ESTIMATION OF INCURRED BUT NOT YET REPORTED CLAIMS BASED ON POISSON DISTRIBUTED REPORTING DELAY

MICHAL GERTHOFER, VÁCLAV SLÁDEK AND PAVEL ZIMMERMANN
University of Economics Prague, Faculty of Informatics and Statistics,
Department of Statistics and Probability, nám. Winstona Churchilla 4, Prague 3, Czech Republic
email: germ00@vse.cz, xslav20@vse.cz, zimmerp@vse.cz

Abstract

A claim frequency model, which estimates number of claims that incurred but are not reported (IBNR), is a key component of an insurer's future liability prediction. Several possible modelling approaches were developed in the past. Chain ladder method is a particularly popular distribution free method, which models development of the cumulative liability through weighted averages of the past development factors. The assumption of this model is that developments of the cumulative claims for each occurrence year are independent. If a trend is identified in occurrence years, predictions are biased. A modification of this method can be applied to relax this assumption. For example, development factors, can be extrapolated to the future by a trend function. Other approaches may be based e.g. on variety of regression specifications. In this contribution, a new approach is presented, which is based on the assumption that number of discrete years until the claim is registered has Poisson distribution. The mean of this distribution is estimated by maximum likelihood method, taking in to account that observations are truncated. An empirical study is presented and our approach is compared with the traditional Chain ladder method as well as the above-mentioned modification.

Key words: *frequency model, Chain ladder, modified Chain ladder, Poisson distribution*

JEL Codes: *G22*

DOI: 10.15611/amse.2017.20.14

1. Introduction

Insurance companies have to estimate its future liability from its business. Estimation is based on statistical and probabilistic methods. One of possible approaches is based on frequency and severity models. This paper is focused on a claims frequency model which estimates number of claims incurred but not reported. One of the basic methods for frequency model is the Chain ladder method described for example in (Boland, 2007) or (Gisler and Wüthrich, 2008). Another approach could be Poisson or Negative Binomial regression (Garrido and col., 2016) and (Bartoszewicz, 2005) which are special cases of Generalized linear models (GLM). These models allow predicting a number of claims incurred but not reported based on dependent variables (age of policyholders, type of vehicle etc.). (Garrido and col., 2016) shows that GLM is also useful when dependence between frequency and severity exists. Poisson and Negative Binomial models were also used in (David and Jemna, 2015).

The frequently used and preferred method is Chain ladder, a non-parametric method with otherwise strict assumptions. Breaking this assumption leads to inaccurate inference. A solution could be a modification of Chain ladder method or completely different approach. The new approach, presented in this paper, is based on assumption that number of discrete years until the claim is reported has Poisson distribution. The parameter of this distribution is estimated by maximum likelihood method.

The aim of this paper is to compare results obtained by Chain ladder with results obtained by its modification and the new parametric approach. Methods are applied to a dataset from Czech Insurers´ Bureau which is described in Section 2. Chain ladder method and its modification are defined in Section 3. The new parametric approach is defined in Section 4. Section 5 contains the application of all considered methods on the dataset. Results comparison and conclusions are in Section 6.

2. Data

Dataset comes from the Czech Insurers´ Bureau, an organization of insurance companies that are authorized to provide liability insurance for damage caused by vehicles (car insurance). The dataset contains records of each individual claim, which occurred between years 2005 and 2015. We only used material claims in our application. Dataset provides information about an accident year, payments amount and year of the first payment about each claim.

Disadvantages of this dataset are missing claims reported but without the first payment and missing year of reporting of the claims so it is not possible to construct triangle where development year is time to report after occurrence of claim. We consider development year in this problem as time to the first payment.

Generally, a claim incurred but not reported is a claim which already occurred but a policyholder has not yet reported it to the insurer (delay in reporting the claim could be weeks, months even years). The insurance company has to booked reserves for these claims based on legislation. In this paper, methods estimate not only claims incurred but not reported but also claims incurred, reported but without first payment (insurance company already has these claims in the database). Let us denote all these claims as IBNR for simplicity.

