Modeling non-life insurance claims

Silvia Yordanova
Sofia University, Sofia, Bulgaria
Pavel Stoynov
Sofia University, Sofia, Bulgaria
todorov@feb.uni-sofia.bg

Abstract: The aim of the paper is to analyse claims received by an Bulgarian insurance company and to model claim count per period and claim size.

Key words: insurance, claim size, claim count

1. Introduction.

In a general insurance context, the typical evolution of a claim may be divided into three parts. From the occurrence of the claim to its notification to the insurance company, the claim is said to be Incurred But Not Reported (IBNR). After notification, the claim is known by the company and there may be some time before the complete payment is made. We call this claim Reported But Not Settled (RBNS). Other acronyms are IBNFR (Incurred But Not Fully Reported), RBNFS (Reported But Not Fully Settled) etc.

Typically, several payments are done for one claim but to simplify the settings we allow only one payment.

With the introduction of Solvency 2 (in 2012) and IFRS 4 Phase 2 (in 2013), the evaluation of size and inter-arrival time of claims as well as of regulatory required solvency capital becomes more important and there are different trials to improve, adjust or extend current techniques for loss reserving.

Generally, these claims reserving techniques are based on aggregated data, conveniently summarized in a run-off triangle per accident year and per development year. The chain-ladder approach (Mack's model in Mack (1993) and Mack (1999)) is the most popular member of this category. A rich literature exists about those techniques and an overview can be found in England & Verrall (2002) or Wuthrich & Merz (2008). However, using aggregated data in combination with the chain-ladder approach gives rise to several issues which are enumerated in Antonio & Plat (2011). Many practical solutions have been proposed, but they have not been applied simultaneously.

A mathematical framework for the development of individual claims was formulated in the last decade of the 20th century by Arjas (1989), Norberg (1993), Haastrup & Arjas (1996) and Norberg (1999). More recently, a semi-parametric model (Zhao & Zhou & Wang (2009)) and a so-called micro-model (Antonio & Plat (2011)) have been introduced.

In this paper, the authors model the individual claim size and number of claims per period. The paper is organized as follows. The statistical model is introduced in Section 2. In Section 3, the data and the assumptions are presented. Results are presented in Sections 4. Finally, Section 5 concludes.

2. The Model.

The input of the model should be a database containing detailed information about the development of individual claims. More specifically, the model will use:

- the occurrence time of every claim;
- the declaration time of every claim;
- the time of payment done for every claim;
- the amount paid for every claim.

The data set is from a European insurance company and concerns a portfolio of general liability insurance policies for private individuals. More details on the data can be found in Section 3.

The Time Structure of the model includes:

- the random variables Q_i which are the reporting delay for claim i, i. e. the difference between the occurrence time of the claim and the time of its notification to the insurance company;
- the random variables T_i which are the payment delay for claim i, i. e. the difference between the notification time and the payment time.

The number of claims K_j for the occurrence period j is supposed to follow a Poisson process with occurrence measure λ_j.

The claim size Y_i for claim $i,$ i. e. the amount paid by insurer for this claim, is supposed to have normal or skewed normal distribution.

3. The Data and the assumptions.

The Database include 9168 fully paid claims to a Bulgarian insurer originated from housing insurance between 2009 and 2015 year. So, the paper focuses on paid claims and not on incurred losses (paid losses plus case estimates) but incorporating information on incurred losses could be a topic for future research. No claim handling expenses are assumed in the data set since we had no information on expenses.

To estimate reserves, the payments should be discounted. However, attention is paid mainly on claim size and time parameters estimationin, so discounting and reserve estimation may be done as a following step of the research.

The distribution of the claims by years is presented in Table 1.

Table 1. Paid claims by years.

Year	Count of claims paid
2009	711
2010	831
2011	562
2012	1403
2013	1110
2014	2615
2015	1936
Total	9168

Occurrence of claims. The random variable K_j representing the number of claims with positive payment(s) for the occurrence period j is supposed to follow a Poisson distribution.

In the model presented in the current paper, the logarithm of the claim size, Y_i is supposed to follow a Univariate Skew Normal (USN) distribution which is introduced below.

The Univariate Skew Normal distribution has been fragmentarily introduced in Roberts & Geisser (1966), but the first formal definition and systematic study of its properties appeared in the seminal work of Azzalini (1985).

