Vybrané kredibilné regresné modely v havarijnom poistení

*Erik ŠOLTÉS – Viera PACÁKOVÁ – Tatiana ŠOLTÉSOVÁ**

Selected Credibility Regression Models in Motor Hull Insurance

Abstract

In this article we review credibility regression models. The theory of credibility is a summary of methods used for a calculation and a systematic correction of net premiums. The credibility models are designed for situations where the average claim amount evolves in time or depends on other measurable factors. Methods of regression analysis were first time used in the theory of credibility by Hachemeister. His model (model HM) was adjusted by De Vylder for the purpose of the practical application. Even though the Hachemeister's model does not result in a compromise premium, which is the basic requirement expected from credibility models, it provided important information for De Vylder's model created in 1985. The objective of this article is to show alternatives of an application of DV85 model in the accident automobile insurance, i.e. in the mandatory contractual insurance.

Keywords: *motor hull insurance, credibility model, regression model, credibility formula, Hachemeister model, De Vylder model, individual premium, collective premium*

JEL Classification: C51, C52, C53, G22

Úvod

Teória kredibility patrí medzi moderné prístupy kalkulácie a permanentnej úpravy netto poistného. Vo vývoji teórie kredibility vznikli rôzne druhy kredibilných modelov, ktoré sa líšia v predpokladoch o priemernej výške poistného plnenia a v závislosti od toho aj v náročnosti výpočtu poistného. Na situácie, v ktorých sa priemerná výška poistného plnenia vyvíja v čase, alebo závisí od iných

* Erik ŠOLTÉS – Viera PACÁKOVÁ – Tatiana ŠOLTÉSOVÁ, Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky, Katedra štatistiky, Dolnozemská cesta 1/b, 852 35 Bratislava 5; e-mail: pacakova@euba.sk; soltes@euba.sk

merateľných faktorov, boli odvodené kredibilné regresné modely. Regresnú analýzu začal v teórii kredibility využívať Charles A. Hachemeister [5]. Jeho prvý model síce neposkytuje uspokojivé výsledky, ale predstavuje základ kredibilných regresných modelov, ktoré navrhli ďalší aktuári, medzi inými aj F. De Vylder [2; 3].

1. Hachemeisterov kredibilný regresný model (model HM)

V kredibilných regresných modeloch sa kredibilné netto poistné nepočíta priamo, ale stanoví sa podľa regresného modelu, pre ktorý sa robí kredibilný odhad parametrov. Parametre takéhoto modelu sú dané vzťahom

$$\hat{\beta}(\theta_j) = \mathbf{Z}_j\hat{\bar{\mathbf{B}}}_j + (\mathbf{I} - \mathbf{Z}_j)\hat{\mathbf{b}} \tag{1}$$

kde $\bar{\mathbf{B}}_j$ je vektor parametrov individuálneho regresného modelu a je nositeľom informácií o j-tom riziku, $\mathbf{b}$ je vektor parametrov kolektívneho regresného modelu a je nositeľom informácií o porovnateľných rizikách, $\mathbf{Z}_j$ je matica kredibility a udáva vierohodnosť vektora $\bar{\mathbf{B}}_j$.

Individuálny regresný model pre j-té poistné riziko vystihuje závislosť priemernej výšky poistného plnenia od istých merateľných a v modeli zahrnutých premenných. Vychádza sa pritom len z údajov pozorovaných pre j-té riziko. Kolektívny regresný model vystihuje túto závislosť za celú skupinu porovnateľných rizík. Pretože sa predpokladá, že výška poistného plnenia v čase t nezávisí od poistných plnení v ostatných obdobiach a rozptyl výšky poistných plnení v čase t je nepriamo úmerný veľkosti poistného kmeňa, parametre individuálneho regresného modelu odhadneme zovšeobecnenou metódou najmenších štvorcov podľa vzťahu

$$\hat{\bar{\mathbf{B}}}_j = (\mathbf{Y}_j'\mathbf{P}_j\mathbf{Y}_j)^{-1}\mathbf{Y}_j'\mathbf{P}_j\mathbf{X}_j \tag{2}$$

pričom maticou váh je inverzná matica k diagonálnej matici $\mathbf{P}_j$, ktorá na diagonále obsahuje veľkosti poistného kmeňa P_{jt} zaznamenané pre j-té riziko počas n období $(t = 1, 2, \ldots, n)$; $\mathbf{X}_j = (X_{jn} \ \ldots \ X_{j1})'$ je vektor priemerných poistných plnení pozorovaných pre j-té riziko a $\mathbf{Y}_j$ je matica vysvetľujúcich premenných. Vychádzajúc z údajov za celú skupinu porovnateľných rizík, vektor parametrov kolektívneho regresného modelu odhadneme podľa vzťahu

$$\hat{\mathbf{b}} = (\mathbf{Y}'\mathbf{P}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{P}\mathbf{X} \tag{3}$$

Ak by sme na výpočet netto poistného použili individuálny regresný model, nezohľadnili by sme cenné informácie o porovnateľných rizikách. Na druhej strane netto poistné stanovené podľa kolektívneho regresného modelu je jednotné pre všetky riziká a neprihliada na individuálne rozdiely. Tento problém sa rieši vážením parametrov individuálneho a kolektívneho modelu, pričom váhu vektora $\vec{\mathbf{B}}_j$ determinuje matica kredibility $\mathbf{Z}_j$. Charles A. Hachemeister na jej výpočet odvodil vzťah

$$\mathbf{Z}_j = \mathbf{A}(\mathbf{A} + \sigma^2(\mathbf{Y}_j'\mathbf{P}_j\,\mathbf{Y}_j)^{-1})^{-1} \tag{4}$$

Odhadom rozptylu σ^2 je priemer reziduálnych rozptylov $\hat{\sigma}_j^2$ ($j = 1, 2, \ldots, J$) individuálnych regresných modelov, to znamená