Table 1: Number of claims reported and with the first payment by accident year

Accident year	Number of claims	Accident year	Number of claims
2005	3,265	2011	2,158
2006	3,065	2012	1,963
2007	3,055	2013	1,944
2008	3,240	2014	1,854
2009	2,560	2015	1,261
2010	2,472		

Source: the author’s work

3. Chain ladder method

Chain ladder method is based on a development triangle. A development triangle is a tool which ordered historical claims by its accident year and by development year. Accident year is a year when a claim occurs, development year is the delay in reporting relative to an accident year. Table 2 represents an example of a triangle where $N_{i,j}$ is the cumulative number of claims occurred in accident year i and reported until development year j.

The aim of Chain ladder is to fill triangle under main diagonal (lower triangle) using development factors f_j, which are estimated from the upper triangle. The basic assumption of

Chain ladder is independence of developing claims over accident years. (Mack, 1993) proved that this assumption provides unbiased prediction.

The estimates of development factors are defined as

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j} N_{i,j}}{\sum_{i=1}^{n-j+1} N_{i,j-1}}, \quad 1 \leq j \leq n-1 \tag{1}$$

where n is the number of accident years. An unbiased estimate of the cumulative number of claims occurred in accident year i and in development year j is obtained by the formula

$$\hat{N}_{i,j} = \hat{f}_j \, N_{i,j-1}. \tag{2}$$

Table 2: Cumulative triangle of the number of claims

	Development year (j)				
Accident year (i)	0	⋯	j	⋯	n-1
1	$N_{1,0}$	⋯	$N_{1,j}$	⋯	$N_{1,n-1}$
⋮	⋮	⋮	⋮	⋮	
i	$N_{i,0}$	⋯	$N_{i,j}$		
⋮	⋮	⋮			
n	$N_{n,0}$				

Source: the author's work

Let us denote claims reported with the first payment but not settled as $R = \sum_{i=1}^{n} N_{i,n-i}$ and ultimate claims (claims with and without the first payment) as $U = \sum_{i=1}^{n} \hat{N}_{i,n-1}$. Estimated number of IBNR claims is then the difference between U and R.

4. Modification of Chain ladder method

If the assumption of Chain ladder is violated, its estimate is biased. (Overestimate or underestimate the reality). The insurance company is then over reserved or is not solvent, both cases are unfavorable for any company. Breach of this assumption can be identified by individual development factors $f_{i,j} = N_{i,j}/N_{i,j-1}$ which represent development of the number of claims for each combination of accident year i and development year j. If the trend is obvious in a series of development factors for a development year j, the assumption is not fulfilled. Several approaches exist how to deal with this problem. This article is focused only on predicting the development factors with a regression method, we call it modified Chain ladder.

Estimation of development factor for development year j is based on a regression model

$$Y \sim f(X,\beta),$$

where $f_{i,j}$ are dependent variable, independent variable is the accident year and β are regression coefficients. There is a need to correctly identify the trend function (linear, parabolic etc.). The estimate of development factor is then a prediction for the following accident year. This method is used only for those development years where the assumption is violated. Others development factors are estimated by (1).

5. Poisson model of time until the first payment

Let us consider that time until the first payment, measured as a count of calendar years between date of occurrence and date of the first payment, is a discrete variable X_i with Poisson distribution and an unknown parameter dependent on the accident year *i* denoted as λ_i.

The problem is that observations in the dataset are truncated. Truncated observations are observations with a value which can occur with non-zero probability but cannot be observed. So, a sample contains only observations from a restricted part of the population. In this actuarial application, observations are collected only up to a certain time point, denoted as τ, representing the date of extraction of the dataset. For each accident year i, we can only observe claims with at least one payment before τ. Maximum discrete number of years between claim occurrence and the first payment, which will be denoted t_i, is the difference between the calendar year of τ and the accident year *i*. Probability distribution function of truncated distribution is defined as

$$P(X_i = x_i | X_i \leq t_i) = \frac{P(x_i)}{F(t_i)},\ for\ i = 1,\ldots,\tau \tag{3}$$

where $F(\cdot)$ is a cumulative distribution function of a random variable with Poisson distribution. The estimate of λ_i based on the maximum likelihood method is obtained by maximization the log likelihood function

$$\sum_{j=1}^{n_i} \ln\left(\left(e^{-\lambda_i}\frac{\lambda_i^{x}}{x_j!}\right)\bigg/\left(e^{-\lambda_i}\sum_{x_j=0}^{t_i}\frac{\lambda_i^{x_j}}{x_j!}\right)\right), \tag{4}$$

where $x_j \leq t_i$ and n_i is number of claims in accident year *i*.

