presents the general theory of the growth curve modelling in the context of loss reserving. The last part is devoted to the proposals made herein: the parametric model and the non-parametric approach. Finally, the case study is presented. In all calculations the **R** software is used [R Core Team, 2012].

1. General definition of the total loss reserve

Let the random variable Y_{ij} with y_{ij} realizations, $i, j = 1,...,n$ be the cumulative loss amount of insurance claims that occurred in year i (accident year) and reported after t_j months in year j (development year). Consider matrix $[y_{ij}]_{nxn}$ of the cumulative loss amount. Elements y_i for $i + j \le n+1$ are observed data, while y_i for $i + j > n+1$ represent future unobserved data. Such a matrix is called *the loss triangle* and has a general form as shown on Figure 1.

i	1	...	n-1	n	p_i
1	y_{11}	...	y_{1n-1}	y_{1n}	p_1
2	y_{21}	...	y_{2n-1}		p_2
$\vdots$		$\iddots$			
n	y_{n1}				p_n

Fig. 1. The cumulated loss triangle

The additional column p_i represents the total earned premium collected in the accident year i. The reported losses are the sum of diagonal elements from the loss triangle. Then, the total loss reserve for a fixed origin year i, defined as the total outstanding loss liabilities for i, is expressed by the following formula:

$$R_i = Y_{in} - Y_{i,n+1-i} \tag{1}$$

Under the assumption that two random variables $Y_{i_1 j_1}$ and $Y_{i_2 j_2}$ are independent if $i_1 \ne i_2$, the total loss reserve is then of the form:

$$R = \sum_{i=1}^{n} Y_{in} - \sum_{i+j=n+1} Y_{ij} \tag{2}$$

The predictors of the loss reserves for origin year i and the total loss reserve can then be written as:

$$\hat{R}_i = Y_{in} - Y_{i,n+1-i} \tag{3}$$

$$\hat{R} = \sum_{i=1}^{n} \hat{Y}_{in} - \sum_{i+j=n+1} Y_{ij} \tag{4}$$

The specific form of the predictor depends on the model used to obtain predictors $\hat{Y}_{ij}$. This paper investigates a non-linear model based on the analytical formula for a specific growth curve.

The total loss reserve prediction accuracy is typically measured using the root mean squared error of prediction (RMSEP):

$$MSEP(R) = E[(\hat{R} - R)^2] \quad (5)$$

The analytical calculations of the MSEP are made in Clark [2002]. Unfortunately, the formula describing the estimator has a rather complex structure. An alternative may be to use the parametric bootstrap technique [see Efron & Tibshirani, 1994]. This, however, goes beyond the scope of this paper.

2. Loss reserving using the growth curve model

In order to illustrate the operation of the loss reserve estimation method based on the growth curve model, an example loss triangle taken from Mack [1993] can be used. The cumulative version of the loss triangle is presented in Table 1.

Table 1. Loss triangle (in thousands)

(i, j)	1	2	3	4	5	6	7	8	9	10	Premium
1	358	1125	1735	2183	2746	3320	3466	3606	3834	3901	10 000
2	352	1236	2170	3353	3799	4120	4648	4914	5339		10 400
3	291	1292	2219	3235	3986	4133	4629	4909			10 800
4	311	1419	2195	3757	4030	4382	4588				11 200
5	443	1136	2128	2898	3403	3873					11 600
6	396	1333	2181	2986	3692						12 000
7	441	1288	2420	3483							12 400
8	359	1421	2864								12 800
9	377	1363									13 200
10	344										13 600

Source: Mack [2003].

The chart presented on Figure 2 indicates that for a fixed origin year i cumulative loss values make a curve that may be interpreted as the growth curve.