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{J} \hat{\sigma}_j^2}{J} \tag{5}$$

a $\mathbf{A}$ je kovariančná matica parametrov kredibilného modelu. Jej neskresleným odhadom je matica

$$\mathbf{H} = \mathbf{\Pi}^{-1}\left[\mathbf{G} - \hat{\sigma}^2(\mathbf{Y}'\mathbf{P}\mathbf{Y})^{-1})\right] \tag{6}$$

k jej výpočtu sa musia uskutočniť tieto čiastkové výpočty

$$\mathbf{\Pi} = \mathbf{I} - \sum_{j=1}^{J}(\mathbf{Y}'\mathbf{P}\mathbf{Y})^{-1}(\mathbf{Y}_j'\mathbf{P}_j\mathbf{Y}_j)(\mathbf{Y}'\mathbf{P}\mathbf{Y})^{-1}(\mathbf{Y}_j'\mathbf{P}_j\mathbf{Y}_j) \tag{7}$$

$$\mathbf{G} = \sum_{j=1}^{J}(\mathbf{Y}'\mathbf{P}\mathbf{Y})^{-1}(\mathbf{Y}_j'\mathbf{P}_j\mathbf{Y}_j)(\hat{\bar{\mathbf{B}}}_j - \hat{\mathbf{b}})(\hat{\bar{\mathbf{B}}}_j - \mathbf{b})' \tag{8}$$

V prípade, že matica $\mathbf{H}$ nie je symetrická, symetrickú maticu $\mathbf{A}$ odhadneme podľa vzťahu

$$\hat{\mathbf{A}} = \frac{1}{2}(\mathbf{H} + \mathbf{H}') \tag{9}$$

2. De Vylderov kredibilný regresný model (model DV85)

De Vylder v roku 1985 navrhol nahradiť maticu kredibility $\mathbf{Z}_j$ diagonálnou maticou kredibility $\mathbf{Z}_j^{DV}$, a teda parametre kredibilného regresného modelu odhadnúť podľa vzťahu

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}_j)^{DV85} = \mathbf{Z}_j^{DV}\hat{\bar{\mathbf{B}}}_j + (\mathbf{I} - \mathbf{Z}_j^{DV})\hat{\mathbf{b}} \tag{10}$$

Na výpočet vektora $\mathbf{z}_j$, ktorý obsahuje diagonálne prvky matice $\mathbf{Z}_j^{DV}$, odvodil vzťah

$$\mathbf{z}_j = (\mathbf{Q}_j \otimes Var[\vec{\mathbf{B}}_j])^{-1} (\mathbf{Q}_j \mathbf{A})^{d\ 1\ 2} \tag{11}$$

pričom

$$\mathbf{Q}_j = (\mathbf{A} + \sigma^2 (\mathbf{Y}_j' \mathbf{P}_j \mathbf{Y}_j)^{-1})^{-1} \tag{12}$$

a

$$Var[\vec{\mathbf{B}}_j] = \mathbf{A} + \sigma^2 (\mathbf{Y}_j' \mathbf{P}_j \mathbf{Y}_j)^{-1} \tag{13}$$

Prvý diagonálny prvok matice $\mathbf{Z}_j^{DV}$ udáva vierohodnosť lokujúcej konštanty individuálneho regresného modelu a $(i + 1)$ diagonálny prvok matice $\mathbf{Z}_j^{DV}$ kvantifikuje kredibilitu i-tého regresného koeficienta individuálneho modelu. Touto úpravou pôvodnej matice kredibility $\mathbf{Z}_j$ De Vylder dosiahol, že kredibilné poistné je kompromisom medzi netto poistným vypočítaným z individuálneho modelu a netto poistným, ktoré je stanovené podľa kolektívneho regresného modelu. Hachemeisterov kredibilný regresný model neposkytoval kompromisné riešenie medzi individuálnym a kolektívnym netto poistným, preto nebol vhodný na aplikáciu.

3. De Vylderov kredibilný regresný model aplikovaný na normovaných regresoroch

Na jednej strane, kredibilné poistné stanovené podľa De Vylderovho modelu z roku 1985 zohľadňuje údaje o analyzovanom riziku, ale aj o porovnateľných rizikách, na druhej strane stráca vlastnosť invariantnosti, ktorú Hachemeisterov odhad poistného mal. Kolektívny, individuálny aj Hachemeisterov kredibilný regresný model sú invariantné, to znamená že ľubovoľná lineárna transformácia matice vysvetľujúcich premenných $\mathbf{Y}_j$ nemá vplyv na výpočet netto poistného. De Vylderov model z roku 1985 túto vlastnosť stráca, preto voľba inej mernej jednotky niektorej z vysvetľujúcich premenných má za následok inú hodnotu netto poistného. Aby sa výsledky štandardizovali, je vhodné použiť normované

[1] Hadamardov súčin $\otimes$ dvoch matíc **B** a **C** rovnakého typu je matica **D**, ktorej prvok v i-tom riadku a j-tom stĺpci (d_{ij}) sa rovná súčinu prvkov b_{ij} a c_{ij} z i-tého riadku a j-tého stĺpca matíc **B** a **C**, t. j. $d_{ij} = b_{ij} \cdot c_{ij}$.

[2] Vektor $(\mathbf{Q}_j \mathbf{A})^d$ obsahuje diagonálne prvky súčinu matíc $\mathbf{Q}_j$ a **A**.

vysvetľujúce premenné. Vo vzťahoch (2) až (13) sa musí matica vysvetľujúcich premenných pre j-té poistné riziko – $\mathbf{Y}_j$ dôsledne nahradiť maticou normovaných vysvetľujúcich premenných $\mathbf{Y}_j^{*}$[3]. Okrem toho pre každé poistné riziko vznikne vlastný kolektívny model, v ktorom budeme modelovať závislosť priemerného poistného plnenia od normovaných vysvetľujúcich premenných, ktoré sú obsiahnuté v matici $\mathbf{Y}_j^{*}$, t. j. maticu $\mathbf{Y}$ dôsledne nahradíme zodpovedajúcou maticou $\mathbf{Y}_j^{*}$.