The parameter λ_i for the last accident year cannot be estimated by using the maximum likelihood method. Because, if t_i is equal 0, changing the value of λ_i does not change the value of function (4), algorithm cannot find maximum of this function. This parameter can be estimated by extrapolating the estimated parameters $\hat{\lambda}_i$ with a regression model (linear, parabolic, hyperbolic etc.) similarly as in the modified Chain ladder.

The estimated number of claims occurred in accident year *i* and reported until development year *j* is obtained by formula

$$\hat{N}_{i,j} = \frac{N_i}{F(t_i)}F(j), \tag{5}$$

where N_i is the number of claims occurred in accident years *i* and reported until the development year t_i. $F(\cdot)$ is a cumulative distribution function of a random variable with Poisson distribution with the parameter λ_i. The ratio $\frac{N_i}{F(t_i)}$ is an estimate of the number of all claims incurred in accident year *i*.

6. Application mentioned methods

The all above-mentioned methods are applied on the portfolio of claims specified in Section 2. The goal is to estimate the number of claims before the first payment, denoted in this paper as IBNR.

6.1. Chain ladder

The dataset contains claims occurred in the time period from 1.1.2005 to 31.12.2015. Table 3 represents incremental triangle of our portfolio. We can see, that only several claims have the first payment in development year 3 or higher. Most of the claims have the first payment until next year after its occurrence. Numbers of claims in the first and second development year suggest that the assumption of Chain ladder is probably violated. The number of claims decreases in the second development year over accident years while as in the first development year no clear trend is observed. But we will ignore that fact and estimate IBNR and Ultimate using Chain ladder.

Table 3: Incremental triangle of the claim numbers in portfolio

Accident year	Time until the first payment 0	1	2	3	4	5	6	7	8	9	10
2005	1,021	1,918	254	63	8	1	0	0	0	0	0
2006	1,080	1,736	191	51	3	2	1	0	1	0	
2007	1,304	1,430	268	43	9	1	0	0	0		
2008	1,411	1,590	191	36	9	2	1	0			
2009	1,181	1,225	101	39	13	1	0				
2010	1,274	995	156	37	10	0					
2011	1,059	962	98	35	4						
2012	989	844	105	25							
2013	1,089	730	125								
2014	1,307	547									
2015	1,261										

Source: the author’s work

Table 4: Cumulative triangle of the claim numbers in portfolio

Accident year	Time until the first payment 0	1	2	3	4	5	6	7	8	9	10
2005	1,021	2,939	3,193	3,256	3,264	3,265	3,265	3,265	3,265	3,265	3,265
2006	1,080	2,816	3,007	3,058	3,061	3,063	3,064	3,064	3,065	3,065	
2007	1,304	2,734	3,002	3,045	3,054	3,055	3,055	3,055	3,055		
2008	1,411	3,001	3,192	3,228	3,237	3,239	3,240	3,240			
2009	1,181	2,406	2,507	2,546	2,559	2,560	2,560				
2010	1,274	2,269	2,425	2,462	2,472	2,472					
2011	1,059	2,021	2,119	2,154	2,158						
2012	989	1,833	1,938	1,963							
2013	1,089	1,819	1,944								
2014	1,307	1,854									
2015	1,261										

Source: the author’s work

Based on the cumulative triangle in Table 4 and (2), we estimate the development factors. They are shown in Table 5 and the filled incremental triangle is shown in Table 6. The bold values in Table 6 represent the predicted IBNR.

Table 5: Development factors for the considered portfolio

Development year	1	2	3	4	5	6	7	8	9	10
$\hat{f}_j$	2.02	1.07	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Source: the author’s work

