INSURANCE RESERVES ESTIMATION BY BOOTSTRAP

Bohdan Linda, Jana Kubanová, Pavla Jindrová

Abstract: The claim reserving calculation is one of the basic problems of the successful function of the insurance companies. These reserves can be calculated both classical and simulation way. The second way – use of resampling method is presented in the paper. Application of bootstrap methods in connection with problems solved in insurance theory is described. The parameters of interest are estimated, the ways how to calculate the point and interval estimates are shown.

Keywords: Point Estimate, Interval Estimate, Chain Ladder Method, Development Factor, Resampling, Bootstrap Method.

JEL Classification: C15, G17.

Introduction

One of conditions under that the insurance company can be competitive and successful in the insurance market is to dispose of reasonable financial reserves. These reserves are used for insurance benefits payments. Very important question is the amount of money that is reserved for these payments. It must be neither too much nor too little; the first extreme leads to situation that finance are conserved and they can't be exchanged, the second extreme can cause insolvency.

One of the most common methods, how to determine the insurance reserves, is the Chain ladder method, described in [5] or [6].

1 Statement of a problem

1.1 Basis of claim reserving theory

When any accident occurs, it is conventional, that payment of the insurance benefits is not fully realized during the accident year, but certain part of the insurance benefit can be paid during following years – it means that raises any interval between the moment of accident and the moment of total insurance benefit payment. This insurance benefit can be paid off in parts during same years. The car insurance, the accident insurance, the property insurance, the travel insurance are examples of above mentioned kinds of insurances.

Estimate and creation of the optimal claim reserves is the necessary condition for following covering of these claims. In the insurance theory - two basic types of claim reserves for past exposures are distinguished [6]:

- IBNR (Incurred but not reported). The reserve for the insurance benefit for the claims that have occurred but haven't been reported yet corresponds with it.
- RBNS (Reported but not settled). This means not settled insured accident corresponding with the reserve for the insurance benefit from the claims that

have been reported but have not been settled. The payment is expected in the future.

The principle of claim reserving calculation assumes knowledge of the past payments.

The most often applied claim reserving calculation methods are:

- Chain-ladder method.
- Bornhuetter-Ferguson method.
- Poison model for claim counts.

More models based on different mathematical principles were developed for every of above mentioned method. Every of these models can be solved by various processes, starting with exact over approximative and terminating with the simulation methods.

When we want to know accuracy of the estimates, we need to introduce a stochastic model and we can select one of following approaches: bootstrap method, Monte Carlo simulations, Bayesian approach. This paper is devoted to the less known but simple and effective method – bootstrap.

1.2 Description of Chain ladder method

The "Run Off Triangle" is the basic scheme for the Chain ladder method (CL) application. Originally, the scheme (resp. the table) is created by the incremental insurance benefits $X_{i,j}$, that were paid off in the year j for the accidents that happened in the year i. But the CL method doesn't use these incremental data $X_{i,j}$, but the cumulative insurance benefits $C_{i,j}$, which express the total insurance benefits paid for the accident that happened in the year i from the development year 0 till j. The structure of the data is presented in the table 1. The Chain ladder method is based upon the idea that the variable $C_{i,j+1}$ is the function of the variable $C_{i,j}$, which can be described as $C_{i,j+} = {}^{r}(C_{i,j})$.

The application of the Chain ladder method is conditioned by the following assumptions [5]:

- The cumulative claims $C_{i,j}$ of the different accident years i are independent.
- Both the last accident year and the last development year are given by n.
- Development of the sums of paid insurance benefits is characterized by the "development coefficient of the insurance benefit" λ_i .
- Homogenous portfolio.
- Stability of inflation during development years.

To be able to refund the claims, occurred in the year i, we have to estimate the values $C_{i,j}$ in the grey part of the table 1. Instead of using the common simple deterministic model $C_{i,j+1} = \lambda_j \cdot C_{i,j}$, the other model expressing relation between variables $C_{i,j}$ and $C_{i,j+1}$ and defining the autoregressive process was applied:

$$C_{i,j+1} = \lambda_j \cdot C_{i,j} + \sigma_{,j} \sqrt{C_{i,j}} \, \varepsilon_{,j+} \,. \tag{1}$$

