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 255, 2011

*Anna Szymańska**

THE INFLUENCE OF SAMPLE SIZE ON THE ESTIMATION OF NET PREMIUMS AND NET PREMIUMS' SIZE IN CIVIL RESPONSIBILITY CAR INSURANCE

Abstract. In the paper an application of Bayesian estimators to the posterior tarrification of car civil responsibility (CR) insurance is presented. The net premiums are determined by means of the expected value rule and quantile of order ε rule. The received premiums are compared for different number of sample for the Pareto type distribution of damage size.

Key words: posterior tarrification, Bayes estimators, Pareto distribution.

I. INTRODUCTION

In the car CR insurance premiums are determined in two stages. The first consists in calculating the basic premium using prior factors, the second in posterior tarrification. In the first stage the insurer determins the basic premium on the basis of known risk factors such as driver's age or car production year. In the second stage the insurer estimates the percentage of the basic premium which an individual driver should pay [Lemaire 1995].

The aim of this paper is to present posterior tarrification methods which are based on the size of driver's past damages. In this case the distribution's asymmetry is of great importance. In the paper the net premium and net premium rate are determined by means of the expected value rule and quantile of order 0,5 of the Pareto type distribution of damage size are compared. The net premium rates are estimated by means of Bayesian estimators for different sample sizes generated from the Pareto distribution.

II. THE RULES OF DETERMINING NET PREMIUMS IN INSURANCE

In property and personal insurances the estimation of the net premium is based on the forecast of the number and size of claims i.e. on the assessment of insurance risk. Thus, the net premium is a function of random variable describing damage size.

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

Let $\Pi(X)$ denote the net premium level for protection against the loss of size X, where X is a random variable. The most common rules of determining the premium functional are the following:

1. Pure premium rule (net premium equivalence)

$$\Pi(X) = \mathrm{E}X \tag{1}$$

2. Expected value rule

$$\Pi(X) = (1+Q)EX \tag{2}$$

where Q ≥0 is called a safety coefficient.

3. Variance rule

$$\Pi(X) = EX + QVarX, Q \geq 0 \tag{3}$$

4. Percentile rule (quantile of order ε)

$$\Pi(X) = \min\{x : F(x) \geq 1-\varepsilon\} = F_X^{-1}(1-\varepsilon) \tag{4}$$

In CR car insurance the individual net premium in period t+1 is equal to:

$$\Pi(X) = (EX)\cdot(E\Lambda)\cdot b_{t+1}(k_1,...,k_t) \tag{5}$$

where $\Pi(X)$ – individual net premium in period t+1, (EX) – expected value of a single damage size , $(E\Lambda)$ – expected value of the number of claims, $b_{t+1}(k_1,...,k_t)$ – rate of the estimated net premium [Lemaire 1995].

III. DAMAGE SIZE DISTRIBUTION

Let X_j be random variable denoting the size of claims in year j for a given policy and let $(X_1, X_2,\dots,X_t)$ be the vector of claim size observations for t years for a given policy. Let the distribution of random variable X_j be dependent on parameter θ. We assume that the parameter θ of the insured risk is steady throughout the whole insurance period and is a realization of random variable Θ .

Let us assume that random variables X_j *for* given $\Theta = \theta$, are independent with equal expectations and insured generate losses independently.

Let X_{t+1} be unknown damage size in the year t+1 for the policy given by the vector of observations (X_1, X_2,…,X_t), which can be estimated by means of the Bayesian estimator.

Let random variable X – denoting damage size follow the Pareto distribution with parameters α i β, with the density of the form

$$f(x) = \frac{\alpha\beta^{\alpha}}{(\beta + x)^{\alpha+1}}, \; x > 0, \tag{6}$$

The posterior distribution of parameter Θ has the gamma distribution with parameters $\hat{\alpha}$ and $\hat{\beta}$ of the form

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

where estimators of parameters α and β determnied by means of the moments method are equal to

$$\tilde{\alpha} = \frac{2S_x^2}{S_x^2 - \bar{x}^2} \quad \text{and} \quad \tilde{\beta} = \bar{x}\frac{S_x^2 + \bar{x}^2}{S_x^2 - \bar{x}^2} \tag{8}$$

where $\bar{x}$ is the mean of portfolio claims, S_x^2 is the variance of the portfolio claim size [Domański 2000].

