A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 235, 2010

Anna Szymańska[*]

APPLICATION OF BAYESIAN ESTIMATORS TO THE ESTIMATION RATES OF NET PREMIUM IN CR AUTOMOBILE LIABILITY INSURANCE

Abstract. In the paper an application of the Bayesian estimators to the posterior tariffication in CR automobile liability insurance is presented. Net premiums are determined by means of the expected value method assuming that the damage size follows the Pareto distribution. The rates of net premium results for the Pareto distribution with different parameters are compared.
Key words: Bayesian estimators, claim size, posterior tariffication, bonus-malus.

I. INTRODUCTION

In CR car liability insurance the premiums are determined in two stages. The first one consists in calculating basic premium based on priori factors (i.e. observable risk factors e.g. car make and year of production, engine capacity, driver's sex and age). The second one is the posterior tariffication (driver's damage history). Thus, in the first stage insurer determines the basic premium and in the second stage insurer estimates the fraction of the Basic Premium which should be paid by the driver (Lemaire 1995).

There are many possible *posterior* tariffications. Insurance companies usually apply the bonus-malus systems based on the number of damages caused by the insured in the past. The bonus for one year of no-damage driving is a 10% reduction of the basic premium. The penalty for at least one damage is a 10% rise.

The aim of this paper is to present *posterior* tariffication methods based not on the number of damages but on the size of them. The premiums are estimated by means of Bayesian estimators. The premiums for different damage size Pareto distributions are compared.

At the begining of the 20-th century the net premium was calculated as a weighted mean of collective premium μ and individual premium $\bar{x}_i$ estimated on the basis of claim history in the past, i.e.

[*] Ph.D., Chair of Statistical Methods, University of Łódź.

$$m(\theta_i) = Z_i \bar{x}_i + (1 - Z_i)\mu \qquad (1)$$

where $Z_i \in [0,1]$. The premium $m(\theta_i)$ so defined is called *credibility premium* for the *i*-th contract while Z_i is a *credibility coefficient*. The estimator of variable $X_{i,t+1}$ is called its' variable's *predictor,* and its value is called a prognosis for $X_{i,t+1}$ based on observations $x_{i1},...,x_{it}$ (Jasiulewicz 2005). If the credibility coefficient Z is close to zero that means that the collective premium should be trusted more than the individual mean based on observations. The credibility coefficient Z close to one means that the insurer can trust the individual mean based on observations. The credibility coefficient Z is close to one when the insurance history of a given policy is long and little diversified in time or the contracts are very diversified. The basis of the credibility theory is the Bayesian statistical analysis with quadratic loss function.

Let us assume that variable X_j denotes the claim amount of one insured person in the j-th insurance year and its density function is $f(x_j)$. The distribution of variable X_j depends on parameter θ. The risk parameter θ of the insured is constant throughout the insurance period and is the value of random variable Θ which follows the density distribution $\pi(\theta)$.

Let $f(x_j|\theta)$ be the density function of conditional distribution of random variable X_j on condition $\Theta = \theta$. The individual net premium of the insured with risk parameter $\Theta = \theta$ is the expected value of the conditional distribution $f(x_j|\theta)$ and is equal to $m(\theta) = E(X_j|\theta)$.

In order to estimate the future claim $(X_{t+1}|\theta)$ of the insured with unknown parameter θ we minimize the expected value of the quadratic loss function $(X_{t+1} - d(X_1,...,X_t))^2$, where $d(x_1,...,x_t)$ denotes an arbitrary real function defined on the observation space (Domański, Pruska 2000).

The forecast of the future claim X_{t+1} with minimum mean square error for given $X_1 = x_1,...,X_t = x_t$ is the conditional expected value $E(X_{t+1}|X_1,...,X_t)$.

II. BAYESIAN PREDICTORS OF CLAIM SIZE

Let X_j be random variable denoting claim size in year j for a given policy, and $(X_1, X_2,\ldots,X_t)$ be the vector of observations of claim sizes through t years for a given policy. Let the distribution of random variable X_j be dependent on parameter θ. We assume that the risk parameter θ of insured is constant throughout the whole insurance period and is the outcome of a random Θ, random variables X_j for established $\Theta = \theta$ are independent and have equal expected values and the insured's generate losses independent.