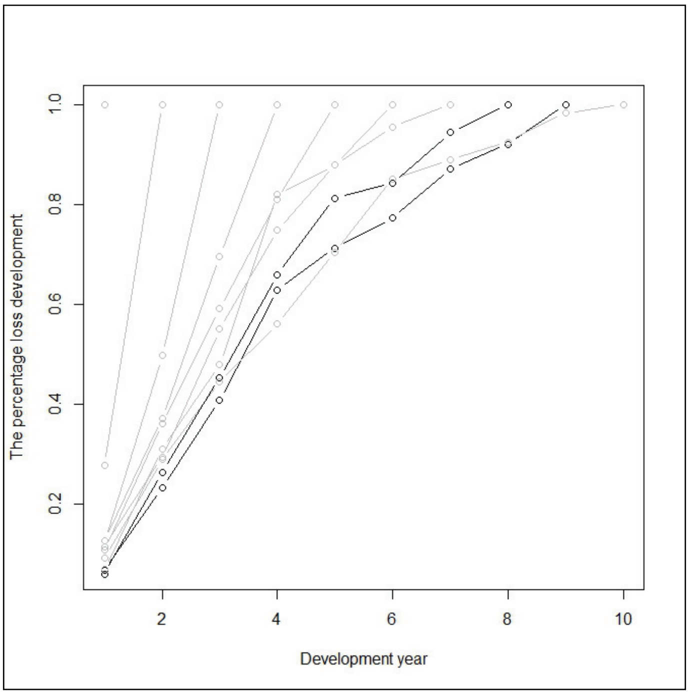

Fig. 2. Loss development in origin years

Since cumulative losses in subsequent development years grow exponentially rather than linearly, the pattern can be described by a growth curve, marked further as $G(\cdot)$.

With a given form of the growth curve $G(\cdot)$, the total loss reserve is predicted using the model of the ultimate loss Y_{in} in the origin year i and the pattern of the loss emergence $G(\cdot)$. The pattern shows the percentage loss development from 0% to 100%. As the biggest number of observations occur for the first origin year, the growth curve parameters are estimated for $i=1$ only. It is further assumed that the pattern of the loss emergence in the other periods will be identical, i.e. the assumed growth curve will have the same parameters in subsequent origin years.

The method proposed by Clark [2002] assumes the following form of the ultimate loss model:

$$Y_{in} = p_i \cdot u \cdot [G(j;\Theta) - G(j-1;\Theta)] + \xi_i \tag{6}$$

where u is the ultimate loss ratio for the loss triangle, ξ_i − the disturbance with $E(\xi)=0$ and $Var(\xi)=\sigma^2$ and $G(j;\Theta)$ − the proper growth curve with the parameter vector Θ . Thus, the parameter vector of the model has the following form:

$$(u,\Theta)' \tag{7}$$

because it is assumed that σ^2 is known.

In order to estimate parameters u and Θ in Model (3), the maximum likelihood estimation (MLE) method is used. In order to find the MLE estimators $\hat{u}$ and $\hat{\Theta}$ analytically without using an iterative algorithm, three strict assumptions are adopted:

(A1) – the loss in any period has a constant ratio $\frac{Variance}{Mean} = \sigma^2$,

(A2) – σ^2 is known,

(A3) – incremental losses follow an over-dispersed Poisson distribution with the probability function $P(Y=y) = \dfrac{\lambda^{\frac{y}{\sigma^2}} e^{-\lambda}}{\left(\frac{y}{\sigma^2}\right)!}$ and the two first moments of the form $E(Y) = \lambda\sigma^2$, $Var(Y) = \lambda\sigma^4$. Under assumptions (A1)-(A3), the calculations are simplified substantially and the analytical derivation of the MLE estimator $\hat{u}$ is possible. The log-likelihood function is as follows:

$$l(\lambda,\sigma; y_1,\dots,y_m) = \sum_i^m \log[\frac{\lambda^{\frac{y_i}{\sigma^2}} e^{-\lambda}}{(\frac{y_i}{\sigma^2})!}] = \sum_i^m \{\frac{y_i}{\sigma^2}\log(\lambda) - \lambda - \log[(\frac{y_i}{\sigma^2})!]\} \tag{8}$$

