ON SOME MODIFICATIONS OF CAPE COD METHOD FOR LOSS RESERVING

Alicja Wolny-Dominiak

Abstract

In the paper we focus on the method for estimating IBNR (Incurred But Not Reported) loss reserves in non-life insurance based on growth curve modeling (Clark, 2003), called by author as the Cape Cod method. The literature proposes a wide variety of methods to estimate IBNR reserves, mostly using the chain-ladder technique. In this paper we investigate the method based on two-stage estimation of the expected amount of loss to emerge: the estimation of the ultimate loss by year and the estimation of the pattern of loss emergence through growth curve modelling, in which few restricted assumptions are adopted. Firstly, the method assumes that the actual incremental loss emergence follows an over-dispersed Poisson distribution and the coefficient of variance is constant. Secondly, as the pattern of loss to emerge, log-logistic and Weilbull growth curves are assumed. The aim of this paper is to assess the loss reserves applying the nonlinear model to fit the pattern of loss emergence, in which different form of growth curve is adopted as well as nonparametric spline fitting.

Key words: loss reserving, growth curve, nonlinear least square, spline

JEL Code: C20, C14

Introduction

The problem of loss reserves estimation in non-life insurance is crucial in the solvency of insurance company. There are a wide variety of methods to estimate IBNR reserves (simply loss reserving or claims reserving)proposed in literature, see e.g. (Wüthrich & Merz, 2008). The aim of this paper is to propose some modifications of the Cape Cod method for total loss reserve estimation based on growth curve modeling proposed by (Clark, 2003) and extended by (Zhang et al., 2012). Generally speaking, in first modification we use different algorithm for estimation of the parameters in growth curve. In the second modification we focu on the growth curve modeling. As the extention of applying two-parametric Weibull and log-logistic curves we use three parametric Gompertz curve and as additionally nonparametric spline fitting.

1 Cape Cod method for loss reserving

In loss reserving, cumulative loss amount is under consideration. Denote the random variable cumulative loss amount as Y. Consider a matrix $[y_i(t_j)]_{nxn}$ of the cumulative loss amount for insurance claims that occured in year i (accident year) and reported after t_j months after in the year j (development year). The elements $y_i(t_j)$ for $i + j \leq n+1$ are observed data while $y_i(t_j)$ for $i + j > n+1$ represent the future unobserved data. Such matrix is called loss triangle and has a general form

Fig. 1: The cumulated loss triangle

i	t_1	$\cdots$	t_{n-1}	t_n	p_i
1	$y_1(t_1)$	$\cdots$	$y_1(t_{n-1})$	$y_1(t_n)$	p_1
2	$y_2(t_1)$	$\cdots$	$y_2(t_{n-1})$		p_2
$\vdots$		$\iddots$			
n	$y_n(t_1)$				p_n

The additional column p_i represents earned premium in the accident year i. The reported losses are a sum of diagonal elements from the loss trangle. The aim of loss reserving is to estimate the ultimate loss Y for every accident i. In Cape Cod method proposed by (Clarki, 2003) this estimation is done by modeling aggregate loss amount based on ultimate loss by year i and the pattern of loss emergence. This pattern shows the percentage loss development from 0% to 100% in months and is modeled by well known growth curves, e.g. taken from biology. Assuming some accident year i and the same pattern of loss emergence for every $i = 1,..,n$, the loss reserve for year i is

$$Y_i = p_i \cdot u \cdot [G(t_j;\Theta) - G(t_{j-1};\Theta)] + \xi_i, \qquad (1)$$

where u is an ultimate loss ratio for the loss triangle and $G(t_j;\Theta)$ is the growth curve of cumulated losses with vector parameters Θ. In order to estimate parameters u and Θ in model (1), the Maximum Likelihood (ML) is used. To find out MLE estimators $\hat{u}$ and $\hat{\Theta}$

analytically without the use of an iterative algorithm, three strict assumptions are adopted. In first two, it is assumed that the loss in any period has a constant ratio $\frac{Variance}{Mean} = \sigma^2$ and σ^2 is known. Straight calculations gives

$$\hat{\sigma}^2 = \frac{1}{m-p} \sum_i \frac{(y_i - \bar{y})^2}{\bar{y}}, \tag{2}$$

where m is a number of reported cumulative losses and p is a number of parameters in growth curve. The third assumption is that incremental losses follow an over-dispersed Poisson distribution with probability function

$$P(Y = y) = \frac{\lambda^{\frac{y}{\sigma^2}} e^{-\lambda}}{\left(\frac{y}{\sigma^2}\right)!}. \tag{3}$$