4. Aplikácia

V tejto časti príspevku ukážeme možnosti aplikácie De Vylderovho kredibilného regresného modelu z roku 1985 v poistení motorových vozidiel. Vychádzali sme z databázy údajov, ktorá je vytvorená zo štatistiky poistných udalostí v havarijnom poistení istej anonymnej poisťovne.

Tabuľka 1

Priemerné štvrťročné poistné plnenie (X_t) a veľkosť poistného kmeňa (P_t) v jednotlivých regiónoch a v celej poisťovni

t	*Bratislava (BA)*		*Nitra (NR)*		*Trenčín (TN)*		*Banská Bystrica (BB)*		*Košice (KE)*		*Spolu*	
	X_t	P_t	X_t	P_t	X_t	P_t	X_t	P_t	X_t	P_t	$\bar{X}_t$	$P_{t\bullet}$
16	3 302	1 383	2 344	627	3 132	836	3 204	1 062	2 936	979	3 055	4 887
15	3 482	1 442	1 973	656	2 740	917	3 575	1 434	2 350	1 030	2 989	5 479
14	1 909	1 461	2 731	695	3 606	999	3 846	1 732	3 335	1 082	3 109	5 969
13	4 235	1 488	2 996	805	3 911	1 081	3 805	1 964	3 519	1 133	3 771	6 471
12	3 674	1 524	3 198	968	3 299	1 162	3 245	2 139	3 077	1 185	3 313	6 978
11	4 365	1 595	2 896	1 121	3 482	1 244	2 923	2 266	2 960	1 236	3 326	7 462
10	5 955	1 726	2 813	1 264	3 552	1 325	3 571	2 353	3 171	1 288	3 900	7 956
9	6 137	2 017	3 442	1 395	3 798	1 407	3 989	2 407	3 437	1 339	4 288	8 565
8	5 848	2 360	3 704	1 521	4 110	1 481	4 053	2 434	3 800	1 361	4 429	9 157
7	5 140	2 617	3 000	1 633	4 361	1 559	3 805	2 450	3 959	1 433	4 142	9 692
6	4 448	2 908	2 702	1 704	3 972	1 648	4 108	2 483	4 114	1 500	3 950	10 243
5	5 288	3 283	2 554	1 816	4 156	1 756	3 726	2 515	3 369	1 558	4 019	10 928
4	4 551	3 524	3 254	1 892	3 898	1 819	3 917	2 455	3 959	1 631	4 007	11 321
3	5 518	3 841	3 268	1 978	3 776	1 909	3 875	2 558	3 608	1 692	4 248	11 978
2	4 782	4 099	2 640	2 013	3 956	1 973	2 855	2 627	3 793	1 691	3 760	12 403
1	4 393	4 360	2 636	2 058	3 738	2 044	2 426	2 730	2 830	1 704	3 386	12 896

Uvažovali sme s piatimi poistnými rizikami, ktoré prislúchajú piatim regiónom (BA, NR, TN, BB a KE) v územnom členení Slovenska v účtovníctve poisťovne, teda predpokladáme, že v každom regióne je iné poistné riziko, a preto

[3] Hviezdičkou (*) sme označili normované vysvetľujúce premenné, matice, ktoré tieto premenné obsahujú, a všetky odhady a charakteristiky, ktoré boli vypočítané z normovaných vysvetľujúcich premenných.

v každom regióne by mala byť stanovená iná výška poistného. V aplikácii abstrahujeme od ďalších podstatných faktorov, ako napríklad objem motora, resp. výkon motora, počet najazdených kilometrov, účel, na aký sa auto používa atď. Disponovali sme údajmi za roky 1999 – 2002, čo je veľmi krátky časový rad ročných údajov, preto sme sa rozhodli pracovať so štvrťročnými údajmi. Pritom predpokladáme, že poistné sa platí štvrťročne a naším cieľom je stanoviť netto poistné na prvý štvrťrok 2003. Vychádzali sme z údajov o priemernej výške poistného plnenia a veľkosti poistného kmeňa, ktoré sú zhrnuté v tabuľke 1. V stĺpci „spolu“ sú uvedené agregované štatistiky za všetky porovnateľné riziká (regióny), v našom prípade za celú poisťovňu.

Netto poistné pre j-té riziko (región) v čase t by malo zodpovedať priemernému poistnému plneniu, preto budeme modelovať priemerné výšky poistných plnení. Keďže jediný relevantný regresor, ktorý poznáme, je čas t, odhadneme trend priemerných poistných plnení pre j-tý región zovšeobecnenou metódou najmenších štvorcov s váhami $1/P_{jt}$. Po analyzovaní rôznych regresných funkcií sme zvolili kvadratický trend, ktorý najlepšie vysvetľoval variabilitu priemerných štvrťročných poistných plnení. Matica vysvetľujúcich premenných má preto tvar

$$\mathbf{Y} = \mathbf{Y}_{\mathrm{j}} = \begin{pmatrix} 1 & t & t_j^2 \end{pmatrix} = \begin{pmatrix} 1 & 16 & 256 \\ 1 & 15 & 225 \\ 1 & 14 & 196 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{pmatrix}$$

Poznámka: Použili sme označenie t = 16, 15, ..., 1, čo je výhodné, pretože netto poistné na nasledujúce obdobie (t = 0) udáva priamo lokujúca konštanta a nie sú potrebné ďalšie výpočty.