We assume that the insurer knows the value of the insured's claims from the period of t years. Let X_{t+1} be the unknown size of claims in t+1 year for a policy described by the vector of observations (X_1, X_2,…, X_t) which can be estimated by means of a Bayesian estimator.

Let random variable X have the exponential distribution with parametr θ with the density function

$$f(x|\theta) = \theta e^{-\theta x}\ , x > 0, \theta > 0 \tag{2}$$

If parameter Θ is a random variable following the gamma distribution with parameters α and β, with the density function

$$\pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta\theta}\theta^{\alpha-1}, \quad \alpha > 0, \beta > 0, \theta > 0, \tag{3}$$

the random variable X is distributed with density function

$$f(x) = \int_0^{\infty} f(x|\theta)\pi(\theta)d\theta = \int_0^{\infty} \theta e^{-\theta x} \frac{\beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta\theta}\theta^{\alpha-1} d\theta =$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} \theta^{\alpha} e^{-(\beta + x)\theta} d\theta = \frac{\alpha\beta^{\alpha}}{(\beta + x)^{\alpha+1}}$$

$$f(x) = \frac{\alpha\beta^{\alpha}}{(\beta + x)^{\alpha+1}} \quad , x > 0, \tag{4}$$

It is the density function of random variable of the Pareto distribution with parameters α and β , respectively with the expected value and variance [Hossack, Pollard , Zehnwirth , 1999]

$$EX = \frac{\beta}{\alpha - 1}, \quad \alpha > 1 \text{ i } D^2X = \frac{\alpha\beta^2}{(\alpha - 1)^2(\alpha - 2)}, \quad \alpha > 2. \tag{5}$$

The posterior density $\pi_x^*(\theta)$ calculated by means of the Bayes formula is given by

$$\pi_x^*(\theta) = \frac{\pi(\theta)\prod_j f_\theta(x_j)}{f(x_1,...x_t)} = \frac{\left(\tilde{\beta} + \sum_j x_j\right)}{\Gamma\left(\tilde{\alpha} + t\right)} e^{-\theta\left(\tilde{\beta} + \sum_j x_j\right)} \theta^{\tilde{\alpha}+t-1} \tag{6}$$

The posteriori distribution of parameter Θ is the gamma distribution with parameters $\hat{\alpha}$ i $\hat{\beta}$ given by

$$\hat{\alpha} = \alpha + t \quad \text{i} \quad \hat{\beta} = \beta + \sum x_j \tag{7}$$

The predictive distribution of claim size X_{t+1} on condition $X_1 = x_1,...,X_t = x_t$ has the form

$$f\left(x \middle| x_1,...,x_t\right) = \int f\left(x \middle| \theta\right) \pi\left(\theta \middle| x_1,...,x_t\right) d\theta =$$

$$= \int \theta e^{-\theta x} \frac{\hat{\beta}^{\hat{\alpha}}}{\Gamma\left(\hat{\alpha}\right)} e^{-\hat{\beta}\theta} \theta^{\hat{\alpha}-1} d\theta = \int \frac{\hat{\beta}^{\hat{\alpha}}}{\Gamma\left(\hat{\alpha}\right)} e^{-\left(\hat{\beta}+x\right)\theta} \theta^{\left(\hat{\alpha}+1\right)-1} d\theta = \tag{8}$$

$$= \frac{\hat{\beta}^{\hat{\alpha}}}{\left(\hat{\beta}+x\right)^{\hat{\alpha}+1}} \frac{\Gamma\left(\hat{\alpha}+1\right)}{\Gamma\left(\hat{\alpha}\right)} \cdot 1 = \frac{\hat{\alpha}\hat{\beta}^{\hat{\alpha}}}{\left(\hat{\beta}+x\right)^{\hat{\alpha}+1}}$$

This is therefore the Pareto distribution with parameters $\hat{\alpha}$ and $\hat{\beta}$. The expected value of this distribution is given by the formula

$$E\left(X_{t+1} \middle| X_1,...,X_t\right) = \frac{\hat{\beta}}{\hat{\alpha}-1} = \frac{\beta + \sum x_j}{\alpha + t - 1} \tag{9}$$