Two first moments are then of the form $E(Y) = \lambda\sigma^2$ and $Var(Y) = \lambda\sigma^4$. Estimating parameters of (3) through maximum likelihood, the log-likelihood function is as follows:

$$l(\lambda, \sigma; y_1, ..., y_m) = \sum_i^m \log\left[\frac{\lambda^{\frac{y_i}{\sigma^2}} e^{-\lambda}}{\left(\frac{y_i}{\sigma^2}\right)!}\right] = \sum_i^m \left\{\frac{y_i}{\sigma^2} \log(\lambda) - \lambda - \log\left[\left(\frac{y_i}{\sigma^2}\right)!\right]\right\}. \tag{4}$$

As the parameter σ^2 is assumed known, the function (4) simplified to the form

$$l(\lambda; y_1, ..., y_m) = \sum_i^m [y_i \log(\lambda) - \lambda]. \tag{5}$$

Using (1), the log-likelihood is

$$l(u; y_1, ..., y_m) == \sum_i^m \{y_i \log(p_i \cdot u \cdot [G(t_j; \Theta) - G(t_{j-1}; \Theta)]) - p_i \cdot u \cdot [G(t_j; \Theta) - G(t_{j-1}; \Theta)]\}. \tag{6}$$

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

$$\hat{u} = \frac{\sum_{i=1}^m y_i}{\sum_{i=1}^m p_i \cdot [G(t_j; \Theta) - G(t_{j-1}; \Theta)]}. \tag{7}$$

Of course except $\hat{u}$, there is still a need to estimate the vector of parameters Θ. Under given $\hat{u}$ and $\hat{\Theta}$, total loss reserve under the Cape Cod method according to (1) is

$$\hat{Y} = \sum_i p_i \cdot \hat{u} \cdot [\hat{G}(t_j; \hat{\Theta}) - \hat{G}(t_{j-1}; \hat{\Theta})]. \tag{8}$$

The calculations are greatly simplified and the analytical derivation of MLE estimator $\hat{u}$ is possible under mentioned assumptions. In (Clark, 2003) two mathematical formulas of growth curve are applied, two-parametric Weibull and log-logistic. We extend this assumptions into three-parametric Gompertz curve. In this case the problem of estimation is rather hard to solve analytically, thus we use nonlinear least square (NLS) technique of parameters estimation for nonlinear model (1) in place of ML (Bates and Watts, 1988). The consequence of this is the normality assumption for incremental losses.

2 Growth curve fitting

As it can be seen in (1), there is a need to model the pattern of loss emergence in every accident year. We examine two approaches in modeling the pattern: one is model-based as in (Clark, 2003), but with three-parameter curve and the other is model-free spline fitting.

Considering model-based approach, since the cumulative losses in development years grow probably exponancially rather then linearly, the pattern can be described by growth curve. In literature, like in (Zwietering et al., 1990), the analytical forms of this type of curves are given. In (Clark, 2003), the pattern of loss emergence is modeled by two-parameters log-logistics and Weibull curves. The analytical form of these curves is as follows:

$$G_L(t;\Theta) = \frac{t^{\varpi}}{t^{\varpi} + \theta^{\varpi}}, \ \Theta = (\varpi, \theta) \tag{9}$$

$$G_W(t;\Theta) = 1 - \exp[-(\frac{t}{\theta})^{\varpi}], \ \Theta = (\varpi, \theta) \tag{10}$$

In our research we examine three-parameter Gompertz curve of the form:

$$G_G(t;\Theta) = \rho \exp[-\exp(\theta - \varpi t)], \ \Theta = (\varpi, \theta, \rho) \tag{11}$$

As the mathematical formula is known, the vector of parameters Θ can be estimated using a nonlinear regression model $y_i = G(t) + \xi$, where ξ is normally distributed disturbance and $E(\xi) = 0, Var(\xi) = \sigma_{\xi}^2$. Concerning methodology, we apply NLS. The problem in this model-based method is a proper specification of start values. To obtain these values we use local weighted regression (LWR), as in (Cleveland, 1979). Either NLS algorithm or LWR algorithm are implemented in R package {stats} (R Core Team, 2012), thus the estimation is easy to conduct.

The alternative approach in fitting the pattern of loss emergence is to abandon the parametric framework and avoid using models with the functional form fixed in advance. Instead, various nonparametric methods may be used such as the isotonic regression and Pool-

Adjacent-Violators Algorithm (PAVA) described by (de Leeuw et al., 2009) and (Gamrot, 2012), or spline functions discussed e.g. by (Judd, 1998). We presume the cubic spline as an example, but there is no obstacle to apply different smoothing. As in the case of parametric models (9)-(11), spline fitting is available in {stats}.

3 Modeling loss reserves – the case study

The loss triangle we examine is taken from (Mack, 1993). The cumulative version is presented in Table 1.

Tab. 1: Loss traingle (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	10000
2	352	1236	2170	3353	3799	4120	4648	4914	5339		10400
3	291	1292	2219	3235	3986	4133	4629	4909			10800
4	311	1419	2195	3757	4030	4382	4588				11200
5	443	1136	2128	2898	3403	3873					11600
6	396	1333	2181	2986	3692						12000
7	441	1288	2420	3483							12400
8	359	1421	2864								12800
9	377	1363									13200
10	344										13600

Source: (Mack, 2003)

In the first step of our estimation procedure we assess parameters of the growth curve from (9)-(11).