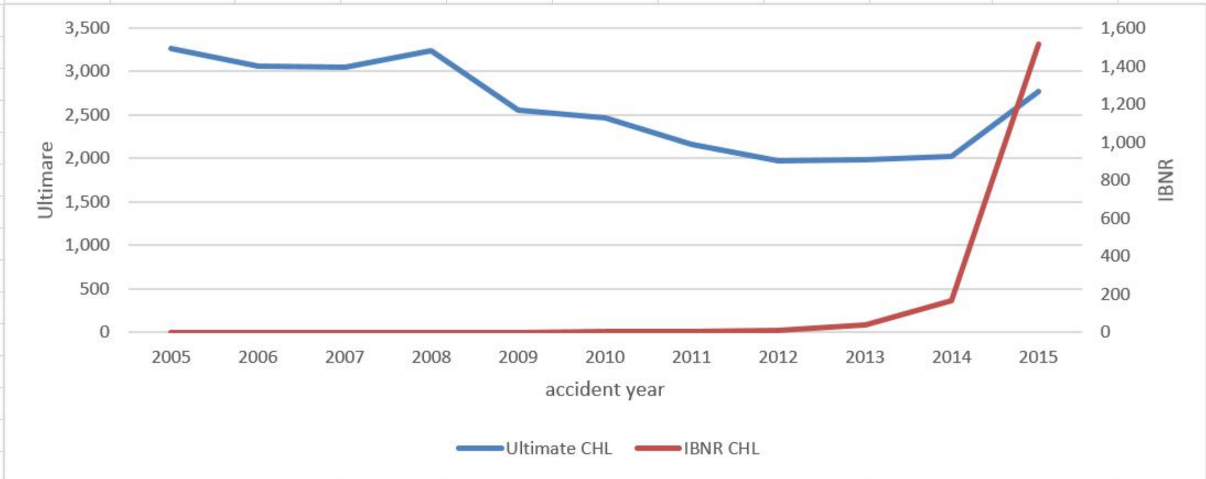
The predicted number of claims based on Chain ladder method is 1,725. We can see that estimated number of all claims occurred in 2015 is much higher than estimated number of all claims occurred in previous accident years. The difference is evident especially in the development year j = 1 where the predicted number of claims is 2.36 times higher than in 2014. It is obvious that Chain ladder overestimates the reality for this portfolio. If an insurance company accept these results, it will hold more reserves than necessary. On Figure 1 we can see that the predicted number of all claims increased for accident year 2015, which is not consistent with the observed decreasing trend in previous periods.

Table 6: Filled incremental triangle based on Chain ladder

	Time until the first payment										
Accident year	0	1	2	3	4	5	6	7	8	9	10
2005	1,021	1,918	254	63	8	1	0	0	0	0	0
2006	1,080	1,736	191	51	3	2	1	0	1	0	**0**
2007	1,304	1,430	268	43	9	1	0	0	0	**0**	**0**
2008	1,411	1,590	191	36	9	2	1	0	**0**	**0**	**0**
2009	1,181	1,225	101	39	13	1	0	**0**	**0**	**0**	**0**
2010	1,274	995	156	37	10	0	**0**	**0**	**0**	**0**	**0**
2011	1,059	962	98	35	4	**1**	**0**	**0**	**0**	**0**	**0**
2012	989	844	105	25	**6**	**1**	**0**	**0**	**0**	**0**	**0**
2013	1,089	730	125	**30**	**6**	**1**	**0**	**0**	**0**	**0**	**0**
2014	1,307	547	**126**	**30**	**6**	**1**	**0**	**0**	**0**	**0**	**0**
2015	1,261	**1,289**	**174**	**42**	**8**	**1**	**0**	**0**	**0**	**0**	**0**

Source: the author's work

Figure 1: Development of Ultimate and IBRN according to an accident year



Source: the author's work

6.2. Modified Chain ladder

Individual factors give us information about breaking the assumption of Chain ladder. The assumption is broken if the pattern of developing claims is similar over accident years. Figure 2 shows the development in time of the first five individual factors. Values of the first development factor can be found on the main axis (left y-axis) and values of the other development factors can be found on the minor axis (right y-axis). It is obvious that the first and the second development factor decrease in time, other factors are constant. The assumption is violated, so modified Chain ladder should be used.