IV. ESTIMATING THE NET PREMIUM RATES

The Bayesian estimator of the parameter Θ, with assumption of the quadratic loss function of the form $L(\theta, a) = (\theta - a)^2$ (where $a = d(x)$ is the decision function), is the conditional expectation of the posterior damage size distribution [Krzyśko 1997] and has the form :

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

Assuming that in equation (5), $EX = \dfrac{\tilde{\beta}}{\tilde{\alpha} - 1}$ and $E\Lambda = 1$ we have

$$\Pi(x) = \frac{\tilde{\beta}}{\tilde{\alpha} - 1} \cdot b_{t+1}(x_1, ..., x_t) \tag{10}$$

Thus, the driver who after t years reported claims of size $\sum_{j=1}^{t} x_j$, should pay the rate of the net premium equal to:

$$b_{t+1}(x_1, ..., x_t) = \frac{\tilde{\alpha} - 1}{\tilde{\beta}} \Pi(X) \cdot 100\% \tag{11}$$

Individual net premium estimated through the expected value rule enlarged by the safety margin Q, is equal to:

$$\Pi(X) = (1 + Q)E\left(X_{t+1} \middle| X_1, ..., X_t\right) = (1 + Q)\frac{\tilde{\beta} + \sum x_j}{\tilde{\alpha} + t - 1} \tag{12}$$

By means of (11) and (12) the driver who after t years reported claim size equal to $\sum_{j=1}^{t} x_j$ in the year t+1 should pay the net premium rate equal to :

$$b_{t+1}(x_1, ..., x_t) = (1 + Q)\frac{(\tilde{\alpha} - 1)(\tilde{\beta} + \sum x_j)}{\tilde{\beta}(\tilde{\alpha} + t - 1)} \cdot 100\% \tag{13}$$

If $Q = 0$, it implies that the net premium rate estimated according to the expected value rule amounts to:

$$b_{t+1}(x_1, ..., x_t) = \frac{(\tilde{\alpha} - 1)(\tilde{\beta} + \sum x_j)}{\tilde{\beta}(\tilde{\alpha} + t - 1)} \cdot 100\% \tag{14}$$

Let us assume that the loss function has the form:

$$L(\theta, a) = |\theta - a| \tag{15}$$

The Bayesian estimator of the parameter Θ with respect to the loss function given by (15) is the median of the posterior distribution of the parameter Θ [Krzyśko 1997]:

$$Me = \hat{\beta}\left(\sqrt[\hat{\alpha}]{2} - 1\right) \tag{16}$$

The individual net premium in the period $t+1$ is equal to:

$$m(\theta) = \tilde{\beta}\left(\sqrt[\tilde{\alpha}]{2} - 1\right) \cdot b_{t+1}(x_1, ..., x_t) \tag{17}$$

Thus, the driver who after t years reported claims of the amount $\sum_{j=1}^{t} x_j$, should pay the rate of the net premium equal to:

$$b_{t+1}(x_1, ..., x_t) = \frac{1}{\tilde{\beta}\left(\sqrt[\tilde{\alpha}]{2} - 1\right)} m(\theta) \cdot 100\% \tag{18}$$

Individual net premium estimated according to the prcentile rule is equal to:

$$m(\theta) = \hat{\beta}\left(\sqrt[\hat{\alpha}]{2} - 1\right) = \left(\tilde{\beta} + \sum x_j\right)\left(\sqrt[\tilde{\alpha}+t]{2} - 1\right) \tag{19}$$

Taking account of (18) and (19) the rate of the premium estimated according to the prcentile rule is equal to:

$$b_{t+1}(x_1, ..., x_t) = \frac{\left(\tilde{\beta} + \sum x_j\right)\left(\sqrt[\tilde{\alpha}+t]{2} - 1\right)}{\tilde{\beta}\left(\sqrt[\tilde{\alpha}]{2} - 1\right)} \cdot 100\% \tag{20}$$

V. EMPIRICAL INVESTIGATION

In the research carried out the rates of the net premiums were estimated by means of the Bayesian estimators for the premiums determined with the classical expected value method (formula 14) and percentile method (formula 20). The research was carried out for the Pareto distribution with the parameters similar to real data on the CR car insurance market. We assumed that the driver who in the first year caused damages below 3 thousand zlotys, next year will be punished with a 10% higher premium – which mimics well the market conditions. In the research, a population with the Pareto distribution was generated, the parameters were equal to : $\alpha = 2{,}3$; $\beta = 2{,}8$; $EX = 4{,}93535$; $DX = 5{,}2745$; $x_{0,5} = 3{,}7681$.