Because parameter σ^2 is assumed as known, Function (5) is reduced to:

$$l(\lambda; y_1,\dots,y_m) = \sum_i^m [y_i \log(\lambda) - \lambda] \tag{9}$$

Using (3), the log-likelihood is:

$$l(u; y_1,\dots,y_m) == \sum_i^m \{y_i \log(p_i \cdot u \cdot [G(t_j;\Theta) - G(t_{j-1};\Theta)]) - \\ - p_i \cdot u \cdot [G(t_j;\Theta) - G(t_{j-1};\Theta)]\} \tag{10}$$

Solving the equation $\frac{\partial l}{\partial u} = 0$, the MLE estimator $\hat{u}$ is expressed as:

$$\hat{u} = \frac{\sum_{i=1}^{m} y_i}{\sum_{i=1}^{m} p_i \cdot [G(t_j;\Theta) - G(t_{j-1};\Theta)]} \tag{11}$$

Naturally, apart from $\hat{u}$, there is still a need to estimate the vector of parameters Θ. In Clark [2002], as well as in Zhang et al. [2012], the loss emergence pattern is modelled using two-parametric growth curves, based on the log-logistic and Weibull distributions, which are defined as follows:

$$G_L(j;\Theta) = \frac{j^{\varpi}}{j^{\varpi} + \theta^{\varpi}}, \ \Theta = (\varpi, \theta) \tag{12}$$

$$G_W(j;\Theta) = 1 - \exp[-(\frac{j}{\theta})^{\varpi}], \ \Theta = (\varpi, \theta) \tag{13}$$

The formulae presented above are used for development years $j = 1,...,n$. Inserting the selected form of the curve into Formula (10) and solving scoring equations $\frac{\partial l}{\partial \varpi} = 0$ and $\frac{\partial l}{\partial \theta} = 0$, the MLE estimators of parameters $\hat{\varpi}$ and $\hat{\theta}$ are obtained. Plugging them into Formula (4), the predictor of the total loss reserve under Model (6) is:

$$\hat{R} = \sum_{i=1}^{n} p_i \cdot \hat{u} \cdot [\hat{G}(j;\hat{\Theta}) - \hat{G}(j-1;\hat{\Theta})] - \sum_{i+j=n+1} Y_{ij} \tag{14}$$

3. Loss reserving modification using the growth curve model

The method presented above assumes the analytical form of the growth curve, which may be the Weibull or the log-logistic curve known from biological sciences. Both these curves are two-parametric functions with a relatively simple form derived from the exponential function. However, it is possible to assume many more analytical forms of the curve [e.g. see Zwietering et al., 1990]. An equally popular but a three-parametric growth curve referred to as the Gompertz curve defined as in (15) below is selected for the purposes of this paper.

$$G_G(j;\Theta) = \rho \exp[-\exp(\theta - \varpi j)], \ \Theta = (\varpi, \theta, \rho) \tag{15}$$

Based on three parameters, the curve maps the real pattern of the loss emergence better, but the downside is that determination of the curve parameter estimators analytically is rather complex. This problem can be solved flexibly by applying a numerical algorithm instead of the MLE method. The non-linear least squares (NLS) technique of parameter estimation as in Bates and Watts [1988] is used herein. The appropriate model is thus defined as follows:

$$Y_{in} = p_i \cdot u \cdot [G_G(j;\Theta) - G_G(j-1;\Theta)] + \xi_i \tag{16}$$

where ξ_i is normally distributed disturbance with $E(\xi) = 0$ and $Var(\xi) = \sigma_\xi^2$. As a result, the vector of parameters of Model (14) has the following form:

$$(u, \Theta, \sigma_\xi^2)' \tag{17}$$

The estimators of the above-mentioned parameters are obtained by minimizing the following sum of squares:

$$\sum_{i+j\le n+1} (Y_{ij} - p_i \cdot u \cdot [G_G(j;\Theta) - G_G(j-1;\Theta)]) -> \min \tag{18}$$

The minimum value of (15) occurs if the appropriate gradient is zero.