Tab. 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 Loglogistic	1,37	0,10	13,41	0,00
θ Loglogistic	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

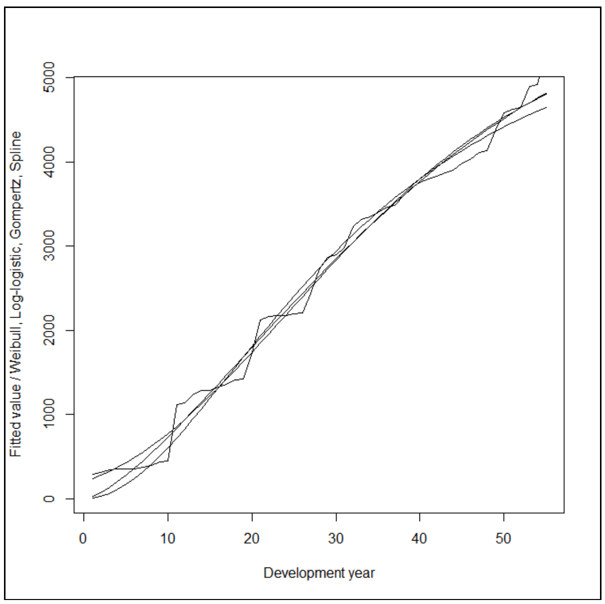
In results presented in Table 2 all parameters are statistically significant at level 5%. The fitted pattern of loss emergence is then as in Table 3

Tab. 3: Fitted patterns of loss emergence

Development year j	$\hat{G}_W(t)$	$\hat{G}_L(t)$	$\hat{G}_G(t)$	$\hat{S}$ - Spline
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 graphical presentation of fitted growth curves is shown in Fig. 2.

Fig. 2: Plot of fitted patterns of loss emergence



In every model, the goodness-of-fit statistics is calculated using (2), similarly as in (Clark, 2003). The definitely lowest value is for spline $\hat{\sigma}_s^2 = 0.81$ in comparing to three other models: $\hat{\sigma}_G^2 = 18.37$, $\hat{\sigma}_W^2 = 81.63$, $\hat{\sigma}_L^2 = 281.4$. Therefore, in further calculations we adapt $\hat{S}$ values from Table 3.

Tab. 4: Estimated loss reserves

$p_i \cdot \hat{u}$	Growth - spline	Reported	Estimated reserves $\hat{y}_i$
5 791,46	17,30%	3 901	1 001,93
6 023,12	18,57%	5 339	1 118,46
6 254,78	22,88%	4 909	1 430,98
6 486,44	25,52%	4 588	1 655,26
6 718,09	28,25%	3 873	1 898,05
6 949,75	39,33%	3 692	2 733,53
7 181,41	49,71%	3 483	3 569,80
7 413,07	58,12%	2 864	4 308,72
7 644,73	69,73%	1 363	5 330,88
7 876,39	84,08%	344	6 622,85
Total 68 339,23			**Total** 29 670,45

Finally applying (6), (7) and growth curve with spline, the estimated loss reserve is in total $\hat{Y} = 29\,670\,450$.

Conclusion

The estimation of loss reserves using growth curve modeling is helpful alternative to investigate the pattern of losses emergence. The nonlinear model gives flexible method of estimation, in which changing the way of fitting growth curve is straighforward. In our research the normality of the disturbance was assumed, but the other distributions are possible by using generalized nonlinear least square. In ongoing research the variability of total loss reserve is under consideration.

Acknowledgment

This research is supported by the grant of Polish National Science Centre (nr NN 111461540).

References

Bates, M., Watts, G. (1988). Nonlinear 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*, *03*(Fall), 41-91.

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

Mack, T. (1993). Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates. *ASTIN Bulletin, 23,* 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), 361-366.

Judd, K. L. (1998). Numerical Methods in Economics. Cambridge MS: MIT Press.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*, 1-24.

Gamrot, W. (2012). Simulation-assisted Horvitz-Thompson statistic and isotonic regression. In J. Ramik, D. Stavarek (Eds.), *Proceedings of 30th International Conference Mathematical Methods in Economics 2012* (pp. 207-212). Karvina: Silesian University of Opava, School of Business Administration.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

Zhang, Y., Dukic, V. & Guszcza, J. (2012). A Bayesian non-linear model for forecasting insurance loss payments. *Journal of Royal Statistics, 175*(Part2), 637-656.

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

Contact

Alicja Wolny-Dominiak

University of Economics in Katowice, Poland

ul. 1 Maja 50

40-287 Katowice

alicja.wolny-dominiak@ue.katowice.pl