Coefficients λ_i , σ_i , $\varepsilon_{i,j+1}$ can be estimated step by step by the variables

$$\hat{\lambda} = \sum_{\substack{i=0 \\ n-j-1 \\ \sum j=0}}^{n-j-1} C_{i,j} ; \quad \sigma_{j} = \sqrt{\frac{1}{n-j-1}} \sum_{i=1}^{n-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{\lambda}_{j} \right) ;$$
(2a, 2b)

$$\varepsilon_{.,j+} = \frac{\frac{C_{i,j+}}{C_{i,j}} - 1}{\sigma} \sqrt{C_{i,j}} \qquad j = 0, 1, ..., n-1, \quad i = 0, 1, ..., n;$$
(3)

Tab. 1: Cumulative insurance benefits

accident	develop	development year j							
year i	0	1	2	•••	n-2	<i>n</i> -1	n		
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	•••	$C_{0,n-2}$	$C_{0,n-1}$	$C_{0,n}$		
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	•••	$C_{1,n-2}$	$C_{1,n-1}$			
2	$C_{2,0}$	$C_{2,1}$	$C_{2,2}$	•••	$C_{2,n-2}$				
•••	•••	•••	•••	•••					
n - 1	$C_{n-1,0}$	$C_{n-1,1}$							
n	$C_{n,0}$								

Source of data: [6]

1.3 Basic idea of bootstrap method

The bootstrap method was elaborated by Bradley Efron in 1977 [1]. The principle is to produce new samples from the original data set with the same size. The original data analysis procedure is repeated many times and the bootstrap replications of the parameter of interest are obtained. The principal advantage of the method is construction of the artificial data sets without making any assumptions about probability distribution [1].

When $C = (C_1, C_2, ..., C_n)$ is the random sample from an unknown distribution F, the bootstrap samples can be produced in two ways:

- When an unknown distribution belongs to any family of distributions differing in values of a parameter ψ (it is marked as F_{ψ}), the parameter ψ is estimated by any suitable statistics $\hat{\psi} = \imath(C)$. The bootstrap samples $C^* = \mathcal{I}_1^*, ..., C_n^*$ from the distribution $F_{\hat{\psi}}$ are generated. This process is called the parametric bootstrap.
- When we don't have any other information about the distribution F, the empirical distribution function \hat{F} is calculated from the random sample C. The bootstrap samples C^* are then generated from this distribution . This process is called the non-parametric bootstrap.

Let $\lambda = t(F)$ be any parameter of unknown distribution F and $\hat{\lambda} = g(C, F)$ is its estimate. The distribution function of this estimate is marked G(C, F). The statistics $\lambda = \xi(C^*, \hat{F})$ is called the bootstrap replication of the estimate $\hat{\lambda}$ and its distribution function $G^*(C^*, F)$.

1.4 The bootstrap quantile interval

When $G^*(X^*, \hat{F})$ is the distribution function of the bootstrap replications Θ , the $(1-2\alpha)$ percentage confidence interval can be expressed

$$\langle G^{*-}(\alpha); G^{*-}(1-\alpha) \rangle$$
 resp. $\langle \Theta_{-}; \Theta_{-+} \rangle$,

where Θ_{-} is $(100 \cdot \alpha)$ -th percentile [4]. This formula is exactly true only when the random sample originates from the distribution F.

We can keep at disposal only finite number of replications $\hat{\Theta}^*$ in reality and to obtain only estimates of the interval $\langle \Theta : \Theta_{-\kappa} \rangle$.

R bootstrap samples $X_1^*,...,X_n^*$ are generated and corresponding values of bootstrap replications Θ ,..., Θ are calculated of them. These replications are arranged in order according to size. $(100 \cdot \alpha)$ -th percentile Θ is estimated by the help of the value $\Theta_{R,\alpha}$, eg. $(R \cdot \alpha)$ -th value in the sequence of all bootstrap replications Θ ,..., Θ arranged in order according to size. The interval

$$\langle \Theta_{R,\alpha} | \Theta_{R,1-\epsilon} \rangle$$
 (4)

is then the estimate of the $(1-2\alpha)\cdot 100\%$ confidence interval $\langle \Theta_{-\kappa} \rangle$.

2 Problem solving

2.1 Bootstrap approach to claim reserves estimate

Situation according to the chapter 2.3. paragraph b) is assumed, it means that the distribution of the random variables $C_{i,j}$ is unknown and the non-parametric bootstrap is used. Resampling is done and bootstrap replications of the estimate of the parameter λ_j are calculated. Resampling is performed in a different way - not the original data are bootstrapped, but residuals $\varepsilon_{i,j}$ mentioned in the model (1) are used and estimated according to the formula (3) [11].