A random variable X with probability density function given by

$$f_X(x) = 2\phi(x)\Phi(\alpha x),$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the probability density function of standard normal distribution and $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$ is the cumulative distribution function of the standard normal distribution is called standard skewed normal distribution. The parameter α is a shape parameter usually called skewness. If $\alpha = 0,$ the distribution co-insides with the standard normal distribution. When $\alpha > 0,$ the distribution is right skewed. When $\alpha < 0,$ the distribution is left skewed.

In the general case, skewed normal distribution with three parameters: location $(\mu),$ scale (σ) and skewness $(\alpha).$ In this case, the probability density formula is

$$f_X(x) = 2\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha\frac{x-\mu}{\sigma}\right).$$

Many extensions of the Univariate Skew Normal distribution to the multivariate case have been suggested by diferent authors (among others, see Azzalini (1985) and Azzalini & Dalla Valle (1996)).

4. The Results.

The parameter of the distribution of the random variable K_j (number of claims per period) observations are considered for each period where period can be either year or month. The graphical representation of the claims per year is given at Figure 1.

The histogram of claims per months is given at Figure 2.

The number of the claims per month for 2009 year is given in Table 2.

The number of the claims per month for 2010 year is given in Table 3.
The number of the claims per month for 2011 year is given in Table 4.

Figure 1. The graphical representation of the claims per year.

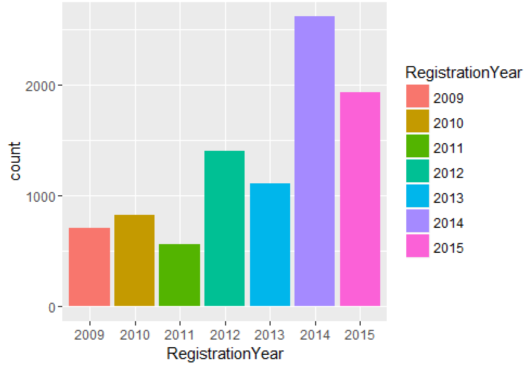


Figure 2. The histogram of the claims count per month for the period 2009-2015.

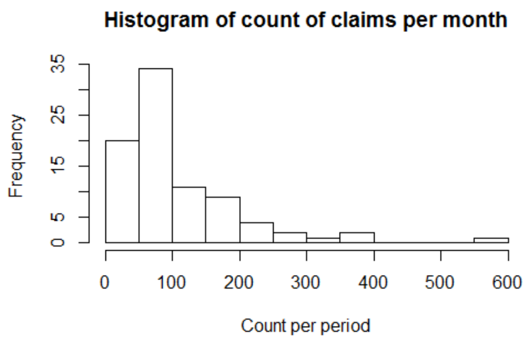


Table 2. Paid claims per month in 2009.

Month	Count of claims paid
1	65
2	62
3	62
4	48
5	33
6	81
7	109
8	70
9	48
10	43
11	38
12	52
Total	711

Table 3. Paid claims per month in 2010.

Month	Count of claims paid
1	41
2	138
3	75
4	44
5	93
6	53
7	90
8	47
9	33
10	106
11	54
12	57
Total	831

Table 4. Paid claims per month in 2011.

Month	Count of claims paid
1	24
2	26
3	42
4	41
5	36
6	48
7	79
8	64
9	31
10	98
11	36
12	37
Total	562

Table 5. Paid claims per month in 2012.

Month	Count of claims paid
1	41
2	227
3	147
4	67
5	365
6	193
7	83
8	67
9	35
10	60
11	59
12	59
Total	1403

Table 6. Paid claims per month in 2013.

Month	Count of claims paid
1	72
2	73
3	154
4	80
5	91
6	119
7	134
8	96
9	65
10	99
11	70
12	56
Total	1110

Table 7. Paid claims per month in 2014.

Month	Count of claims paid
1	86
2	95
3	97
4	144
5	314
6	382
7	571
8	254
9	203
10	202
11	114
12	153
Total	2615

Table 8. Paid claims per month in 2015.