In practice, it is easy to solve Expression (18) using the NLS technique implemented in the **R** software package in the form of the function **nls** {stats}. The only problem is the correct setting of initial values, which has a considerable impact on the estimation duration time. Following Cleveland [1979], locally weighted regression (LWR) is applied herein. Plugging estimators $(\hat{u}, \hat{\Theta}, \hat{\sigma}_\xi^2)'$ into Formula (14), the total loss reserve predictor is obtained.

An alternative approach is to abandon the parametric framework and avoid using models with the functional form fixed in advance. Instead, various non-parametric methods may be used. This paper proposes the following procedure:

1) fit the loss development pattern $G_S(j)$, $j = 1,...,n$ using cubic splines [see Judd, 1998],
2) estimate parameter u according to Formula (7):

$$\hat{u} = \frac{\sum_{i=1}^{m} y_i}{\sum_{i=1}^{m} p_i \cdot [\hat{G}(j) - \hat{G}(j-1)]}$$

3) predict the total loss reserve using Equation (14).

There is no obstacle to applying different smoothing techniques, such as the isotonic regression and the Pool-Adjacent-Violators Algorithm (PAVA) described by de Leeuw et al. [2009] and Gamrot [2012].

4. The case study

The loss triangle from Table 1 is taken to illustrate the proposed solutions. The goal is to predict the loss reserves $\hat{R}_1,...,\hat{R}_{10},\hat{R}$. The first step is to:

- estimate the parameters of the log-logistic G_L, Weibull G_W and Gompertz G_G growth curves,
- fit cube splines $\hat{G}_S$.

The obtained results are listed in Table 2 below.

Table 2. Parameters of growth curves

Growth curie	Estimate	s.e.	t-value	Pr(>|t|)
ϖ Weibull	1.32	0.08	15.88	0.00
θ Weibull	55.20	11.28	4.89	0.00
ϖ Log-logistic	1.37	0.10	13.41	0.00
θ Log-logistic	68.07	18.14	3.75	0.00
ρ Gompertz	0.93	0.04	24.10	0.00
θ Gompertz	1.22	0.04	30.45	0.00
ω Gompertz	0.05	0.00	15.67	0.00

In the results presented in Table 2 all parameters are statistically significant at level 5%. The fitted loss emergence pattern is then as in Table 3.

Table 3. Fitted patterns of loss emergence

Development year j	$\hat{G}_W(j)$	$\hat{G}_L(j)$	$\hat{G}_G(j)$	$\hat{G}_S(j)$
1	76.94%	76.87%	77.65%	73.53%
2	73.79%	73.74%	74.72%	72.26%
3	67.22%	67.21%	68.29%	67.95%
4	63.80%	63.82%	64.81%	65.31%
5	58.52%	58.57%	59.28%	62.58%
6	49.34%	49.45%	49.39%	51.50%
7	41.73%	41.84%	41.11%	41.12%
8	33.96%	34.03%	32.84%	32.71%
9	16.57%	16.42%	16.17%	21.10%
10	6.10%	5.86%	8.12%	6.75%

The fitted growth curves are presented graphically on Figure 3.