From the population, samples of sizes 50, 100, 200, 300, 500, 1000 were drawn 10000 times. For the samples drawn classical and positional measures were calculated and the Pareto distribution parameters were estimated. The results are presented in Table 1.

Table 1. Measure values for samples drawn from the Pareto distribution with parameters $\alpha = 2{,}3$; $\beta = 2{,}8$; $EX = 4{,}9535$; $D^2X = 27{,}8200$; $x_{0,5} = 3{,}7681$

n	$\widetilde{x}$	$MSE(\hat{\theta})$	$BIAS(\hat{\theta})$	$D^2(\hat{\theta})$	$\widetilde{S}_x^2$	$\widetilde{\alpha}$	$\widetilde{\beta}$
50	4,9516	0,5455	0,016065	0,545242	26,9150	2,382373	2,873168
100	4,9557	0,2889	0,020165	0,288493	28,7841	2,361328	2,857008
150	4,9498	0,1715	0,014265	0,171297	25,5445	2,399690	2,887117
200	4,9507	0,1358	0,015165	0,13557	27,3993	2,376418	2,867439
250	4,9509	0,1098	0,015365	0,109564	27,8628	2,371028	2,862818
300	4,9520	0,0936	0,016465	0,093329	27,8308	2,371540	2,863905
500	4,9498	0,0549	0,014265	0,054697	27,3832	2,376491	2,866981
1000	4,9550	0,0285	0,019465	0,028121	28,0685	2,369203	2,863580

Source: own calculations

Table 2. Measure values for samples drawn from the Pareto distribution with parameters $\alpha = 2{,}3$; $\beta = 2{,}8$; $EX = 4{,}9535$; $D^2X = 27{,}8200$; $x_{0,5} = 3{,}7681$

n	$\widetilde{Me}$	$MSE(\hat{\theta})$	$BIAS(\hat{\theta})$	$D^2(\hat{\theta})$
50	3,8009	0,1797	0,0328	0,17862416
100	3,7866	0,1273	0,0185	0,12695775
150	3,7817	0,1037	0,0136	0,10351504
200	3,7781	0,0883	0,0100	0,08820000
250	3,7757	0,0788	0,0076	0,07874224
300	3,7753	0,0709	0,0072	0,07084816
500	3,7745	0,0553	0,0064	0,05525904
1000	3,7720	0,0390	0,0039	0,03898479

Source: own calculations.

where

n – sample size;

r – number of repetitions;

$\widetilde{x} = \frac{1}{r}\sum_{j}^{r} \overline{x}_j$ – arithmetic mean estimator for sample size n out of r-repetitions;

$\overline{x}_j = \frac{1}{n}\sum_{i}^{n} x_i^{(j)}$ –arithmetic mean for j-th repetition;

$\widetilde{Me} = \frac{1}{r}\sum_{j}^{r} Me_j$ – arithmetic mean of medians out of r- repetitions;

Me_j – mediana of j-th repetition;

$\widetilde{S}_x^2 = \frac{1}{nr}\sum_k^{nr} x_k^2 - (\widetilde{x})^2$ – variance estimator for sample of size n out of r – repetitions;

$\widetilde{\alpha}$ and $\widetilde{\beta}$ estimators of the Pareto distributions parameters;

$MSE(\hat{\theta}) = E((\hat{\theta}-\theta)^2) = \frac{1}{r}\sum_{j=1}^{r}\left(\overline{x}_j - EX\right)^2$ mean square error of estimator $\hat{\theta} = EX$