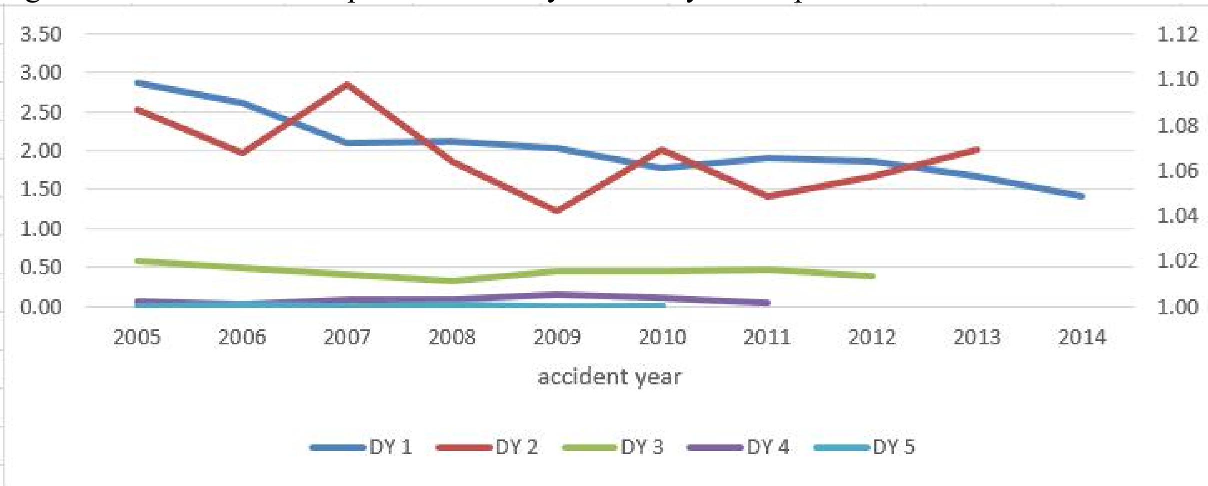
As a trend occurs for the first and the second development year, we apply the regression models for estimation of their values. For development years j>2 the Chain ladder estimator (1) was applied. The development of the factors is approximately linear. The estimated regression equation for the first development factor is

$$\hat{f}_{i,1} = 2.76 - 0.13i,$$

and for the second development factor is

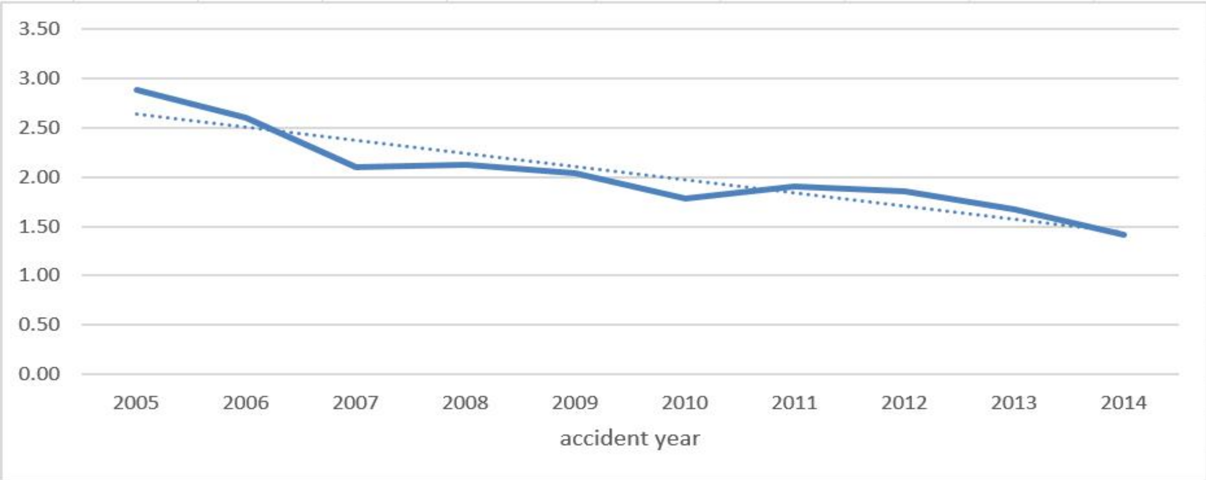
$$\hat{f}_{i,2} = 1.08 - 0.0033i.$$

Figure 2: Individual development factors by accident year for portfolio considered.