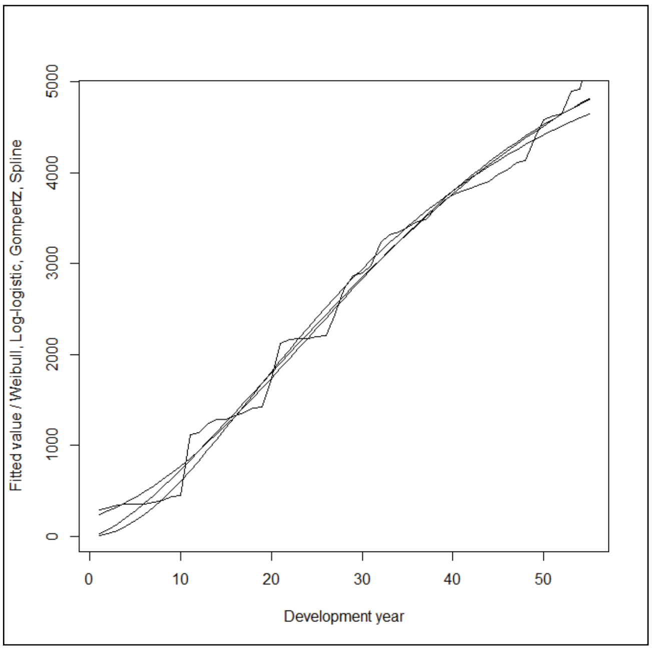

Fig. 3. Plot of fitted patterns of loss emergence

The goodness-of-fit statistics is calculated in every model. Definitely, the lowest value is for spline fitting $\hat{\sigma}_s^2 = 0.81$, compared to the results obtained for $\hat{\sigma}_G^2 = 18.37$, $\hat{\sigma}_W^2 = 81.63$, $\hat{\sigma}_L^2 = 281.4$. As for models, the best results are obtained using the one with the Gompertz curve. Therefore, the $\hat{G}_S(j)$ values from Table 3 are used in further calculations.

Table 4. Predictors of loss reserves

Origin year i	$p_i \cdot \hat{u}$	Growth - spline	Reported	$\hat{R}_i$
1	5 791.46	17.30%	3 901	1 001.93
2	6 023.12	18.57%	5 339	1 118.46
3	6 254.78	22.88%	4 909	1 430.98
4	6 486.44	25.52%	4 588	1 655.26
5	6 718.09	28.25%	3 873	1 898.05
6	6 949.75	39.33%	3 692	2 733.53
7	7 181.41	49.71%	3 483	3 569.80
8	7 413.07	58.12%	2 864	4 308.72
9	7 644.73	69.73%	1 363	5 330.88
10	7 876.39	84.08%	344	6 622.85

Finally, the total loss reserve predictor is $\hat{R} = 29\,670\,450$. The essential information is naturally the size of the prediction error. The estimator can be determined using the bootstrap technique, which is the subject of the Author's further work.

Conclusion

The estimation of loss reserves using the growth curve modelling is a useful alternative in the investigation of the pattern of the emergence of losses for a single LOB. The non-linear model gives a flexible method of estimation in which changing the way of fitting the growth curve is straightforward. It is proved that using the growth curve in reserves makes it also possible to apply a non-parametric approach. Cubic splines are just an example – they can be replaced with another technique. In the parametric approach the model proposed in this paper requires the assumption of disturbance normality. This, however, does not exclude an easy transition to other distributions. It is sufficient to use a different procedure for the model parameter estimation instead of the NLS technique and go on to the generalized non-linear least squares method. Therefore, it may be stated that the use of the growth curve in the loss reserve prediction creates sample opportunities for application.

Literature

Bates M., Watts G. (1988), *Non-linear Regression Analysis and Its Application*, John Wiley & Sons, New York.

Clark D.R. (2003), *Ldf Curve-fitting and Stochastic Reserving: A Maximum Likelihood Approach*, "Casualty Actuarial Society E-Forum", Fall, s. 41-91.

Cleveland W.S. (1979), *Robust Locally Weighted Regression and Smoothing Scatterplots*, "Journal of American Statistical Association", 74(368), s. 829-836.

Efron B., Tibshirani R.J. (1994), *An Introduction to the Bootstrap*, CRC press.

England P.D., Verrall R.J. (2002), *Stochastic Claims Reserving in General Insurance*, "British Actuarial Journal", 8(03), s. 443-518.