The literature recommends to use adjusted residuals calculated according to the following formula:

$$z_{i,j+} = \varepsilon_{i,j+} \left(-C_{i,j} \left(\sum_{i=1}^{n-j-} C_{i,j} \right)^{-1} \right)^{\frac{1}{2}}.$$
 (5)

The process that follows is to get the bootstrap replications λ_j^* . By the help of adjusted bootstrap residuals $z_{i,j+}^*$ the bootstrap sample of $C_{i,j}^*$ is gained. The first step is to put $C_{i,0}^* = \mathbb{Z}_{i,0}$. To calculate the other values $C_{i,j}^*$ (for i+i< i) we use the following formula (6).

$$C_{i,j+}^* = \lambda C_{i,j}^* + \sigma \sqrt{C_{i,j}^*} z_{i,j+}^*$$
(6)

The formula (7) shows how to calculate the bootstrap replications λ_j of the development coefficient $\hat{\lambda}$:

$$\lambda_{j}^{*} = \sum_{i=1}^{n-j-} C_{i,j+}^{*} \sum_{i=1}^{n-j} C_{i,j}^{*} , j = 0, 1, ..., n-1$$
 (7)

The final step of the method is to estimate the missing values $C_{i,j}^*$ of the predicted cumulative claims (lower triangle matrix). The following formula (8) helps us to finish calculations.

$$C_{i,j}^* = C_{i,n-i} \prod_{k=n-i}^{j-1} \lambda_{j}, \qquad j = 1, 2, ..., n$$
 (8)

The bootstrap process is repeated R-times and for each pair of indices i,j (i+1>1). The vector of λ_j , j=0,1,...,n-1, and triangle matrix of the predicted cumulative claims $C_{i,j}^*$ are gained after each repetition. The bootstrap values $C_{i,j}^*$ enables us to estimate the unknown value of cumulative claims $C_{i,j}$ - it is the average from the appropriate $C_{i,j}^*$ bootstrap replications.

2.2 Confidence interval

2.2.1 The quantile confidence interval for λ

The result of R bootstrap replications are R calculated values of λ_j , j=0,1,...,n-1. According to the formula (4) the limits of the bootstrap confidence interval are the $R \cdot \alpha$ -th ordered value of R replications of the variable λ_j and similarly the $R \cdot (1-\alpha)$ -th ordered value of R replications of the variable λ_j .

Approximation of the 1 - 2α quantile interval is then $\langle \lambda_{R,\alpha}; \lambda_{R,1-} \rangle$.

2.2.2 The quantile confidence interval for cumulative claim reserves

The calculated values of $C_{i,j}^*$ are considered. After R bootstrap simulation the lower and upper limits of the bootstrap confidence interval for the cumulative claim reserves $C_{i,j}$ are $C_{i,j}^*$ and $C_{i,j-R,1-}^*$, where $C_{i,j-R,\alpha}^*$ is the $R \cdot \alpha$ -th ordered value and $C_{i,j-R,1-}^*$ is the $R \cdot (1-\alpha)$ -th ordered value of R replications of the variable $C_{i,j}^*$.

Approximation of the 1 - 2α quantile interval is then

$$\langle C_{i,j,R,\alpha}^*; C_{i,j,R,l-}^* \rangle.$$
 (9)

2.3 Presentation of the bootstrap method – the concrete example

To present the bootstrap process of the claim reserves estimation the data published in Pacáková (2004) were used. Both the incremental data and cumulative data of insurance benefits are stated in the book. For our need only cumulative data are presented in the table 2 – white upper triangle.

2.3.1 Classical and bootstrap Chain ladder method

The missing cumulative claim reserves were estimated by the help of deterministic CL method in the left half of the table and by the help of bootstrap CL method in the

right half of the table 2. The results obtained by the deterministic CL method were calculated according to the simple deterministic model, $C_{i,j+1} = \lambda_j \cdot C_{i,j}$ in which the coefficients λ_j were estimated according to the formula (2a). They are written in the "left grey" part of the table 2.