For such distributions $f\left(x \,\middle|\theta\right)$ and $\pi(\theta)$ we have: $\mu = \dfrac{\beta}{\alpha - 1}$, $Z = \dfrac{t}{t + \alpha - 1}$. Hence, the individual premium in year t+1 is the Bayesian credibility predictor of the form

$$m(\theta) = \frac{\beta + \sum x_j}{\alpha + t - 1}. \tag{10}$$

III. NET PREMIUM ESTIMATION

In CR automobile liability insurance the individual net premium in period t+1 amounts to

$$m(\theta) = (EX)\cdot(E\Lambda)\cdot b_{t+1}(x_1,...,x_t) \qquad (11)$$

where $m(\theta)$ - individual net premium in period t+1, (EX) - the expected value of a single claim size , $(E\Lambda)$ - the expected value of the number of claims in accounting period (in one year stretch), $b_{t+1}(x_1,...,x_t)$ - the posterior premium (Lemaire 1995).

Assuming that $EX = \frac{\beta}{\alpha - 1}$ and $E\Lambda = 1$ the equation (11) has the form

$$m(\theta) = \frac{\beta}{\alpha - 1}\cdot b_{t+1}(x_1,...,x_t) \qquad (12)$$

Hence, a driver who after t years reported claims of the size of $\sum_{j=1}^{t} x_j$, should pay the posterior premium equal to

$$b_{t+1}(x_1,...,x_t) = \frac{\alpha - 1}{\beta} m(\theta)\cdot 100\% \qquad (13)$$

The simplest rule for calculating premium in CR automobile liability insurance is the rule of the expected value (Lemaire 1995). According to this rule, the estimated individual net premium enlarged by a safety margin Q is equal to

$$m(\theta) = (1+Q)E\left(X_{t+1}\middle|X_1,...,X_t\right) = (1+Q)\frac{\beta + \sum x_j}{\alpha + t - 1} \qquad (14)$$

Assuming that Q=0 from formulas (13) and (14) it follows that a driver who after t years reported claims equal to $\sum_{j=1}^{t} x_j$ in the t+1 year should pay the posterior premium equal to

$$b_{t+1}(x_1,...,x_t) = \frac{(\alpha-1)(\beta+\sum x_j)}{\beta(\alpha+t-1)} \cdot 100\%. \tag{15}$$

IV. EMPIRICAL EXAMPLE

Table 1. and Table 2. present posterior premiums estimated by means of formula (15) for claim size distributions of the Pareto type with various parameters.

Table 1. The posterior net premium $b_{t+1}(x_1,...,x_t)$ in the t+1 year depending on insurance duration time t and the sum $\sum_{j=1}^{t} x_j$ of claims reported in the years 1,...,t. The parameters of the damages' size distribution following the patterns:

A: $EX = 5{,}41;\ DX = 91{,}23;\ n = 10000;$ B: $EX = 5{,}09;\ DX = 78{,}91;\ n = 10000$

C: $EX = 4{,}81;\ DX = 68{,}59;\ n = 10000;$ D: $EX = 4{,}25;\ DX = 49{,}29;\ n = 10000$

$\sum_{j=1}^{t} x_j$ (thous. zl)	T															
	1				2				3				4			
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
below 1	100	100	100	100	67	67	67	67	50	50	50	50	40	40	40	40
1-2	116	116	117	119	77	78	78	79	58	58	59	60	46	47	47	48
2-3	131	133	134	138	87	89	90	92	66	66	67	69	53	53	54	55
3-4	147	149	151	157	98	99	101	105	73	75	76	79	59	60	61	63
4-5	162	165	168	175	108	110	112	117	81	83	84	88	65	66	67	70
5-6	178	181	185	194	118	121	124	130	89	91	93	97	71	73	74	78
6-7	193	198	202	213	129	132	135	142	97	99	101	107	77	79	81	86
7-8	209	214	219	232	139	143	147	155	104	107	110	116	84	86	88	93
8-9	224	230	236	251	150	154	158	168	112	115	119	126	90	92	95	101
9-10	240	247	254	270	160	165	169	180	120	124	127	135	96	99	102	108
10-11	255	263	271	288	170	176	181	193	128	132	136	145	102	105	109	116
11-12	271	279	288	307	181	186	192	205	136	140	144	154	108	112	115	123
12-13	286	296	305	326	191	197	203	218	143	148	153	164	115	119	122	131

Source: Author's calculations.