T a b u ľ k a 2

Odhady parametrov individuálnych modelov a ich základné charakteristiky

Parameter	*Bratislava*	*Nitra*	*Trenčín*	*B. Bystrica*	*Košice*
$\hat{B}_0$	**4 075.740**	**2 448.644**	**3 668.910**	**2 616.211**	**3 185.140**
$\hat{B}_1$	386.262	187.640	119.142	290.663	163.551
$\hat{B}_2$	–29.920	–12.375	–10.316	–16.555	–12.690
$\hat{\sigma}_j^2 = MSE$	1 194 254 368	173 236 567	87 344 450	411 660 136	193 519 062
r^2	53.2 %	32.5 %	61.2 %	39.9 %	45.4 %
$D-W$	1.860	1.685	1.941	1.961	1.966

Parametre individuálnych regresných modelov boli odhadnuté v štatistickom balíku Statgraphics Plus (príloha 1 až 5) a sú uvedené v tabuľke 2.

Všetky individuálne regresné modely sú štatisticky významné a aj ich parametre (lokujúca konštanta a regresné koeficienty) sú signifikantné na hladine významnosti 0,1 (príloha 1 až 5). Variabilita priemerných poistných plnení je najlepšie vysvetlená v regióne Trenčín, a to na 61,20 % (r_2 = 0,6120), najhoršie v regióne Nitra, na 32,52 %.

Aby sme modely mohli použiť na prognostické účely, overili sme predpoklady o náhodnej zložke. Vo všetkých modeloch bola vystihnutá heteroskedasticita, čo znamená že sme váhy zvolili vhodne. Na hladine významnosti 0,05 sa potvrdila normalita náhodných chýb aj ich nezávislosť, a to vo všetkých individuálnych modeloch.

Kolektívny regresný model sme odhadli v tvare

$$\hat{\mu}_t = \mathbf{3\,387{,}556} + 220{,}803t - 16{,}418t^2$$

Jeho parametre sú štatisticky významné aj na hladine významnosti 0,01; model ako celok je štatisticky významný na ľubovoľnej hladine významnosti väčšej ako 0,0005 a vysvetľuje variabilitu závislej premennej na 68,88 % (príloha 6). Model vystihol heteroskedasticitu náhodných chýb, čo potvrdili Cochranov test aj Goldfeldov-Quandtov test. Durbinova-Watsonova štatistika má hodnotu 1,541 a je v oblasti prijatia nulovej hypotézy, t. j. $D-W \in (h;\,4-h)$, pričom pre výberový súbor s rozsahom n = 16, pre dva regresory (k = 2) a hladinu významnosti α = 0,01 má horná hranica hodnotu h = 1,086. Normálne rozdelenie náhodných chýb sa tiež potvrdilo.

Pretože individuálne modely aj kolektívny model spĺňajú všetky predpoklady potrebné na prognostickú aplikáciu regresných modelov, môžeme ich použiť na výpočet individuálneho poistného a kolektívneho poistného. Ich hodnoty predstavujú lokujúce konštanty, ktoré sú v tabuľke 2 a v kolektívnom modeli zvýraznené. Na obrázku 1 sú znázornené individuálne regresné modely a kolektívny regresný model, pričom netto poistné na 1. štvrťrok 2003 predstavuje hodnota v období $t = 0$.

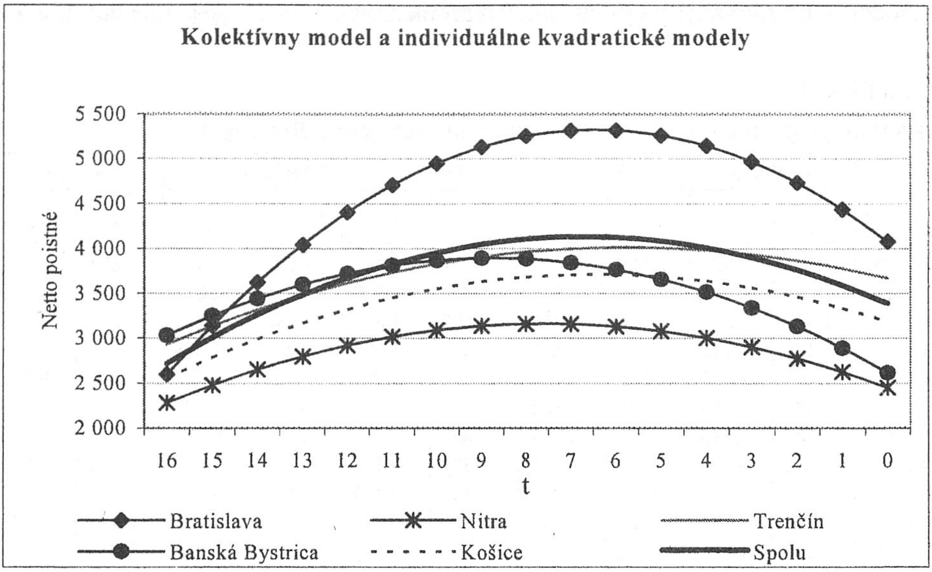
Na odhad individuálnych modelov aj kolektívneho regresného modelu sme použili aj normované vysvetľujúce premenné

$$t_j^* = \frac{t_j - \overline{t_j}}{\hat{\sigma}_{jt}} \quad \text{a} \quad t_j^{2*} = \frac{t_j^2 - \overline{t_j^2}}{\hat{\sigma}_{jt^2}}$$

Takto vzniknú premenné, ktoré sú pre každé riziko odlišné, čo je spôsobené tým, že priemery $\overline{t_j}$ a $\overline{t_j^2}$ aj štandardné odchýlky $\hat{\sigma}_{jt}$ a $\hat{\sigma}_{jt^2}$ sú vážené charakteristiky s váhami P_{jt}.