Tab. 2: Cumulative values of claim reserves calculated by deterministic CL method and bootstrap CL method

		Detern	Deterministic Chain ladder method						Boot	strap Ch	ain lada	ler meth	od	
			development year j						C	levelopr	nent yea	ır j		
	i	0	1	2	3	4	5	i	0	1	2	3	4	5
	0	566	1 049	1 270	1 407	1 460	1 483	0	566	1 049	1 270	1 407	1 460	1 483
ır i	1	501	993	1 186	1 345	1 409	1 431	1	501	993	1 186	1 345	1 409	1 431
t year	2	543	1 055	1 287	1 471	1 534	1 558	2	543	1 055	1 287	1 471	1534	1558
accident	3	652	1 323	1 633	1 842	1 921	1 951	3	652	1 323	1 633	1841	1920	1950
acc	4	739	1 479	1 799	2 030	2 116	2 149	4	739	1 479	1799	2028	2115	2148
	5	752	1 478	1 798	2 028	2 115	2 148	5	752	1478	1797	2026	2113	2146
		λ_0	λ_I	λ_2	λ_3	λ_4								
		1.966	1.216	1.128	1.043	1.016								

Source of data: [6], author

The bootstrap values were calculated after 1000 bootstrap simulations. The bootstrap algorithm for $C_{i,j}$ values calculation was programmed in the software Matlab. The claim reserves were estimated according to the formula (1) and calculated results are in the "right grey" part of the table 2.

Tab. 3: Difference between incremental values of classical and bootstrap method

<i>,,</i>	J 1								
	development year j								
	i	0	1	2	3	4	5		
į	0	0	0	0	0	0	0		
ar	1	0	0	0	0	0	0		
yea	2	0	0	0	0	0	-0.303		
ent	3	0	0	0	0	0.744	1.003		
accident	4	0	0	0	1.581	0.868	1.200		
ac	5	0	0	0.907	2.470	1.709	2.023		

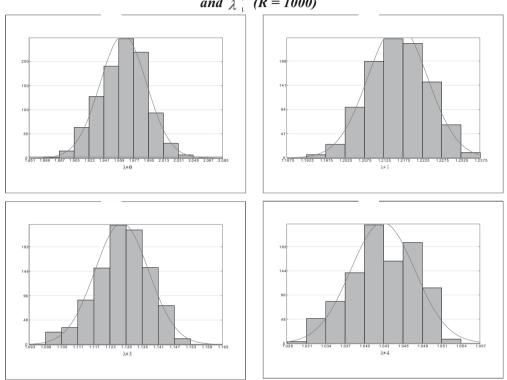
Source of data: authors

Claim reserves presented in the table 2 are in cumulative form. They were transformed to the incremental values and differences between results of two above mentioned methods calculated. These differences are presented in the table 3. It can be stated that almost all differences (except one) are positive, it signifies the reserves estimate by the bootstrap method is lower than the reserves estimate by the classical Chain ladder method. The insurance company should save some money or make some other profitable investments when bootstrap estimate is used.

2.3.2 Probability distribution estimate for λi and Ci,j and convergence

Histogram of 1000 bootstrap replication of development factors λ_1 , λ_2 , and λ_3 , were used to estimate the probability distribution of these parameters. We can assume that approximation of this distribution by the normal (Gaussian) probability distribution is adequate. These histograms are illustrated in the figure 1. Probability distribution of the development factor λ_4 is approximated by the uniform distribution.

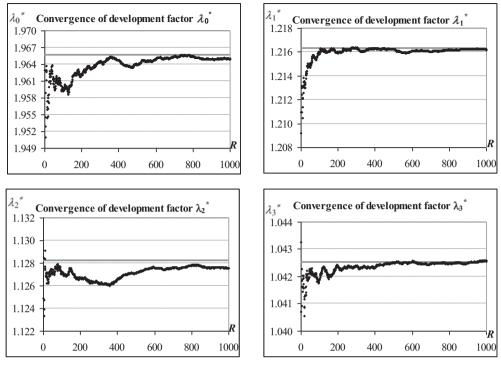
Fig. 1: Histograms of bootstrap replications of the development factor λ_{j} , λ_{j} , λ_{j} , and λ_{j} (R = 1000)



Source of data: authors

The question that is asked very often is how many bootstrap simulations are optimal to be done? The answer is not unique, the simulation have to continue until the values start to stabilize. Convergence of the development factors λ , (i = 0, 1, 2, 3) when 1000 simulations were made and 1000 replication calculated is presented in the figure 2. The values started to be unchanging after 500 bootstrap replications.

Fig. 2: Convergence for development factors λ_0^* , λ_1^* , λ_2^* , λ_3^*



Source of data: authors

We can now compare the difference between deterministic CL and bootstrap CL methods when the development factor λ is estimated.