Source: the author's work

Figure 3: First development factor and its fitted values (dotted line)

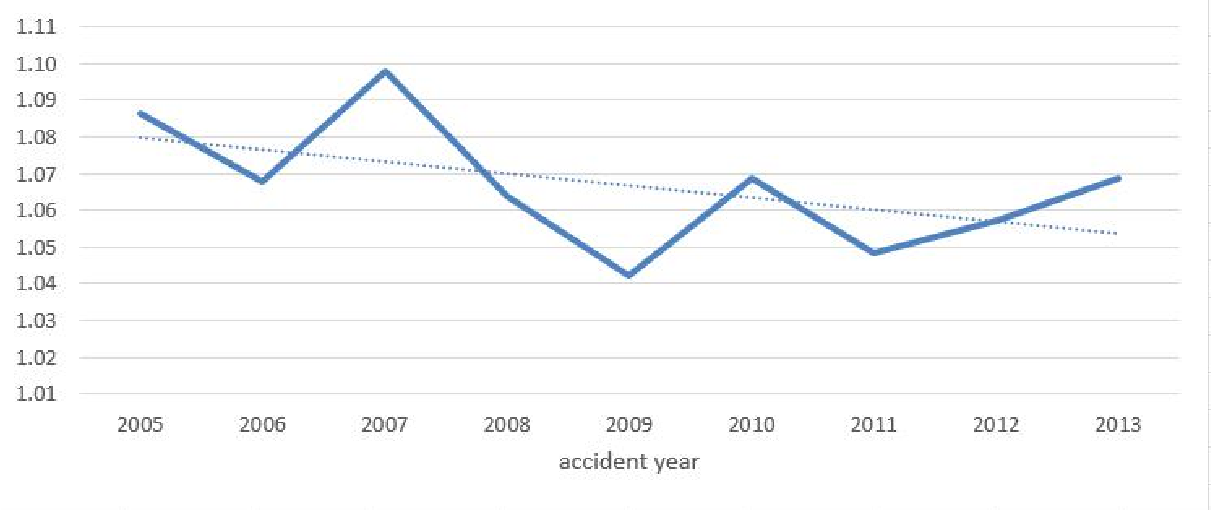


Source: the author's work

Development factors of modified Chain ladder (MCHL) are different from development factors of Chain ladder (CHL) in the first and the second development factor. The first development factor of CHL has value 2.02 and factor of MCHL has value 1.31. The second development factor of CHL has value 1.07 and factor of MCHL has value 1.05.

Table 7 contains filled incremental triangle by this approach. IBNR for accident years 2005-2013 remain unchanged in comparison with CHL, because the developments factors are the same. IBNR claims occurred in 2014 are 130 and IBNR claims occurred in 2015 are 507, almost 3 times lower against Chain ladder. Ultimate of MCHL is 1,768, which is 1.57 times lower against Chain ladder. Maybe, this result is more optimistic than reality. We can see, that Ultimate for accident years 2012-2013 are on the same level and IBNR for 2014 and 2015 are lower.

Figure 4: Second development factor and its fitted values (dotted line)



Source: the author's work

Table 7: Filled incremental triangle by modified Chain ladder

Accident year	Time until the first payment 0	1	2	3	4	5	6	7	8	9	10
2005	1,021	1,918	254	63	8	1	0	0	0	0	0
2006	1,080	1,736	191	51	3	2	1	0	1	0	**0**
2007	1,304	1,430	268	43	9	1	0	0	0	**0**	**0**
2008	1,411	1,590	191	36	9	2	1	0	**0**	**0**	**0**
2009	1,181	1,225	101	39	13	1	0	**0**	**0**	**0**	**0**
2010	1,274	995	156	37	10	0	**0**	**0**	**0**	**0**	**0**
2011	1,059	962	98	35	4	**1**	**0**	**0**	**0**	**0**	**0**
2012	989	844	105	25	**6**	**1**	**0**	**0**	**0**	**0**	**0**
2013	1,089	730	125	**30**	**6**	**1**	**0**	**0**	**0**	**0**	**0**
2014	1,307	547	**93**	**30**	**6**	**1**	**0**	**0**	**0**	**0**	**0**
2015	1,261	**391**	**83**	**27**	**5**	**1**	**0**	**0**	**0**	**0**	**0**

Source: the author's work

6.3. Poisson model

The last approach to estimate the number of IBNR claims is the new approach, Poisson model of a number of development years until the first payment. This method is not based on development factors, i.e. is not recursive. Instead, parameters of a Poisson distribution are fitted for each accident year taking into account truncation of the data collected.

