Uniwersytet Ekonomiczny w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 242 · 2015

Ekonomia 3

Alicja Wolny-Dominiak[1]

Uniwersytet Ekonomiczny w Katowicach
Wydział Ekonomii
Katedra Metod Statystyczno-Matematycznych
w Ekonomii
alicja.wolny-dominiak@ue.katowice.pl

Tadeusz Czernik

Uniwersytet Ekonomiczny w Katowicach
Wydział Finansów i Ubezpieczeń
Katedra Matematyki Stosowanej w Ekonomii
tadeusz.czernik@ue.katowice.pl

Jan Acedański

Uniwersytet Ekonomiczny w Katowicach
Wydział Ekonomii
Katedra Metod Statystyczno-Matematycznych w Ekonomii
jan.acedanski@ue.katowice.pl

MODEL REGRESYJNY WARTOŚCI POJEDYNCZEJ SZKODY UWZGLĘDNIAJĄCY POLISY BEZSZKODOWE

Streszczenie: Celem pracy jest zaproponowanie rozkładu prawdopodobieństwa, w którym dowolny rozkład dyspersyjnej rodziny rozkładów wykładniczych zostaje rozszerzony o wartości zerowe (ozn. ZA-rozkład). Rozkład ten wykorzystywany jest dalej do konstrukcji modelu regresyjnego mającego zastosowanie w prognozowaniu wartości pojedynczej szkody w masowym portfelu polis ubezpieczeniowych (np. komunikacyjne czy nieruchomości).

Słowa kluczowe: ZA-rozkład, GLM, portfel polis.

Wprowadzenie

W działalności ubezpieczeniowej, w szczególności w ubezpieczeniach majątkowych, istotnym zagadnieniem jest prognozowanie wartości pojedynczej szkody w portfelu majątkowych polis ubezpieczeniowych. Uzyskane prognozy stosowane są dalej w takich zagadnieniach ubezpieczeniowych, jak: taryfikacja [Antonio i Beirlant, 2006; Wolny-Dominiak, 2009; Ohlsson i Johansson, 2010;

[1] Praca częściowo finansowana przez grant Narodowego Centrum Nauki (nr NN 111461540).

Antonio i Valdez, 2012), szacowanie rezerw szkodowych [Taylor, 2000; Wolny, 2003, s. 678-684; Pobłocka, 2011, s. 173-189] czy zarządzanie ryzykiem. Jednym z podejść w prognozowaniu jest podejście modelowe, w którym analizuje się wartości szkód dla pojedynczych polis wraz z czynnikami, które mogą mieć wpływ na wystąpienie szkody oraz jej wysokość. Najczęściej budowane są modele regresyjne, a w szczególności uogólnione modele liniowe (ozn. GLM) [por. Nelder i Wedderburn, 1972]. W ubezpieczeniowych modelach GLM zmienną objaśnianą jest m.in. wartość pojedynczej szkody. W przypadku gdy w portfelu ubezpieczeniowym występuje duża liczba polis, dla których nie wystąpiła żadna szkoda (dalej nazywanych polisami bezszkodowymi), rodzi się problem określenia rozkładu zmiennej losowej reprezentującej wartość szkód dla pojedynczej polisy w modelu GLM. Jednym z możliwych podejść jest pominięcie polis bezszkodowych i wykorzystywanie informacji dotyczących jedynie tych polis, dla których wystąpiła co najmniej jedna szkoda zakładając asymetryczny rozkład gamma [MacCullagh i Nelder, 1989; Ostasiewicz i Dębicka, 2004; Ohlsson i Johansson, 2010]. Jednak wydaje się, iż prowadzi to do utraty pewnych istotnych informacji zawartych w portfelu polis, jeśli chodzi np. o czynniki wpływające na wystąpienie szkody oraz częstość szkód dla danej polisy.

Celem pracy jest zaproponowanie rozkładu prawdopodobieństwa, w którym rozszerzamy dowolny rozkład dla $Y > 0$ o wartości zerowe (ozn. ZA-rozkład) oraz konstrukcja modelu regresyjnego służącego do prognozowania wartości pojedynczej szkody w portfelu polis. Problemem predykcji z wykorzystaniem modeli liniowych będącymi przypadkami szczególnymi modelu GLM w problemach ekonomicznych zajmowali się m.in. Rao [2003] i Żądło [2008].