Tab. 4: Difference between values of development factor

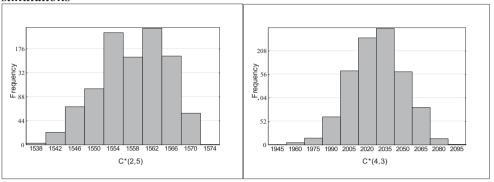
	λ_0	λ_1	λ_2	λ_3	λ_4
deterministic CL(λ)	1.965678	1.21629	1.128239	1.042515	1.01575
bootstrap CL ()	1.96499	1.21620	1.12754	1.04257	1.01580

Source of data: authors

The second row of the table 4 presents values of the development factor λ_i (i = 0, 1, 2, 3, 4) calculated according to the simple deterministic model, $C_{i,j+1} = \lambda_j \cdot C_{i,j}$ and formula 2a). In the third row we can see values of the development factor—obtained after 1000 bootstrap simulations. We are interested in difference between values of development factor, when two above stated methods were used. We can sum up the results and claim, that the difference after application both methods is very small, at the fourth or fifth decimal place.

The probability distribution of the cumulative claim reserves $C_{i,j}^*$ can be displayed by histogram. We selected histogram only of two variables to demonstrate the problem, bootstrap claim reserves $C_{2,5}^*$ and $C_{4,3}^*$, but histogram of the other variables were generally similar. We can claim that it is possible to approximate the distribution of the cumulative claim reserves by the normal (Gaussian) distribution.

Fig. 3: Histograms of bootstrap replications of the $C^*_{2,4}$ $C^*_{4,3}$ after 1000 bootstrap simulations



Source of data: authors

2.3.3 Confidence interval for the development factor and claim reserves

One of the tasks of this paper was to find the 90% confidence interval for the development factor λ_j . The lower and upper limits of this interval are stated in the table 5 (the third row and the fourth row). In the second row (bold) we can see the point estimates of the developing factor for periods j and j+1.

Bootstrap confidence intervals are not symmetrical, the left part of the interval is longer for λ_0 , λ_2 , λ_3 estimates.

Tab. 5: 90% confidence interval for λ

R=1000	λ_0	λ_1	λ_2	λ 3	λ 4
	1.9657	1.2163	1.12824	1.04254	1.0158
lower limit $\lambda_{R,\alpha}$	1.9176	1.2037	1.1102	1.0338	1.0158
upper limit $\lambda_{R,1-}$	2.0098	1.229	1.1434	1.0498	1.0158

Source of data: authors

The 90% confidence interval for claim reserves is shown in the table 6. We can observe the original triangle of incremental values (upper triangle) and the lower triangle with three values. The middle one represents the predicted claims, calculated after 1000 bootstrap simulations, lower and upper values represent lower and upper limit of the 90% bootstrap confidence interval, obtained after 1000 bootstrap replications as well.

The table 7 presents how many percentage of the value $X_{i,j}^*$ creates the length of the right part of the 90% confidence interval. It means that the real value of $X_{i,j}$ may vary about stated percentage from the estimate $X_{i,j}^*$. We can see that these values are indispensable. This error is of course the maximal error and we fall into this error with a small probability. We can estimate the dependence between the size of error and reliability coefficient by the help of bootstrap.

Tab. 6: 90% confidence interval for claim reserves – incremental values

	0	1	2	3	4	5
0	566	483	221	137	53	23
1	501	492	193	159	64	22.200 22.200 22.200
2	543	512	232	184	49.700 62.616 73.200	10.984 24.161 34.884
3	652	671	310	180.000 208.276 234.100	47.124 78.375 108.724	-1.551 30.240 61.049
4	739	740	301.300 319.762 338.700	192.338 229.416 264.838	45.722 86.326 125.922	-7.904 33.312 73.596
5	752	690.100 725.673 759.400	272.027 319.478 367.327	169.849 229.208 291.949	25.641 86.249 151.541	-28.208 33.280 99.592

Source of data: authors

Tab. 7: Percentage of forecasted value

Tub. 7. I ereemage of forecasica value								
·	0	1	2	3	4	5		
0								
1						0		
2					17	44		
3				12	39	102		
4			6	15	46	121		
5		5	15	27	76	199		

Source of data: authors

2.3.4 Claim reserving estimates

The insurance companies are questing for the most reliable method how to determinate the claim reserves for the following periods. We can see in the table 8 estimates that were calculated in different ways. The claim reserves calculated by the deterministic CL method are stated in the second column of the table. These reserves for the period 6 – 10 result from the data referring from insurance benefits paid during the periods 0–5. Results obtained by the alternative method - the bootstrap Chain ladder method are introduced in the third column. The lower and upper limit of the 90% bootstrap confidence interval for the claim reserves are presented in the fourth and fifth columns. The data in the last column include the information about the difference in the claim reserves estimates between classical and bootstrap Chain ladder method application. We can see that the reserves calculated by bootstrap Chain ladder method are lower than the reserves calculated by the classical Chain ladder method.