Month	Count of claims paid
1	160
2	245
3	254
4	156
5	160
6	184
7	103
8	175
9	177
10	135
11	122
12	65
Total	1936

The number of the claims per month for 2012 year is given in Table 5.
The number of the claims per month for 2013 year is given in Table 6.
The number of the claims per month for 2014 year is given in Table 7.
The number of the claims per month for 2015 year is given in Table 8.
The density of claim count per month is presented at Figure 3.
The histogram of claim size for 2009 is presented at Figure 4.
The histogram of claim size for 2010 is presented at Figure 5.
The histogram of claim size for 2011 is presented at Figure 6.
The histogram of claim size for 2012 is presented at Figure 7.
The histogram of claim size for 2013 is presented at Figure 8.
The histogram of claim size for 2014 is presented at Figure 9.
The histogram of claim size for 2015 is presented at Figure 10.

Figure 3. The density of claim count per month

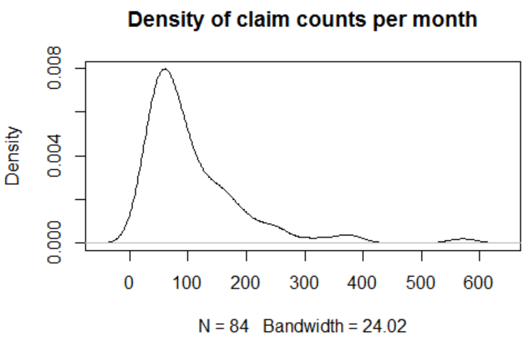


Figure 4. The histogram of claim size for 2009

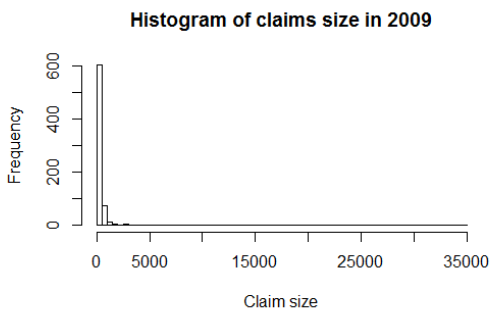


Figure 5. The histogram of claim size for 2010

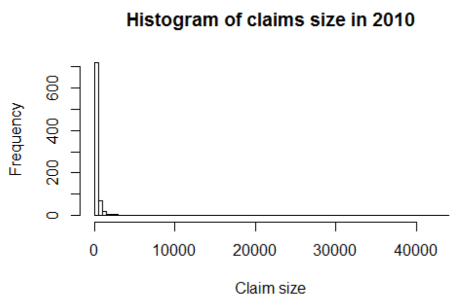


Figure 6. The histogram of claim size for 2011

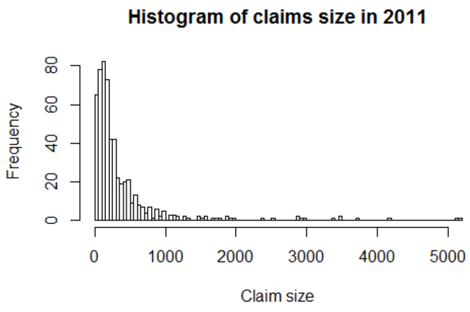


Figure 7. The histogram of claim size for 2012

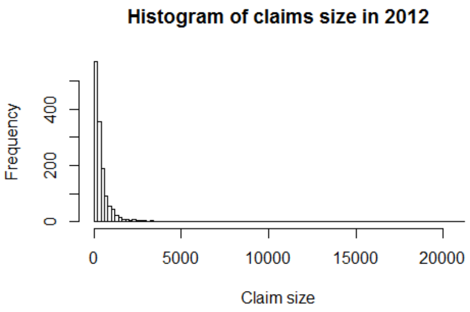


Figure 8. The histogram of claim size for 2013

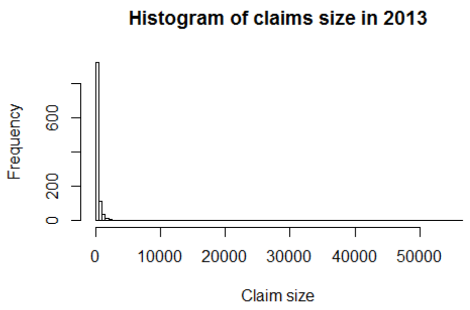


Figure 9. The histogram of claim size for 2014

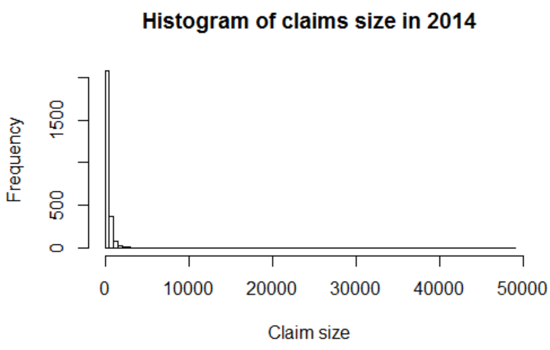
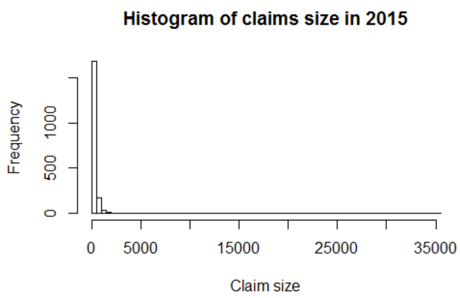
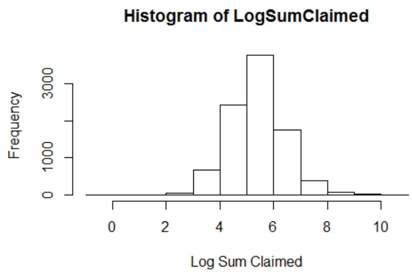


Figure 10. The histogram of claim size for 2015



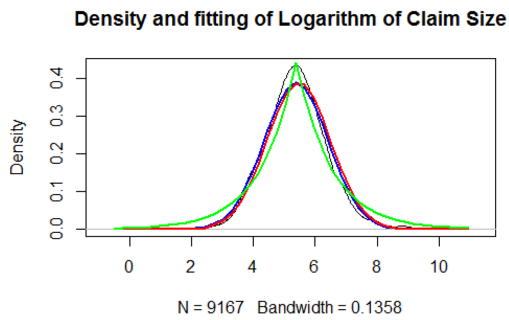
Histogram of logarithm of claim size is presented at Figure 11.

Figure 11. Histogram of logarithm of claim size



The density of the logarithm of the claim size and its fitting with univariate skewed normal distribution are presented at Figure 12.

Fig 12. Density of the logarithm of the claim size and its fitting with univariate skewed normal distribution.