W pierwszej części pracy charakteryzujemy ogólnie ZA-rozkład oraz rozpatrujemy przypadek rozkładu gamma. Dalej prezentujemy model regresyjny do szacowania składki czystej dla pojedynczej polisy. Postać tego modelu jest zbliżona do modelu GLM, jednak formalnie nie jest to model GLM, gdyż nie jest spełnione założenie stanowiące, iż rozkład zmiennej objaśnianej pochodzi z dyspersyjnej rodziny rozkładów wykładniczych. W ostatniej części pracy przedstawiamy studium przypadku na danych zaczerpniętych w literatury przedmiotu [de Jong i Heller, 2008]. Do obliczeń wykorzystujemy program R [R_Core_Team, 2013].

1. Charakterystyka ZA-rozkładu

W celu zdefiniowana ogólnej postaci ZA-rozkładu prawdopodobieństwa oznaczmy przez Y zmienną losową o wartościach nieujemnych, $Y \geq 0$. Dystrybuanta ZA-rozkładu zmiennej Y ma wtedy postać:

$$F_Y(y) = \begin{cases} 0 \quad dla\ y \le 0 \\ v + (1-v)F_{Y>0}(y) \;\; dla\ y > 0 \end{cases} \tag{1}$$

lub równoważnie

$$F_Y(y) = vI_{y>0} + (1-v)F_{Y>0}(y), \tag{2}$$

gdzie: $supp(F_{Y>0}) = \mathbb{R}^{+}$, $F_Y(y / Y > 0) = F_{Y>0}(y)$ jest dowolnym rozkładem, v oznacza prawdopodobieństwo nie wystąpienia szkody oraz

$$I_{Y>0} = \begin{cases} 0 \;\; dla \;\; y \le 0 \\ 1 \;\; dla \;\; y > 0 \end{cases}.$$

Zatem funkcja gęstości jest wyrażeniem postaci:

$$f_Y(y) = v\delta_0 + (1-v) f_{Y>0}(y), \tag{3}$$

gdzie: $\delta_0 = \delta(y)$ jest miarą atomową zwaną również deltą Diraca [Dirac, 1982]. Miara dana jest wzorem:

$$\delta(y) = \begin{cases} +\infty, y = 0 \\ 0, y \ne 0 \end{cases},$$

i zachodzi:

$$\int_{-\infty}^{+\infty} \delta(y) dy = 1. \tag{4}$$

Korzystając ze wzorów (3) oraz (4) mamy:

$$\int_{-\infty}^{+\infty} f_Y(y) dy = v \int_{-\infty}^{+\infty} \delta(y) dy + (1-v) \int_{-\infty}^{+\infty} f_{Y>0}(y) dy = 1. \tag{5}$$

Rozpatrywana zmienna *Y* jest mieszaną zmienną losową posiadającą komponentę dyskretną oraz ciągłą. Łatwo wyprowadzić dwa pierwsze momenty ZA-rozkładu. Wartość oczekiwaną z *g*(*Y*) obliczamy ze wzoru (o ile wartość oczekiwana istnieje):

$$E[g(Y)] = \nu \int_{-\infty}^{+\infty} g(y) d(I_{Y>0}) + (1-\nu) \int_{-\infty}^{+\infty} g(y) dF_{Y>0} .$$

Dalej

$$E[g(Y)] = \nu g(0) + (1-\nu) \int_{-\infty}^{+\infty} g(y) f_{Y>0}(y) dy ,$$

gdzie: $E[g(Y) \mid Y > 0] = \int_{-\infty}^{+\infty} g(y) f_{Y>0}(y) dy$.

Skąd łatwo znajdujemy:

$$E[Y] = (1-\nu) \int_{-\infty}^{+\infty} y f_{Y>0}(y) dy = (1-\nu) E[Y \mid Y > 0] = (1-\nu)\mu ,$$

dla parametru μ będącym wartością oczekiwaną Y pod warunkiem $Y > 0$. Interpretując, parametr ten jest wartością oczekiwaną wartości pojedynczej szkody, o ile do niej doszło. Przechodząc do wariancji wyprowadźmy:

$$E[Y^2] = (1-\nu) E[Y^2 \mid Y > 0],$$
$$D^2(Y) = E[Y^2] - (E[Y])^2 = (1-\nu) E[Y^2 \mid Y > 0] - (1-\nu)^2 (E[Y \mid Y > 0])^2 .$$

Ostatecznie znajdujemy:

$$D^2(Y) = (1-\nu) D[Y^2 \mid Y > 0] - (1-\nu)\nu (E[Y \mid Y > 0])^2 .$$

Zatem wariancja zmiennej Y ma postać:

$$D^2(Y) = (1-\nu)[\sigma^2 - \nu\mu^2] , \qquad (6)$$

gdzie: $D[Y^2 \mid Y > 0] = \sigma^2$ jest wariancją Y pod warunkiem $Y > 0$.