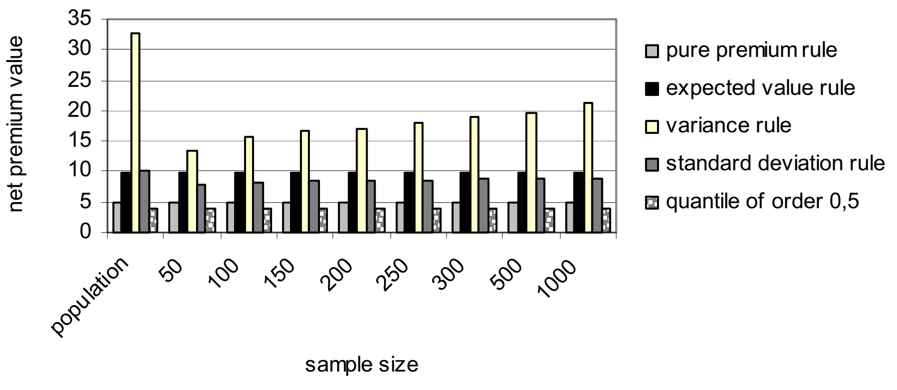
$BIAS(\hat{\theta}) = b(\hat{\theta}) = E\left(\hat{\theta}\right) - \theta = \widetilde{x} - EX$ bias of estimator $\hat{\theta} = EX$

$D^2\left(\hat{\theta}\right) = MSE(\hat{\theta}) - b^2(\hat{\theta})$ estimator's variance

Table 3. Values of net premium calculated with different methods

	Rule of determining net premium				
	Pure premium	Expected value α = 1	variance α = 1	Standard deviation α = 1	Quantile of order 0,5
Population	4,935335	9,87067	32,75569	10,20984	3,7681
n=50	4,9715	9,9430	13,5189	7,8951	3,8096
n=100	4,9467	9,8934	15,6311	8,2154	3,7989
n=150	4,9426	9,8852	16,5939	8,3560	3,7767
n=200	4,9558	9,9116	16,9599	8,4205	3,7773
n=250	4,9623	9,9246	17,8799	8,5564	3,7796
n=500	4,9618	9,9236	19,0573	8,7162	3,7763
n=1000	4,9505	9,9010	19,5903	8,7767	3,7746

Source: own calculations.



Graph 1. Values of net premium calculated with different methods

Source: own calculations.

Table 4. The rate of net premium $b_{t+1}(x_1,...,x_t)$ in year t+1 with respect to the insurance duration time t, sums $S=\sum_{j=1}^{t} x_j$ of claims reported in the years 1,...,t and sample sizes for premiums determined with the expected value rule

Sum of claims S (thous. zl)	n – sample size											
	50				100				200			
	t=1	t=2	t=3	t=4	t=1	t=2	t=3	t=4	t=1	t=2	t=3	t=4
Below 3	110	77	60	49	110	77	59	48	110	77	60	49
3–4	129	91	70	57	128	90	69	57	129	91	70	57
4–5	147	104	80	65	147	103	80	65	147	104	80	65
5–6	166	117	90	74	166	116	90	73	166	117	90	73
6–7	185	130	101	82	184	130	100	81	185	130	100	82
7–8	204	143	111	90	203	143	110	89	204	143	111	90
8–9	222	157	121	98	222	156	120	98	222	156	121	98
9–10	241	170	131	107	241	169	130	106	241	170	131	107
10–11	260	183	141	115	259	182	140	114	260	183	141	115
11–12	279	196	151	123	278	195	151	122	279	196	151	123
12–13	297	209	162	132	297	208	161	131	297	209	161	131
Above 13	316	223	172	140	315	222	171	139	316	222	172	140
	n – sample sizes											
Sum of claims S (thous. zl)	300				500				1000			
	t=1	t=2	t=3	t=4	t=1	t=2	t=3	t=4	t=1	t=2	t=3	t=4
Below 3	110	77	60	48	110	77	60	49	110	77	60	48
3–4	129	90	70	57	129	91	70	57	128	90	70	57
4–5	147	104	80	65	147	104	80	65	147	103	80	65
5–6	166	117	90	73	166	117	90	73	166	117	90	73
6–7	185	130	100	82	185	130	100	82	185	130	100	81
7–8	203	143	110	90	204	143	111	90	203	143	110	90
8–9	222	156	121	98	222	156	121	98	222	156	120	98
9–10	241	169	131	106	241	170	131	107	241	169	131	106
10–11	260	183	141	115	260	183	141	115	259	182	141	114
11–12	278	196	151	123	279	196	151	123	278	196	151	123
12–13	297	209	161	131	297	209	161	131	297	209	161	131
Above 13	316	222	171	139	316	222	172	140	316	222	171	139

Source: own calculations.