O b r á z o k 1

Porovnanie kolektívneho a individuálnych regresných modelov



Priemernú hodnotu a štandardnú odchýlku premennej t pre j-tý región sme vypočítali podľa vzťahov

$$\overline{t_j} = \frac{\sum_{t=1}^{n} t \cdot P_{jt}}{P_{t_\bullet}}$$

a $$\hat{\sigma}_{jt} = \sqrt{\overline{t_j^2} - (\overline{t_j})^2}\,, \text{ kde } \quad \overline{t_j^2} = \frac{\sum_{t=1}^{n} t^2 \cdot P_{jt}}{P_{t_\bullet}}$$

a analogicky sme to urobili pre premennú t^2.

T a b u ľ k a 3

Vážené priemery a štandardné odchýlky premenných t a t^2

Charakteristika	***Bratislava***	***Nitra***	***Trenčín***	***Banská Bystrica***	***Košice***
$\overline{t_j}$	6.663	6.850	7.303	7.685	7.698
$\hat{\sigma}_{jt}$	4.446	4.239	4.450	4.367	4.535
$\overline{t_j^2}$	64.159	64.892	73.134	78.127	79.821
$\hat{\sigma}_{jt^2}$	72.386	68.260	73.489	72.244	76.527

Priemerné hodnoty a štandardné odchýlky premenných t a t^2 pre každý región sú uvedené v tabuľke 3.

Hodnoty normovaných vysvetľujúcich premenných t_j^* a t_j^{2*} obsahuje tabuľka 4.

T a b u ľ k a 4

Vysvetľujúce premenné t, t^2 a ich normovania pre jednotlivé regióny

		Bratislava		*Nitra*		*Trenčín*		*Banská Bystrica*		*Košice*	
t	t^2	t_{BA}^*	t_{BA}^{2*}	t_{NR}^*	t_{NR}^{2*}	t_{TN}^*	t_{TN}^{2*}	t_{BB}^*	t_{BB}^{2*}	t_{KE}^*	t_{KE}^{2*}
16	256	2.100	2.650	2.158	2.800	1.954	2.488	1.904	2.462	1.831	2.302
15	225	1.875	2.222	1.923	2.346	1.730	2.067	1.675	2.033	1.610	1.897
14	196	1.650	1.821	1.687	1.921	1.505	1.672	1.446	1.632	1.390	1.518
13	169	1.425	1.448	1.451	1.525	1.280	1.305	1.217	1.258	1.169	1.165
12	144	1.201	1.103	1.215	1.159	1.056	0.964	0.988	0.912	0.949	0.839
11	121	0.976	0.785	0.979	0.822	0.831	0.651	0.759	0.593	0.728	0.538
10	100	0.751	0.495	0.743	0.514	0.606	0.366	0.530	0.303	0.508	0.264
9	81	0.526	0.233	0.507	0.236	0.381	0.107	0.301	0.040	0.287	0.015
8	64	0.301	–0.002	0.271	–0.013	0.157	–0.124	0.072	–0.196	0.067	–0.207
7	49	0.076	–0.209	0.035	–0.233	–0.068	–0.328	–0.157	–0.403	–0.154	–0.403
6	36	–0.149	–0.389	–0.200	–0.423	–0.293	–0.505	–0.386	–0.583	–0.374	–0.573
5	25	–0.374	–0.541	–0.436	–0.584	–0.517	–0.655	–0.615	–0.735	–0.595	–0.716
4	16	–0.599	–0.665	–0.672	–0.716	–0.742	–0.777	–0.844	–0.860	–0.815	–0.834
3	9	–0.824	–0.762	–0.908	–0.819	–0.967	–0.873	–1.073	–0.957	–1.036	–0.925
2	4	–1.049	–0.831	–1.144	–0.892	–1.192	–0.941	–1.302	–1.026	–1.256	–0.991
1	1	–1.274	–0.873	–1.380	–0.936	–1.416	–0.982	–1.531	–1.068	–1.477	–1.030

Parametre individuálnych aj kolektívnych regresných modelov, ktoré vychádzajú z normovaných vysvetľujúcich premenných, sme odhadli v systéme Statgraphics Plus a sú uvedené v prvých dvoch stĺpcoch tabuľky 5. Keďže tieto regresné modely majú vlastnosť invariantnosti, ich kvalita sa nemení, t. j. test štatistickej významnosti parametrov, test štatistickej významnosti modelu ako celku, aj testy na overenie predpokladov o náhodnej zložke modelu majú rovnaké výsledky ako testy v prípade použitia vysvetľujúcich premenných t a t^2. Samozrejme, aj netto poistné stanovené podľa týchto modelov je rovnaké ako podľa pôvodných modelov. Netto poistné na nasledujúce obdobie v tomto prípade nepredstavuje lokujúca konštanta, ale hodnota, ktorú získame po dosadení hodnôt $t_j^* = -\bar{t} / \hat{\sigma}_{jt}$ a $t_j^{2*} = -\overline{t_j^2} / \hat{\sigma}_{jt^2}$ (uvedené v prvých dvoch stĺpcoch tabuľky 6) do odhadnutej paraboly.

Na odhad parametrov kredibilných regresných modelov je potrebné ešte vypočítať matice kredibility. Pretože sme použili normované vysvetľujúce premenné, namiesto kovariančnej matice **A** sme, využitím vzťahov (5) až (9), pre každé poistné riziko vypočítali vlastnú kovariančnú maticu $\mathbf{A}_j^*$. Matice kredibility $\mathbf{Z}_j^{DV*}$ (tab. 5) sme získali aplikovaním vzťahov (11) až (13) na normované vysvetľujúce premenné.