Interesującym nasz przypadkiem ZA-rozkładu znajdującym zastosowanie w zagadnieniach ubezpieczeniowych jest przyjęcie założenia, iż rozkład dla części dodatniej $Y > 0$ jest rozkładem gamma z parametrami (k, θ) o gęstości:

$$f^{g}_{Y|Y>0}(y, \mu, \alpha) = \left(\frac{\alpha}{\mu} \right)^{\alpha} \frac{y^{\alpha-1} e^{\frac{-\alpha y}{\mu}}}{\Gamma(\alpha)} . \qquad (7)$$

Dla postaci (7) rozkładu gamma, dwa pierwsze momenty wynoszą:

$$E[Y \mid Y > 0] = \frac{\alpha^2}{\mu},\ D^2(Y \mid Y > 0) = \frac{\alpha^3}{\mu^2}.$$

Możliwa jest również inna parametryzacja rozkładu gama, jednak wzór (7) jest najczęściej stosowany w modelach GLM.

2. Model regresyjny do prognozowania wartości pojedynczej szkody w portfelu ubezpieczeniowym

Niech $Y_1, \dots, Y_n$ będą niezależnymi zmiennymi losowymi reprezentującymi wartość pojedynczej szkody w portfelu polis ubezpieczeniowych przyjętymi w modelu regresyjnym jako zmienne objaśniane o realizacjach $y_1, \dots, y_n$. Klasycznie w modelowaniu wartości pojedynczej szkody przyjmuje się rozkład gamma zmiennej Y_i, $i = 1, \dots, n$. W celu wykorzystania informacji dotyczących polis bezszkodowych, w modelu regresyjnym zakładamy ZA-rozkład zmiennej Y_i postaci:

$$f_{Y_i}(y_i) = v_i \delta_0 + (1 - v_i) f^g_{Y_i > 0}(y_i). \tag{8}$$

Dalej niech $X_1, \dots, X_k$ będą jakościowymi zmiennymi objaśniającymi, które interpretujemy jako czynniki wpływające na wartość pojedynczej szkody w portfelu (np. czynniki charakteryzujące osobę ubezpieczającą oraz czynniki charakteryzujące przedmiot ubezpieczenia, tj. pojazd, mieszkanie, dom). Przyjmujemy również, iż pewne czynniki wpływają również na wystąpienie szkody i dla ustalenia uwagi oznaczmy je poprzez $Z_1, \dots, Z_q$. W proponowanym modelu regresyjnym oznaczmy $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$ jako efekty dla wartości pojedynczej szkody oraz $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_q)^T$ jako efekty dla prawdopodobieństwa nie wystąpienia szkody. Model regresyjny definiujemy następująco:

$$E[Y_i \mid X_1, \dots, X_k, Z_1, \dots, Z_q] = [1 - v_i(\boldsymbol{\gamma}, Z_1, \dots, Z_q)]\mu_i(\boldsymbol{\beta}, X_1, \dots, X_k). \tag{9}$$

W modelu (10) parametry μ_i są uzależnione od zmiennych objaśniających za pomocą funkcji logarytmicznej natomiast parametry v_i – funkcji logitowej. Zatem mamy:

$$\mu_i(\boldsymbol{\beta}, X_1, \dots, X_k) = e^{\mathbf{x}_i^T \boldsymbol{\beta}},$$

$$\nu_i(\boldsymbol{\gamma}, Z_1, \ldots, Z_q) = \frac{e^{\mathbf{z}_i^T \boldsymbol{\gamma}}}{1 + e^{\mathbf{z}_i^T \boldsymbol{\gamma}}} = \frac{1}{1 + e^{-\mathbf{z}_i^T \boldsymbol{\gamma}}},$$

gdzie: $\mathbf{x}_\mathbf{i}^\mathbf{T}$ oznacza *i*-ty wiersz macierzy modelu $\mathbf{X}$ dla zmiennych $X_1, \ldots, X_k$ oraz $\mathbf{z}_\mathbf{i}^\mathbf{T}$ oznacza *i*-ty wiersz macierzy modelu $\mathbf{Z}$ dla zmiennych $Z_1, \ldots, Z_q$.