Table 5. The rate of net premium $b_{t+1}(x_1,...,x_t)$ in year t+1 with respect to the insurance duration time t, sums $S=\sum_{j=1}^{t} x_j$ of claims reported in the years 1,...,t and sample sizes for premiums determined with the percentile rule (quantile of order 0,5)

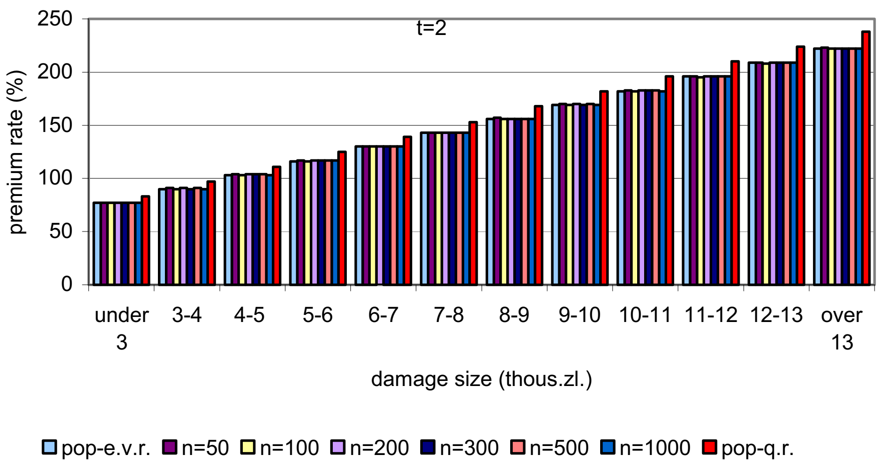
Sum of claims S (thous. zl)	n – sample sizes											
	50				100				200			
	t=1	t=2	t=3	t=4	t=1	t=2	t=3	t=4	t=1	t=2	t=3	t=4
Below 3	110	83	66	55	110	83	66	55	110	83	66	55
3–4	129	97	78	65	129	97	78	65	129	97	78	65
4–5	147	111	89	74	148	111	89	74	148	111	89	74
5–6	166	125	100	84	166	125	100	84	166	125	100	84
6–7	185	139	112	93	185	139	112	93	185	139	112	93
7–8	204	153	123	103	204	153	123	102	204	153	123	103
8–9	222	168	134	112	223	167	134	112	223	168	134	112
9–10	241	182	146	122	241	182	145	121	241	182	146	122
10–11	260	196	157	131	260	196	157	131	260	196	157	131
11–12	279	210	168	141	279	210	168	140	279	210	168	140
12–13	297	224	180	150	298	224	179	150	298	224	180	150
Above 13	316	238	191	159	316	238	191	159	316	238	191	159
Sum of claims S (thous. zl)	n – sample sizes											
	300				500				1000			
	t=1	t=2	t=3	t=4	t=1	t=2	t=3	t=4	t=1	t=2	t=3	t=4
Below 3	110	83	66	55	110	83	66	55	110	83	66	55
3–4	129	97	78	65	129	97	78	65	129	97	78	65
4–5	148	111	89	74	148	111	89	74	148	111	89	74
5–6	166	125	100	84	166	125	100	84	166	125	100	84
6–7	185	139	112	93	185	139	112	93	185	139	112	93
7–8	204	153	123	103	204	153	123	103	204	153	123	103
8–9	223	168	134	112	223	168	134	112	223	168	134	112
9–10	241	182	146	122	241	182	146	122	241	182	146	122
10–11	260	196	157	131	260	196	157	131	260	196	157	131
11–12	279	210	168	140	279	210	168	141	279	210	168	141
12–13	298	224	180	150	298	224	180	150	298	224	180	150
Above 13	316	238	191	159	316	238	191	159	316	238	191	159

Source: own calculations

Table 6. The rate of net premium $b_{t+1}(x_1,...,x_t)$ in year t+1 with respect to the insurance duration time t, sums $S=\sum_{j=1}^{t} x_j$ of claims reported in the years 1,...,t and sample sizes for premiums determined with the expected value rule and the percentile rule (for population)