Table 2. . The posterior net premium $b_{t+1}(x_1,...,x_t)$ in the *t*+1 year depending on insurance duration time *t* and the sum $\sum_{j=1}^{t} x_j$ of claims reported in the years 1,...,*t*. The parameters of the damages' size distribution following the patterns:

E: $EX = 6,37;\ DX = 73,94;\ n = 10000;$ F: $EX = 5,22;\ DX = 41,23;\ n = 10000$

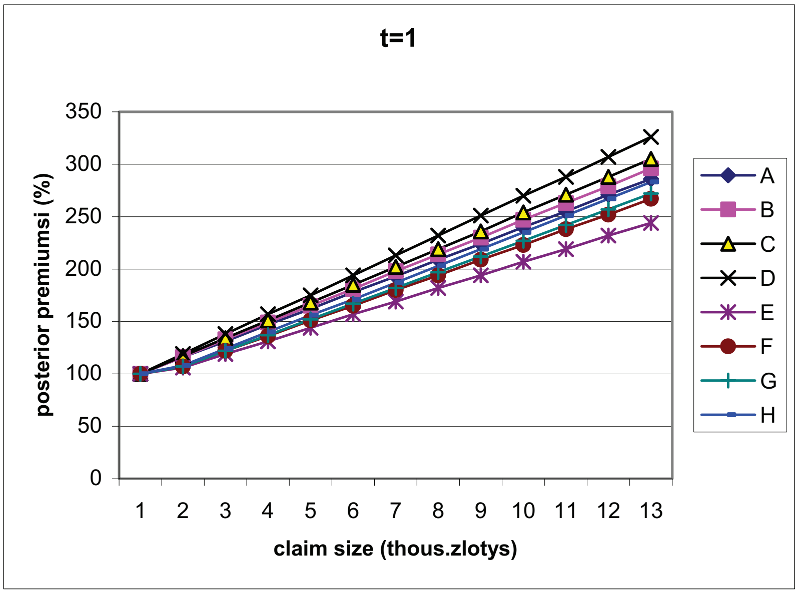
G: $EX = 4,97;\ DX = 35,21;\ n = 10000;$ H: $EX = 4,5;\ DX = 25,06;\ n = 10000$

$\sum_{j=1}^{t} x_j$ (thous. Zl)	T															
	1				2				3				4			
	E	F	G	H	E	F	G	H	E	F	G	H	E	F	G	H
below 1	100	100	100	100	67	67	67	67	50	50	51	51	40	40	40	41
1-2	106	107	107	108	71	72	72	73	53	54	54	55	43	43	44	44
2-3	119	122	122	124	79	82	82	83	60	61	62	63	48	49	50	51
3-4	131	136	137	140	88	91	92	94	66	69	69	71	53	55	56	57
4-5	144	151	152	156	96	101	102	105	72	76	77	79	58	61	62	63
5-6	157	165	167	171	105	111	112	116	79	83	85	87	63	67	68	70
6-7	169	180	182	187	113	121	122	126	85	91	92	95	68	73	74	76
7-8	182	194	197	203	121	130	132	137	91	98	100	103	73	78	80	83
8-9	194	209	212	219	130	140	143	148	97	105	107	111	78	84	86	89
9-10	207	223	227	235	138	150	153	158	104	113	115	119	83	90	92	96
10-11	219	238	242	251	147	159	163	169	110	120	122	128	88	96	98	102
11-12	232	252	257	267	155	169	173	180	116	127	130	136	93	102	104	109
12-13	244	267	272	283	163	179	183	191	123	135	138	144	98	108	110	115

Source: Author's calculations.