T a b u ľ k a 5

Porovnanie odhadov parametrov kredibilných regresných modelov (DV85*) s odhadmi parametrov kolektívneho a individuálneho regresného modelu v jednotlivých regiónoch

Región (j)	$\hat{\hat{\mathbf{B}}}_{\mathrm{j}}^{*}$	$\hat{\mathbf{b}}_{\mathrm{j}}^{*}$	$\mathbf{Z}_{\mathrm{j}}^{DV*}$	$\hat{\beta}(\theta_j)^{DV85*}$
Bratislava	$\begin{pmatrix} 4\,729,78 \\ 1\,717,18 \\ -2\,165,76 \end{pmatrix}$	$\begin{pmatrix} 3\,805,38 \\ 981,61 \\ -1\,188,45 \end{pmatrix}$	$\begin{pmatrix} 0,576 & 0 & 0 \\ 0 & 0,470 & 0 \\ 0 & 0 & 0,473 \end{pmatrix}$	$\begin{pmatrix} 4\,338,04 \\ 1\,327,58 \\ -1\,650,74 \end{pmatrix}$
Nitra	$\begin{pmatrix} 2\,930,91 \\ 795,44 \\ -844,72 \end{pmatrix}$	$\begin{pmatrix} 3\,834,62 \\ 936,02 \\ -1\,120,71 \end{pmatrix}$	$\begin{pmatrix} 0,582 & 0 & 0 \\ 0 & 0,372 & 0 \\ 0 & 0 & 0,391 \end{pmatrix}$	$\begin{pmatrix} 3\,308,89 \\ 883,74 \\ -1\,012,83 \end{pmatrix}$
Trenčín	$\begin{pmatrix} 3\,784,56 \\ 530,19 \\ -758,08 \end{pmatrix}$	$\begin{pmatrix} 3\,799,31 \\ 982,58 \\ -1\,206,56 \end{pmatrix}$	$\begin{pmatrix} 0,575 & 0 & 0 \\ 0 & 0,393 & 0 \\ 0 & 0 & 0,411 \end{pmatrix}$	$\begin{pmatrix} 3\,790,82 \\ 804,65 \\ -1\,022,06 \end{pmatrix}$
Banská Bystrica	$\begin{pmatrix} 3\,556,46 \\ 1\,269,40 \\ -1\,196,01 \end{pmatrix}$	$\begin{pmatrix} 3\,801,64 \\ 964,30 \\ -1\,186,13 \end{pmatrix}$	$\begin{pmatrix} 0,586 & 0 & 0 \\ 0 & 0,453 & 0 \\ 0 & 0 & 0,464 \end{pmatrix}$	$\begin{pmatrix} 3\,657,88 \\ 1\,102,37 \\ -1\,190,71 \end{pmatrix}$
Košice	$\begin{pmatrix} 3\,431,22 \\ 741,70 \\ -971,10 \end{pmatrix}$	$\begin{pmatrix} 3\,776,71 \\ 1\,001,34 \\ -1\,256,45 \end{pmatrix}$	$\begin{pmatrix} 0,572 & 0 & 0 \\ 0 & 0,392 & 0 \\ 0 & 0 & 0,412 \end{pmatrix}$	$\begin{pmatrix} 3\,579,21 \\ 899,57 \\ -1\,138,91 \end{pmatrix}$

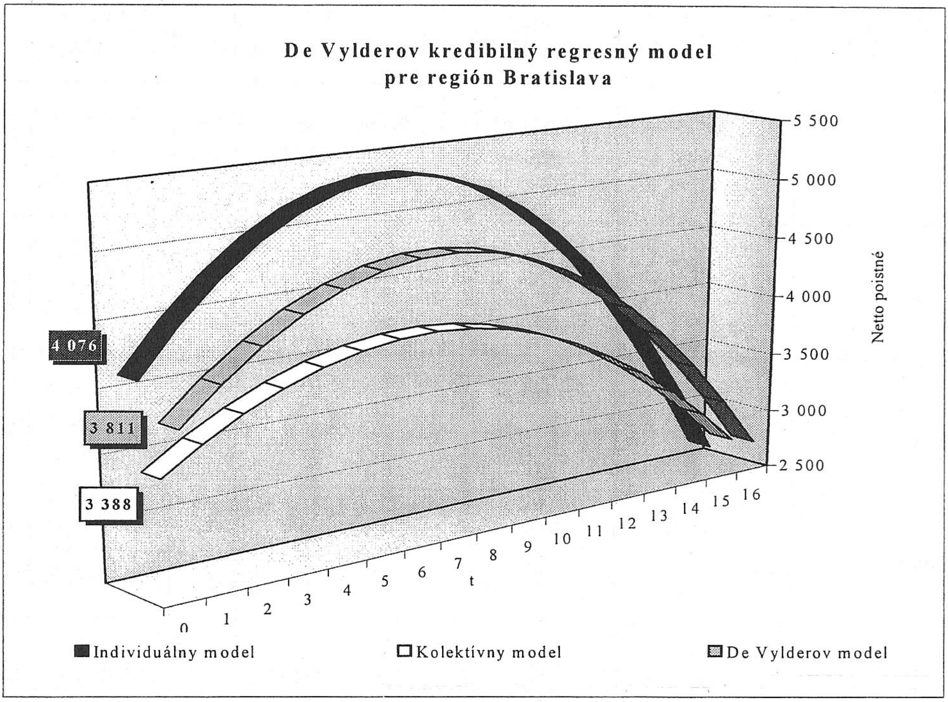
Po výpočte matíc kredibility stačí vektoru $\hat{\hat{\mathbf{B}}}_{\mathrm{j}}^{*}$ prisúdiť váhu $\mathbf{Z}_{\mathrm{j}}^{DV*}$ a vektor $\hat{\mathbf{b}}_{\mathrm{j}}^{*}$ kolektívneho regresného modelu vážiť ($\mathbf{I}-\mathbf{Z}_{\mathrm{j}}^{DV*}$), čím získame odhad vektora parametrov kredibilného regresného modelu DV85* pre j-té riziko. Ten je pre každý región uvedený v poslednom stĺpci tabuľky 5.