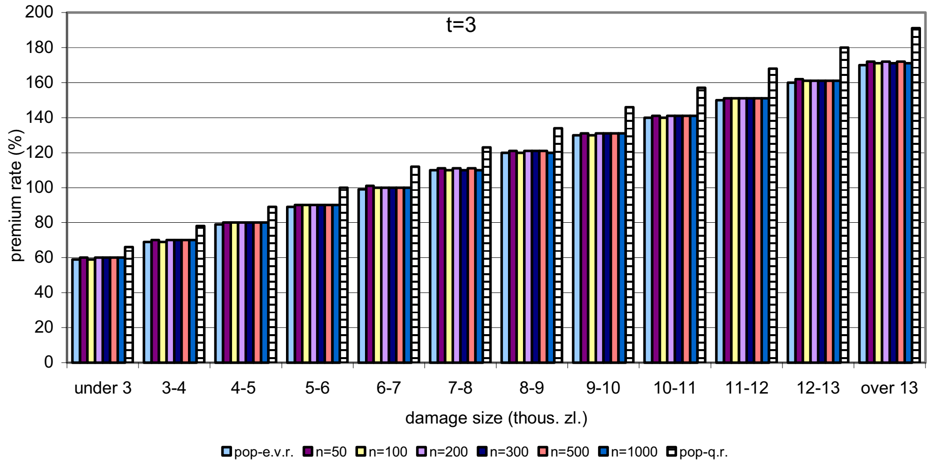
$\sum_{j=1}^{t} x_j$ (thousands zl.)	t							
	1		2		3		4	
	Expected value rule	Percentile rule	Expected value rule	Percentile rule	Expected value rule	Percentile rule	Expected value rule	Percentile rule
below 3	110	110	77	82	59	66	48	55
3–4	129	129	90	97	69	77	56	64
4–5	148	148	103	111	79	88	64	74
5–6	167	167	116	125	89	100	72	83
6–7	186	186	130	139	99	111	81	92
7–8	205	205	143	153	110	122	89	102
8–9	224	224	156	167	120	134	97	111
9–10	243	243	169	182	130	145	105	121
10–11	262	262	182	196	140	156	114	130
11–12	281	281	196	210	150	168	122	140
12–13	300	300	209	224	160	179	130	149
above 13	319	319	222	238	170	190	138	159

Source: own calculations.



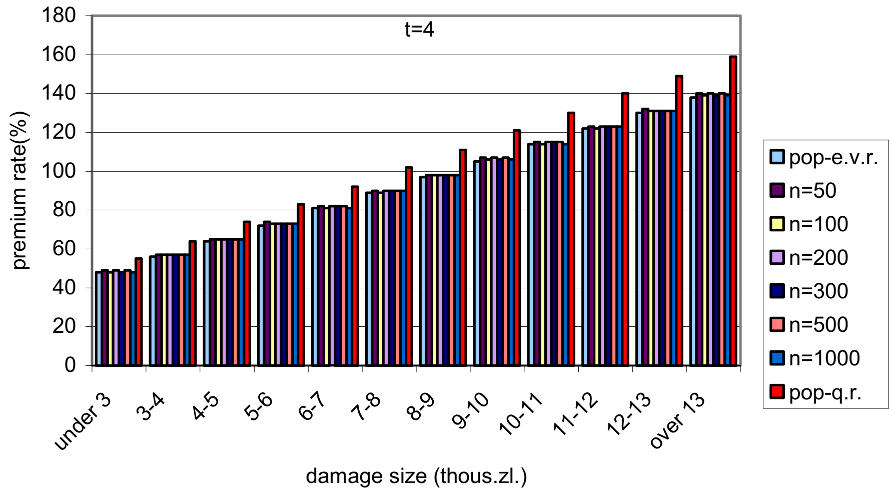
Graph 2. The rate of net premium $b_{t+1}(x_1,...,x_t)$ in year t+1 with respect to the insurance duration time t and the sum of claims reported in the years 1,...,t for premiums determined with the expected value rule [(e.v.r.) for population and samples of different sizes] and percentile rule [(q.r.) for population]

Source: own calculations.