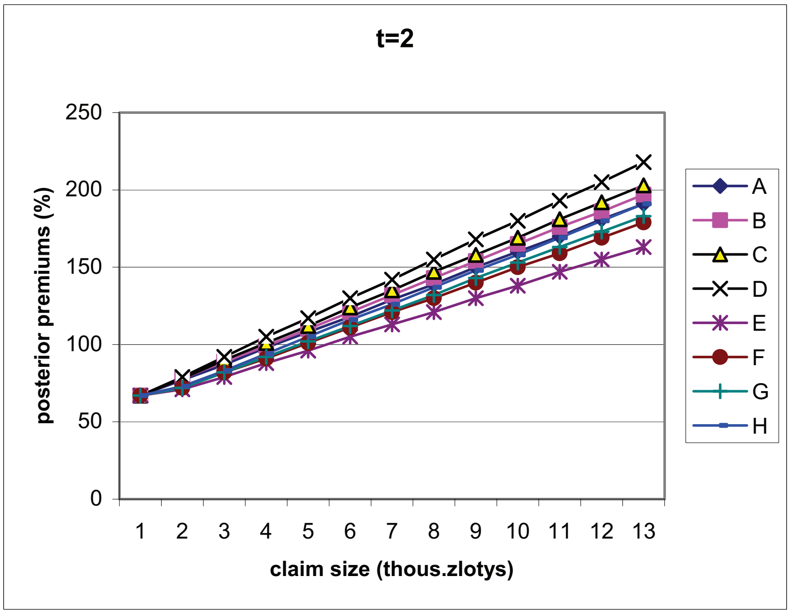
Graphs 1-4 present the posterior premiums from tables (1) and (2).

Graph 1. The posterior net premium in the t+1 year depending on the claim size reported in the years 1,...,t and the parameters of the damages' size distribution.



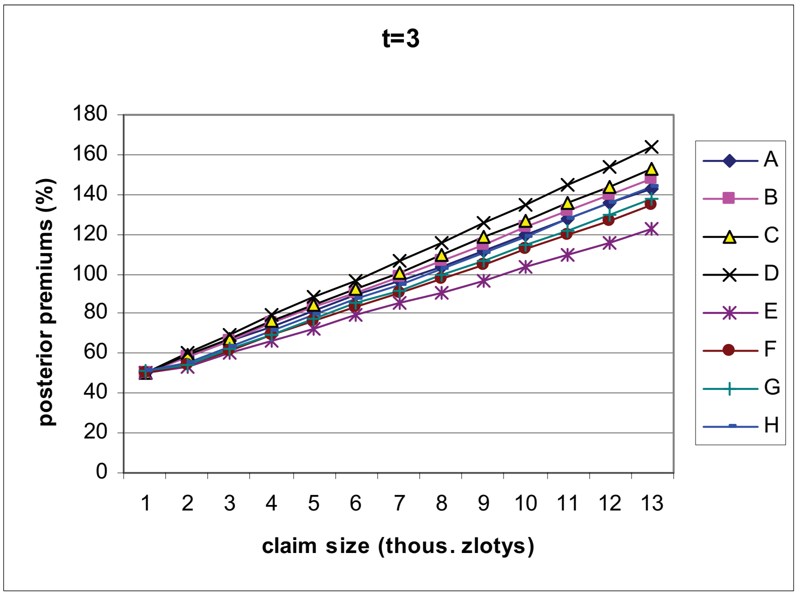
Source: on the basis of tables (1) and (2).

Graph 2. The posterior net premium in the t+1 year depending on the damages' size reported in the years 1,...,t and the parameters of the damages' size distribution.



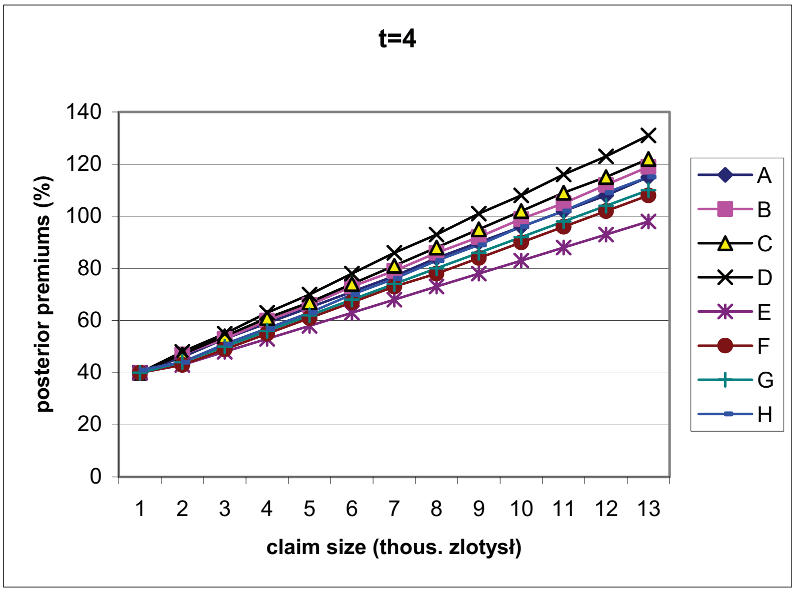
Source: on the basis of tables (1) and (2).