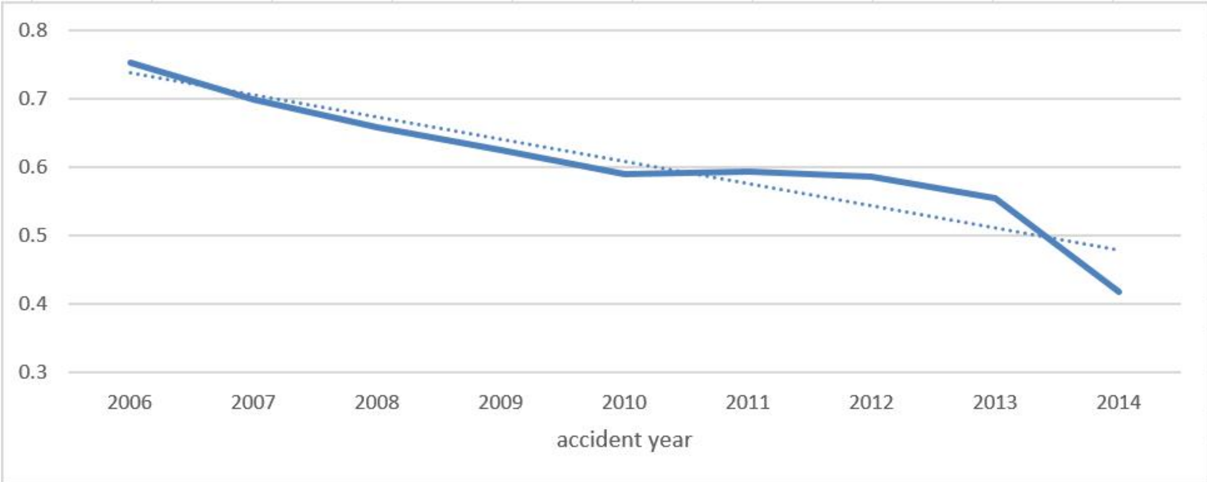
The unknown parameters are estimated by maximum likelihood method. As we mentioned in Section 4, a parameter for the accident year 2015 has to be estimated by another approach. Using a regression model, this value was extrapolated from fitted values in previous accident years.

Linear trend estimated for λ_i is in Figure 5. After a slight increase above the trend in 2010, the estimates decreased below the trend in 2015. In our opinion, it is temporary fluctuation and this rapid decrease will not continue in upcoming periods. Estimated parameters are shown in Table 8.

The accuracy of estimates was tested using the method in the sample. We compared the fitted number of claims for each development year with the real number of claims. Differences between real and fitted values are high for the first accident years (2006-2009).

This inaccuracy is irrelevant for prediction because the number of claims in development year 5 and higher is marginal (based on claims occurred in period 2005-2009). From the year 2010, the accuracy is much better. If we use for modeling only period 2010-2014, value of estimate λ_i for accident year 2015 is 0.434. Difference against using all observations for modelling is only 0.012. We decided to use all accident years, because we obtain little bit more conservative results.

Figure 5: λ_i by the accident year and its linear trend (dotted line)



Source: the author's work

Table 8: Parameters of Poisson distributions

Accident year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
λ_i	0.753	0.699	0.658	0.625	0.590	0.594	0.586	0.554	0.418	0.446

Source: the author's work

Table 9 shows filled incremental triangle of considered portfolio, bold values are predicted IBNR claims. We can see, that development of a number of claims in accident years 2014 and 2015 is similar. The total number of claims occurred in 2015 is 1,971. If we use the second estimate of λ_{2015}, (0.434), the total number of claims is "only" 1,946. The impact of using λ_{2015} equal 0.446 or 0.434 on predicted number of claims occurred in 2015 is small.

Table 9: Filled incremental triangle by Poisson model

Accident year	Time until the first payment										
	0	1	2	3	4	5	6	7	8	9	10
2005	1,021	1,918	254	63	8	1	0	0	0	0	0
2006	1,080	1,736	191	51	3	2	1	0	1	0	**0**
2007	1,304	1,430	268	43	9	1	0	0	0	**0**	**0**
2008	1,411	1,590	191	36	9	2	1	0	**0**	**0**	**0**
2009	1,181	1,225	101	39	13	1	0	**0**	**0**	**0**	**0**
2010	1,274	995	156	37	10	0	**0**	**0**	**0**	**0**	**0**
2011	1,059	962	98	35	4	**1**	**0**	**0**	**0**	**0**	**0**
2012	989	844	105	25	**5**	**1**	**0**	**0**	**0**	**0**	**0**
2013	1,089	730	125	**32**	**4**	**0**	**0**	**0**	**0**	**0**	**0**
2014	1,307	547	**114**	**16**	**2**	**0**	**0**	**0**	**0**	**0**	**0**
2015	1,261	**563**	**126**	**19**	**2**	**0**	**0**	**0**	**0**	**0**	**0**