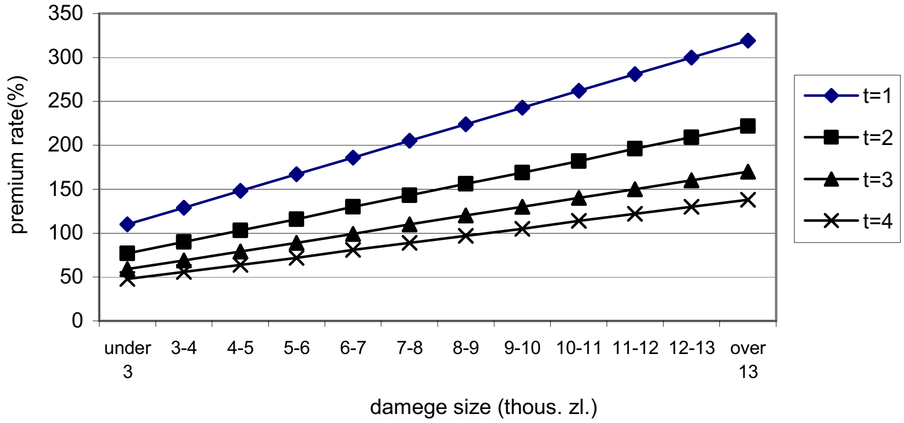
Graph 3. The rate of net premium in year *t*+1 with respect to the insurance duration time *t* and the sum of claims reported in the years 1,...,*t* for premiums determined with the expected value rule [(e.v.r.) for population and samples of different sizes] and percentile rule [(q.r.) for population]

Source: own calculations



Graph 4. The rate of net premium in year *t*+1 with respect to the insurance duration time *t* and the sum of claims reported in the years 1,...,*t* for premiums determined with the expected value rule [(e.v.r.) for population and samples of different sizes] and percentile rule [(q.r.) for population]

Source: own calculations



Graph 5. The rate of net premium in year *t*+1 with respect to the insurance duration time *t* and the sum of claims reported in the years 1,...,*t* for premiums determined with the expected value rule

Source: own calculations.

VI. CONCLUSIONS

On the basis of the research carried out one may state that the pure premium and percentile rules give most stable (independently of the sample size) values of the rate of net premium. (see graph 1). The worst results come from the variance rule. The bigger the sample size the more similar net premium values estimated with different methods from samples for premiums estimated from population parameters. The only exception is the premium estimated with the variance method which is substantially different for samples and population. This means that in the variance rule case the insurer estimating net premium from sample will underestimate it with respect to the portfolio damage size. In CR car insurance, for portfolios with big numbers of risks, determining other parameters than variance may be troublesome. Thus, it is very handy to use a sample drawn from portfolio for example with 500 elements.

One may also state that net premium rates are higher for premiums estimated with the percentile rule – here the quantile of order 0,5. The difference between premium rates found with the help of the methods used gets bigger with the growth of *t*. Thus, the longer the damage history of the insured, the bigger the diffferences between rate premiums found with the methods investigated. However, for both methods the rates get smaller with the growth of *t* (see Graph 5). This is a correct assessment of driver by the system – the driver who in longer time causes damages of a given size is punished with smaller rise than the driver who causes equal damages in a shorter time. The sample size does not have a serious influence upon the estimated premium rates.

In the paper the investigation resluts for only one Pareto distribution were presented. However, the results for the Pareto distribution with other parameters, close to real data, give similar results.

LITERATURE

Domański Cz., Pruska K., *Nieklasyczne metody statystyczne*, PWE, Warszawa 2000.
Krzyśko M., *Statystyka matematyczna cz. II,* Wydawnictwo UAM, Poznań 1997.
Lemaire J. , *Bonus-Malus Systems in Automobile Insurance*, Kluwer Nijhoff, Boston 1995.

Anna Szymańska

WPŁYW WIELKOŚCI PRÓBY NA SZACOWANIE SKŁADEK NETTO ORAZ STAWEK SKŁADKI NETTO W UBEZPIECZENIACH KOMUNIKACYJNYCH OC

W pracy przedstawiono zastosowanie estymatorów bayesowskich do taryfikacji *a posteriori* w ubezpieczeniach komunikacyjnych OC. Składki netto wyznaczono za pomocą zasady wartości oczekiwanej oraz zasady kwantyla rzędu ε. Porównano otrzymane stawki składek dla różnej liczebności próby dla rozkładu wielkości szkód typu Pareto.