Graph 3. The posterior net premium in the t+1 year depending on the damages' size reported in the years 1,...,t and the parameters of the damages' size distribution.



Source: on the basis of tables (1) and (2).

Graph 4. The posterior net premium in the t+1 year depending on the damages' size reported in the years 1,...,t and the parameters of the damages' size distribution.



Source: on the basis of tables (1) and (2).

Table 3. Variants of the damages' size distribution for different t values according to the posterior premium size (from highest to lowest)

Variants of the claim size distribution	t				Distribution parameters		
	1	2	3	4	*EX*	*DX*	*Me*
D	D-highest premiums	D	D	D	4,25	49,29	1,70
C	C	C	C	C	4,81	68,59	1,74
B	B	B	B	B	5,09	78,91	1,76
A	A	A	A	A	5,41	91,23	1,77
H	H	H	H	H	4,5	25,06	2,38
G	G	G	G	G	4,97	35,21	2,44
F	F	F	F	F	5,22	41,23	2,46
E	E-lowest premiums	E	E	E	6,37	73,94	2,56

Table 4. The credibility coefficient *Z* with respect to observations $x_1,\dots,x_t$ depending on insurance duration time *t* and the sum $\sum_{j=1}^{t} x_j$ of claims reported in the years 1,...,*t*.

$\sum_{j=1}^{t} x_j$ (thous. zlotys)	*t*				
	1	2	3	4	5
below 1	0,50	0,66	0,75	0,80	0,83
1-2	0,50	0,66	0,75	0,80	0,83
2-3	0,50	0,66	0,75	0,80	0,83
3-4	0,50	0,66	0,75	0,80	0,83
4-5	0,50	0,66	0,75	0,80	0,83
5-6	0,50	0,66	0,75	0,80	0,83
6-7	0,50	0,66	0,75	0,80	0,83
7-8	0,50	0,66	0,75	0,80	0,83
8-9	0,50	0,66	0,75	0,80	0,83
9-10	0,50	0,66	0,75	0,80	0,83
10-11	0,50	0,66	0,75	0,80	0,83
11-12	0,50	0,66	0,75	0,80	0,83
12-13	0,50	0,66	0,75	0,80	0,83

Source: .Author's calculations.

One can observe that for an arbitrary *t* the highest values of the posterior premiums are for variant D, and lowest for variant E. The higher the median of the claim size distribution the lower the posterior premium.

The highest premiums will be paid by the drivers with short damage history and big size of procured damages. The insurer should correct the estimated posterior premiums by fixing the maximum and minimum rise depending on other insurers' offers.

The posterior premiums estimated on the basis of damage size estimators are better for the insured who causes even a couple of small damages. In the case of the tariffication based on the number of damages caused by the drivers who cause numerous small damages will have big premium rises. This method of tariffication still has some shortcomings. Causing one big damage inflicts big rises and the trust coefficients are satisfying for drivers with long damage history. What is more, the insurer sees the pooled damage size in the course of t years and does not see the damage size distribution in time.

REFERENCES

Domański Cz., Pruska K., *Nieklasyczne metody statystyczne*, PWE, Warszawa 2000.

Hossack I.B., Pollard J.H., Zehnwirth B., *Introductory statistics with applications in general insurance*, Cambridge 1999.

Jasiulewicz H., Teoria zaufania. Modele aktuarialne, Wydawnictwo AE we Wrocławiu, Wrocław 2005.

Lemaire J. , *Bonus-Malus Systems in Automobile Insurance*, Kluwer Nijhoff, Boston 1995.

Anna Szymańska

ZASTOSOWANIE ESTYMATORÓW BAYESOWSKICH DO SZACOWANIA STAWEK SKŁADKI W UBEZPIECZENIACH KOMUNIKACYJNYCH OC

W pracy przedstawiono zastosowanie estymatorów bayesowskich do taryfikacji *a posteriori* w ubezpieczeniach komunikacyjnych OC. Składki netto wyznaczono metodą wartości oczekiwanej przy założeniu, że rozkład wielkości szkód jest rozkładem Pareto. Porównano otrzymane stawki składek dla rozkładu Pareto wielkości szkód o różnych parametrach.