Gamrot W. (2012), *Simulation-assisted Horvitz-Thompson Statistic and Isotonic Regression* [in:] J. Ramik, D. Stavarek (eds.), *Proceedings of 30th International Conference on Mathematical Methods in Economics 2012*, Silesian University of Opava, School of Business Administration, Karvina, p. 207-212.

Judd K.L. (1998), *Numerical Methods in Economics*, MIT Press, Cambridge MS.

de Leeuw J., Hornik K., Mair P. (2009), *Isotone Optimization in R: Pool-adjacent-violators Algorithm (PAVA) and Active Set Methods*, "Journal of Statistical Software", 32(5), 1-24.

Mack T. (1993), *Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates*, "ASTIN Bulletin" 23, s. 213-225.

Mack T. (1999), *The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor*, "ASTIN Bulletin", 29(2), s. 361-366.

Pobłocka A. (2011), *Rezerwa IBNR w ubezpieczeniach majątkowych – praktyczne metody jej szacowania* [*IBNR Claim Reserve in Non-life Insurance – Practical Methods of Estimation*], Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 207, s. 173-189.

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Wien, Austria.

Wüthrich M.V., Merz M. (2008), *Stochastic Claims Reserving Methods in Non-life Insurance*, John Willey & Sons, England.

Wolny-Dominiak A. (2014), *Zmodyfikowana regresja logarytmiczno-normalna w szacowaniu rezerwy szkodowej* [*Modified Log-normal Regression in Loss Reserve Estimation*], „Studia Ekonomiczne", nr 181, Uniwersytet Ekonomiczny w Katowicach, s. 220-234.

Zhang Y., Dukic V., Guszcza J. (2012), *A Bayesian Non-linear Model for Forecasting Insurance Loss Payments*, "Journal of Royal Statistics", 175 (Part 2), s. 637-656.

Zwietering M.H., Jongenburger I., Rombouts F.M., Riet K. van 't (1990), *Modelling of the Bacterial Growth Curve*, "Applied and Environmental Microbiology", 56, s. 1875-1881.

ZASTOSOWANIE KRZYWEJ WZROSTU DO SZACOWANIA REZERWY SZKODOWEJ

Streszczenie: W artykule zaproponowano modyfikację metody predykcji rezerwy szkodowej (IBNR) w ubezpieczeniach majątkowych, w której wykorzystuje się modelowanie krzywej wzrostu. Literatura zawiera szeroką gamę metod predykcji rezerwy IBNR, głównie przy użyciu techniki chain-ladder. Metoda omówiona w niniejszym artykule opiera się na estymacji oczekiwanej wartości skumulowanych szkód przeprowadzanej dwuetapowo: szacowanie wartości szkód dla jednego roku wypadkowego oraz szacowanie krzywej wzrostu opisującej rozwój szkodowości, zakładając dalej, że krzywa jest taka sama dla każdego roku wypadkowego. Do szacowania krzywej wzrostu wykorzystuje się model nieliniowy, w którym przyjmuje się dwa podstawowe założenia. Po pierwsze, zakłada się, że rozkład skumulowanej wartości szkód ma rozkład ODP. Po drugie, zakłada się dwie parametryczne krzywe wzrostu: log-logistyczną oraz Weibulla. W artykule zaproponowano zastosowanie trzyparametrycznej krzywej wzrostu Gompertza. W celu oszacowania parametrów krzywej wykorzystano ważoną metodę najmniejszych kwadratów (NLS), w której przyjęto normalny rozkład skumulowanej wartości szkód. Ponadto zaproponowano alternatywne podejście nieparametryczne do modelowania krzywej wzrostu oparte na splinach. W przykładzie numerycznym wykorzystano rzeczywisty trójkąt szkód zaczerpnięty z literatury. Wszelkie obliczenia przeprowadzono w programie **R,** wykorzystując częściowo pakiet {ChainLadder}.

Słowa kluczowe: rezerwa szkodowa, krzywa Gompertza, model nieliniowy, spliny.