Najvierohodnejšie z parametrov individuálnych modelov sú lokujúce konštanty a ich priemerná kredibilita je 51,8 %. Najmenej vierohodné sú podľa matíc $\mathbf{Z}_{\mathrm{j}}^{DV*}$ regresné koeficienty B_{1j}^{*} s priemernou vierohodnosťou 41,6 %. Poznamenajme, že vysvetľujúca premenná t, resp. t_j^{*} mala menší prínos ako premenná t^2, resp. t_j^{*2} (výnimkou bol model v banskobystrickom regióne). Regresné koeficienty B_{2j}^{*} majú priemernú vierohodnosť väčšiu, a to 43 %. Najväčšiu (47,3 %) vierohodnosť má regresný koeficient B_{2j}^{*} individuálneho modelu pre región Bratislava, pri ktorom bola zaznamenaná aj najnižšia p-hodnota v teste o štatistickej významnosti regresných koeficientov. Najnižšiu, 37,2 %-nú vierohodnosť má regresný koeficient B_{1j}^{*} individuálneho modelu pre región Nitra.

Grafické porovnanie individuálneho, kolektívneho a De Vylderovho kredibilného regresného modelu je pre bratislavský región znázornené na obrázku 2.

O b r á z o k 2

Porovnanie De Vylderovho kredibilného regresného modelu s individuálnym a kolektívnym modelom v bratislavskom regióne



Na obrázku 2 je v období $t = 0$ zobrazená čiastka, ktorá predstavuje kredibilné netto poistné pre región Bratislava na 1. štvrťrok 2003. Pre ďalšie regióny sú tieto hodnoty vypočítané v tabuľke 6.

T a b u ľ k a 6

Netto poistné stanovené podľa individuálneho, kolektívneho a kredibilného regresného modelu (DV85*) na 1. štvrťrok 2003

j	*Hodnoty regresorov pre* $t = 0$		*Netto poistné vypočítané podľa modelu*		
	t_j^*	t_j^{2*}	*individuálneho*	*kolektívneho*	*kredibilného*
BA	−1.499	−0.886	4 075.740	3 387.556	**3 811.433**
NR	−1.616	−0.951	2 448.644	3 387.556	**2 843.745**
TN	−1.641	−0.995	3 668.910	3 387.556	**3 487.449**
BB	−1.760	−1.081	2 616.211	3 387.556	**3 005.815**
KE	−1.697	−1.043	3 185.140	3 387.556	**3 240.193**

* * *

Výsledky získané v tabuľke 6 dokazujú vlastnosť invariantnosti individuálnych a kolektívnych regresných modelov. Ak by sme mali stanoviť jednotné netto poistné pre všetky regióny, malo by byť na úrovni približne 3 388 Sk. Pretože môžeme predpokladať rozdielnu škodovosť v jednotlivých regiónoch, či už v dôsledku rozdielnej hustoty premávky, počtu kritických cestných úsekov alebo štruktúry poistených vozidiel, bolo by vhodné poistné regionálne diferencovať.

Netto poistné na 1. štvrťrok 2003, ktoré by mali vodiči jednotlivých regiónov zaplatiť v prípade regionálnej diferenciácie poistného a za predpokladu nezohľadňovania údajov z iných regiónov, je vypočítané podľa individuálneho regresného modelu. Najvyššie netto poistné na 1. štvrťrok 2003 mali platiť, podľa očakávania, vodiči z bratislavského regiónu, v priemere vo výške 4 076 Sk, a najnižšie poistné mali zaplatiť vodiči z nitrianskeho regiónu – v priemere 2 449 Sk. Ak však nechceme zanedbať cenné údaje o porovnateľných rizikách, stanovíme netto poistné podľa kredibilného regresného modelu.

De Vylderov model z roku 1985 zohľadňuje individuálne, ale aj kolektívne údaje a ponúka netto poistné, ktoré je kompromisom medzi individuálnym a kolektívnym poistným, čo dokazujú výsledky v tabuľke 6. Najvyššie netto poistné, v priemernej výške 3 811 Sk, mali za prvý štvrťrok 2003 zaplatiť vodiči z regiónu Bratislava. Najmenšiu čiastku mali, podľa kredibilného regresného modelu, zaplatiť vodiči nitrianskeho regiónu, kde mala byť priemerná sadzba nižšia takmer o 1 000 Sk.

Literatúra

[1] BÜHLMANN, H. – GISLER, A.: Credibility in the Regression Case Revisited. ASTIN Bulletin, *27*, 1997, č. 1, s. 81 – 98.

[2] DE VYLDER, F.: Regression Model with Scalar Credibility Weights. Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker, Heft 1, 1981, s. 27 – 39.

[3] DE VYLDER, F.: Non-linear Regression in Credibility Theory. Insurance. Mathematics and Economics, *4*, 1985, s. 163 – 172.

[4] DRAPER, N. – SMITH, H.: Applied Regression Analysis. New York: J. Wiley & Sons 1985.

[5] HACHEMEISTER, Ch. A.: Credibility for Regression Models with Application to Trend. In: P. Kahn (ed.): Credibility: Theory and Applications. New York: Academic Press 1975.

[6] PACÁKOVÁ, V.: Aplikovaná poistná štatistika. Bratislava: ELITA 2000.

[7] PACÁKOVÁ, V. – ŠOLTÉS, E.: Možnosti aplikácie kredibilných regresných modelov. Acta Oeconomica Cassoviensia, 2004, č. 8, s. 186 – 196.

[8] ŠOLTÉS, E.: De Vylderove odhady parametrov kredibilného regresného modelu. In: Kvantitatívne metódy v ekonómii a podnikaní. [Zborník príspevkov 8. medzinárodnej vedeckej konferencie.] Bratislava: Vydavateľstvo Ekonóm 2002, s. 561 – 565.