This result can evoke the idea that the insurance company needn't to create so big claim reserve and that it can use some amount of money in a more effective way.

Tab. 8: Claim reserving estimates - comparison

Two or commercial to be compared to								
period	CL	Cl-boots	low.limit	up.limit	dif.Cl-Clboot			
6	1340.233	1338.526	1243.300	1427.600	1.706			
7	652.894	651.430	522.473	775.773	1.464			
8	347.107	345.775	214.020	478.920	1.333			
9	119.572	119.561	17.736	225.136	0.011			
10	33.314	33.280	-28.208	99.592	0.034			

Source of data: authors

Conclusion

The goal of the paper was to present some alternative approaches to the solution of the problem of claim reserves estimate and to show applicability of the bootstrap method. The important advantage of the bootstrap method at claim reserves estimates were pointed out – the absence of strict assumptions. Application of the bootstrap method enables us to make not only point estimates of the parameters of interest, but interval estimates as well. The confidence interval for the development coefficients and for the cumulative reserves $C_{i,j}$, resp. incremental reserves $X_{i,j}$, were calculated. Some advantage of interval estimates was accented in the paper, above all in continuity with economy of the insurance companies and determination of the amount of costs for claim reserves.

Acknowledgement

The article was elaborated with the support of the grant GAČR 402/09/1866, Modeling, simulations and management of insurance risk.

References

- [1] EFRON, B, TIBSHIRANY, R. J. 1993. An Introduction to the Bootstrap. Chapman & Hall/CRC, 1993. ISBN 0-412-04231-2.
- [2] HINKLEY, D. V, Davison, A.C.1997. Bootstrap methods and their application. Cambridge University Press1997. ISBN 0-521-57391-2.
- [3] JINDROVÁ, P. 2011. Trends in the Insurance Business in the Czech Republic. In: Scientific papers of the University of Pardubice, No. 19 (I/2011), Vol. XVL, series D. Pardubice 2011, ISSN 1211-555X, p.70-81
- [4] KUBANOVÁ, J. 2003. Bootstrapové metody. Monografie. Univerzita Pardubice, 2003. ISBN 80-7194-581-1
- [5] MACK,T. 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. Astin Bullerin 23/2, p.213-225
- [6] PACÁKOVÁ, V. 2004. Poistná štatistika. Iura Edition. Bratislava 2004. ISBN 80-8078-00-8.
- [7] PACÁKOVÁ, V., ŠOLTÉS, E, ŠOLTÉSOVÁ, T. 2009. Kredibilný odhad škodovej frekvencie. E+M, roč.XII, 2/2009. str. 122-126, ISSN 1212-3609

- [8] RUBLÍKOVÁ, E. Analýza časových radov. 2007. Iura Edition. Bratislava 2007. ISBN 978-80-8078-139-2.
- [9] PINHEIRO, J. R., ANDRADE E SILVA, J. M., CENTENO, M., L.2000. Bootstrap Methodology in Claim Reserving. Working papers. Centro de Mathemática Aplicada á Previsao e Decisao Económica. Lisboa, 2000.
- [10] VERALL, R. J. Statistical Methods for the Chain Ladder Technique. www.casact.org/pubs/forum/94spforum/94spf393.pdf
- [11] WÜTHRICH, M.V., MERZ, M. 2008. Stochastic Claims Reserving Methods in Insurance. John Wiley 2008. ISBN 978-0-470-72346-3

Contacts Address

doc. RNDr. Bohdan Linda, CSc.

doc. PaedDr. Jana Kubanová, CSc.

Mgr. Pavla Jindrová, Ph.D.

University of Pardubice, Faculty of Economics and Administration, Department of Mathematics

Studentská 95, 53210 Pardubice, Czech Republic

Email: bohdan.linda@upce.cz; Email: jana.kubanova@upce.cz Email: pavla.jindrova@upce.cz Phone number: +420 46 603 6020 Phone number: +420 46 603 6776

Received: 28. 06. 2011 Reviewed: 08. 11. 2011 Approved for publication: 30. 11. 2011