It shows that non-skewed normal distribution gives a good fitting.

5. Conclusions

The study in this paper shows that logarithm of the claim size of house insuranse policies can be fitted by normal distribution without skewness, the number of claims per year or per month can be modeled by Poisson distribution and inter-arrival time between two consecutive claims can be modeled by exponential distribution.

References:

Antonio, K. and R. Plat, R. (2011) Micro-level stochastic loss reserving. Working paper.

Arjas, E. (1989). The claims reserving problem in non-life insurance: some structural ideas ASTIN Bulletin, vol. 19, No. 2, 139-152.

Azzalini, A. (1985). A class of distribution which includes the normal ones. Scandinavian Journal of Statistics, vol. 12, 171-178.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. Biometrika, vol. 83, 715-726.

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. British Actuarial Journal, vol. 8, 443-544.

Haastrup, S. and Arjas, E. (1996). Claims reserving in continuous time: a nonparametric Bayesian approach. ASTIN Bulletin, vol. 26, No. 2, 139-164.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN Bulletin, vol. 23, 213-225.

Mack, T. (1999). The standard error of chain-ladder reserve estimates: recursive calculation and inclusion of a tail-factor. ASTIN Bulletin, vol. 29, 361-366.

Norberg, R. (1993). Prediction of outstanding liabilities in non-life insurance. ASTIN Bulletin, vol. 23, No. 1, 95-115.

Norberg, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. ASTIN Bulletin, vol. 29, No. 1, 5-25.

Roberts, C. and Geisser, S. (1966). A necessary and su_cient condition for the square of a random variable to be gamma. Biometrika, vol. 53, No. 1/2, 275-278.

Wuthrich, M. V. and Merz, M. (2008). Stochastic claims reserving methods in insurance. Wiley Finance.

Zhao, X. B. and Zhou, X. and Wang, J. L. (2009). Semiparametric model for prediction of individual claim loss reserving. Insurance: Mathematics and Economics, vol. 45, No. 1, 1-8