Prílohy

Príloha 1
Odhad individuálneho regresného modelu v regióne Bratislava

Multiple Regression Analysis

Dependent variable: XtBA

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	4075.74	495.417	8.22689	0.0000
t	386.262	153.474	2.51679	0.0258
t2	-29.9197	9.42574	-3.17425	0.0073

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1.76745E10	2	8.83725E9	7.40	0.0072
Residual	1.55253E10	13	1.19425E9		
Total (Corr.)	3.31998E10	15			

R-squared = 53.2367 percent
R-squared (adjusted for d.f.) = 46.0424 percent
Standard Error of Est. = 34558.0
Mean absolute error = 473.508
Durbin-Watson statistic = 1.85954

Príloha 2
Odhad individuálneho regresného modelu v regióne Nitra

Multiple Regression Analysis

Dependent variable: XtN

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	2448.64	266.218	9.19788	0.0000
t	187.64	79.9451	2.34711	0.0354
t2	-12.3751	4.96488	-2.49253	0.0270

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1.08514E9	2	5.4257E8	3.13	0.0776
Residual	2.25208E9	13	1.73237E8		
Total (Corr.)	3.33722E9	15			

R-squared = 32.5163 percent
R-squared (adjusted for d.f.) = 22.1342 percent
Standard Error of Est. = 13161.9
Mean absolute error = 272.344
Durbin-Watson statistic = 1.68528

Príloha 3
Odhad individuálneho regresného modelu v regióne Trenčín

```
Multiple Regression Analysis
-----------------------------------------------------------------------------
Dependent variable: XtT
-----------------------------------------------------------------------------
                                        Standard          T
Parameter                Estimate         Error       Statistic       P-Value
-----------------------------------------------------------------------------
CONSTANT                  3668.91       187.971         19.5185        0.0000
t                         119.142        55.343         2.15279        0.0507
t2                       -10.3157       3.35124        -3.07817        0.0088
-----------------------------------------------------------------------------

                          Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Model                  1.79098E9       2    8.9549E8       10.25       0.0021
Residual               1.13548E9      13   8.73444E7
-----------------------------------------------------------------------------
Total (Corr.)          2.92646E9      15

R-squared = 61.1996 percent
R-squared (adjusted for d.f.) = 55.2303 percent
Standard Error of Est. = 9345.83
Mean absolute error = 186.002
Durbin-Watson statistic = 1.9405
```

Príloha 4
Odhad individuálneho regresného modelu v regióne Banská Bystrica

```
Multiple Regression Analysis
-----------------------------------------------------------------------------
Dependent variable: XtBB
-----------------------------------------------------------------------------
                                        Standard          T
Parameter                Estimate         Error       Statistic       P-Value
-----------------------------------------------------------------------------
CONSTANT                  2616,21       347,489          7,5289        0,0000
t                         290,663       99,4976          2,9213        0,0119
t2                        -16,555       6,01475         -2,7524        0,0165
-----------------------------------------------------------------------------

                          Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Model                  3,55408E9       2   1,77704E9        4,32       0,0365
Residual               5,35158E9      13    4,1166E8
-----------------------------------------------------------------------------
Total (Corr.)          8,90567E9      15

R-squared = 39,9081 percent
R-squared (adjusted for d.f.) = 30,6632 percent
Standard Error of Est. = 20289,4
Mean absolute error = 325,924
Durbin-Watson statistic = 1,96054
```

Príloha 5
Odhad individuálneho regresného modelu v regióne Košice

```
Multiple Regression Analysis
-----------------------------------------------------------------------------
Dependent variable: XtK
-----------------------------------------------------------------------------
                                          Standard           T
Parameter                    Estimate       Error        Statistic       P-Value
-----------------------------------------------------------------------------
CONSTANT                      3185.14     298.485          10.671         0.0000
t                             163.551     85.4548         1.91389         0.0779
t2                           -12.6896     5.06401        -2.50584         0.0263
-----------------------------------------------------------------------------

                           Analysis of Variance
-----------------------------------------------------------------------------
Source              Sum of Squares     Df  Mean Square     F-Ratio       P-Value
-----------------------------------------------------------------------------
Model                    2.09163E9      2    1.04581E9        5.40        0.0196
Residual                 2.51575E9     13    1.93519E8
-----------------------------------------------------------------------------
Total (Corr.)            4.60738E9     15

R-squared = 45.3974 percent
R-squared (adjusted for d.f.) = 36.997 percent
Standard Error of Est. = 13911.1
Mean absolute error = 315.189
Durbin-Watson statistic = 1.96627
```

Príloha 6
Odhad kolektívneho regresného modelu

```
Multiple Regression Analysis
-----------------------------------------------------------------------------
Dependent variable: Xt.
-----------------------------------------------------------------------------
                                          Standard           T
Parameter                    Estimate       Error        Statistic       P-Value
-----------------------------------------------------------------------------
CONSTANT                      3387.56     197.933         17.1147         0.0000
t                             220.803     58.6669         3.76366         0.0024
t2                           -16.4183     3.56822        -4.60124         0.0005
-----------------------------------------------------------------------------

                           Analysis of Variance
-----------------------------------------------------------------------------
Source              Sum of Squares     Df  Mean Square     F-Ratio       P-Value
-----------------------------------------------------------------------------
Model                   1.73907E10      2    8.69533E9       14.38        0.0005
Residual                  7.8589E9     13    6.04531E8
-----------------------------------------------------------------------------
Total (Corr.)           2.52496E10     15

R-squared = 68.8751 percent
R-squared (adjusted for d.f.) = 64.0867 percent
Standard Error of Est. = 24587.2
Mean absolute error = 180.56
Durbin-Watson statistic = 1.54102
```