Source: the author's work

7. Conclusion

Chain ladder method is based on assumption that development of the cumulative claims for each occurrence year is independent. Development factors are estimated and used to fill the cumulative triangle. Sometimes the assumption is violated. Modified Chain ladder can then be used instead. Development factors which cannot be estimated by Chain ladder are then estimated by regression model (individual development factors are dependent variable and accident year is the independent variable). Poisson model approach is based on the assumption that in each accident year i, number of discrete years until the claim has the first payment has Poisson distribution with parameter λ_i. Table 10 contains results from all three above-mentioned methods.

Table 10: Results of all methods

Accident	Chain ladder		Modified Chain ladder		Poisson model	
year	Ultimate	IBNR	Ultimate	IBNR	Ultimate	IBNR
2005	3,265	0	3,265	0	3,265	0
2006	3,065	0	3,065	0	3,065	0
2007	3,055	0	3,055	0	3,055	0
2008	3,240	0	3,240	0	3,240	0
2009	2,560	0	2,560	0	2,560	0
2010	2,473	1	2,473	1	2,472	0
2011	2,159	1	2,159	1	2,159	1
2012	1,970	7	1,970	7	1,969	6
2013	1,981	37	1,981	37	1,981	37
2014	2,018	164	1,984	130	1,986	132
2015	2,776	1,515	1,768	507	1,971	710

Source: the author's work

Figure 6: Ultimate claims for all above-mentioned method

Source: the author's work

The assumption of Chain ladder was violated, so the results are biased. Its estimate of IBNR and Ultimate probably overestimates reality which will lead to overbooking an insurance company. Modified Chain ladder uses same development factors for development years 3 to 10, so predictions are the same for accident years 2006-2013. The difference is only in the last two accident years where modified Chain ladder uses the first development factor 1.31 instead of 2.02 and the second development factor 1.05 instead of 1.07. It leads to estimate 1,008 IBNR claims less than Chain ladder. This method assumes that the number of claims will decrease in time. Poisson model estimates fewer claims for the accident year 2014 than modified Chain ladder, the difference is only 2 claims. The estimate for the accident year 2015 is higher than modified Chain ladder (about 203 claims) and lower than Chain ladder (about 805 claims). Graphical representation of results is shown in Figure 6 and 7.

Figure 7: Predicted IBNR for all above-mentioned method

Source: the author’s work

Poisson model approach estimates Ultimate claims for the last fourth accident years on the same level for this portfolio. For this portfolio, it is more conservative approach than modified Chain ladder which assumes that the decreasing trend in a total number of claims will continue. It seems that Poisson model gives more accurate results for this portfolio. An advantage of the modified Chain ladder and the Poisson model is that these methods take into account the decreasing trend in a number of claims.

Acknowledgements

The support of the grant scheme USE OF MICROMODELS FOR MODELING THE AMOUNT AND RISK OF NON-LIFE INSURANCE RESERVES AT A ONE-YEAR HORIZON IG410036 is gladly acknowledged.

References

[1] Bartoszewicz, B. 2005, Modelling the Claim Count with Poisson Regression and Negative Binomial Regression. In: Baier, D., Wernecke, KD. (eds) Innovations in Classification, Data Science and Information Systems. Berlin: Springer. ISBN 978-3-540-23221-6.

[2] Boland, P. 2007. Statistical and probabilistic methods in actuarial science. New York: Chapman & Hall/CRC. ISBN 1-58488-695-1.

[3] David, M., Jemna D. 2015. Modeling the Frequency of Auto Insurance Claims by Means of Poisson and Negative Binomial Models. Annals of the Alexandru Ioan Cuza University – Economics, vol. 62, pp. 151-168.

[4] Garrido, J., Genest, C., Schulz, J. 2016. Generalized linear models for dependent frequency and severity of insurance claims. Insurance: Mathematics and Economics, vol. 70, pp. 205-215.

[5] Gisler, A., Wüthrich, M. 2008. Credibility for the chain ladder reserving method. Astin Bulletin, vol. 38, pp. 565-600.

[6] Mack, T. 1993. Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates, AST1N Bulletin, vol. 23, pp. 213-225.