Do estymacji parametrów modelu regresyjnego stosujemy metodę największej wiarygodności. Funkcja wiarygodności ma zatem postać:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^{N-N_s} \nu_i \prod_{j=N-N_s+1}^{N} (1-\nu_i) f_{Y>0}(y_j, \mu_j), \tag{10}$$

gdzie: N jest liczbą polis w portfelu natomiast N_s – liczbą szkód ($Y > 0$). Powyższą funkcję wiarygodności można zapisać:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = [\prod_{i=1}^{N-N_s} \nu_i \prod_{i=N-N_s+1}^{N} (1-\nu_i)] \prod_{j=N-N_s+1}^{N} f_{Y>0}(y_j, \mu_j), \tag{11}$$

Zatem funkcja $L(\cdot)$ dekomponuje się na funkcje $L_1(\cdot)$ oraz $L_2(\cdot)$ następująco:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = L_1(\boldsymbol{\beta}) L_2(\boldsymbol{\gamma}), \tag{12}$$

gdzie: $L_1(\boldsymbol{\beta})$ zależy tylko od $\boldsymbol{\beta}$ natomiast $L_2(\boldsymbol{\gamma})$ zależy jedynie od γ. Przechodząc dalej do postaci logarytmicznej we wzorze (5), funkcja log-wiarygodności może być zapisana jako dwie odrębne funkcje:

$$\log L = \log L_1(\boldsymbol{\beta}) + \log L_2(\boldsymbol{\gamma}). \tag{13}$$

Warunki konieczne wystąpienia ekstremum:

$$\frac{\partial \log L}{\partial \beta_i} = \frac{\partial \log L_1}{\partial \beta_i} = 0. \tag{14}$$

$$\frac{\partial \log L}{\partial \gamma_i} = \frac{\partial \log L_2}{\partial \gamma_i} = 0. \tag{15}$$

Hesjan jest macierzą blokowo-diagonalno, ponieważ zachodzi:

$$H(\log L)_{\beta\gamma} = \frac{\partial^2 \log L}{\partial \beta_m \partial \gamma_n} = 0. \tag{16}$$

Z własności wyznacznika (dla macierzy diagonalno-klatkowej) wynika, że cała macierz jest ujemnie określona, jeżeli obie kwadratowe, diagonalno-klatkowe macierze są ujemnie określone. Wraz z (14) i (15) oznacza to, że możemy poszukiwać maksimum osobno dla funkcji $\log L_1(\boldsymbol{\beta})$ oraz $\log L_2(\boldsymbol{\gamma})$.

3. Studium przypadku

Chcąc zilustrować możliwości omawianego modelu wykorzystano zbiór danych z literatury przedmiotu [de Jong i Heller, 2008], dostępny na stronie internetowej [www 1]. Zawiera on dane dotyczące rocznych polis obowiązkowego ubezpieczenia odpowiedzialności cywilnej posiadaczy pojazdów mechanicznych. Zbiór liczy 67 856 polis, z czego w badanym roku szkody odnotowano na 4624 polisach, co stanowi około 6,8% całego zbioru. Oprócz danych na temat wysokości szkód, zbiór zawiera informacje dotyczące typu (zmienna TYPE – 12 możliwych typów) oraz wieku (VEH_AGE – 4 grupy (GR_1 – najmłodszy)) pojazdów, a także płci (SEX), wieku (AGE – 6 grup wiekowych (GR_1 - najmłodszy)) oraz miejsca zamieszkania właścicieli (AREA – 6 obszarów).

W pierwszym kroku analizy szacowano parametry całego modelu uwzględniające wszystkie pięć zmiennych objaśniających, korzystając z metody największej wiarygodności. Zgodnie z rozważaniami przedstawionymi w poprzednich punktach pracy estymowano oddzielnie dwa modele cząstkowe (wzór 14). Prawdopodobieństwa wystąpienia szkody modelowano korzystając z klasycznego modelu logitowego [Greene, 2003], natomiast wysokość szkody opisywana była uogólnionym modelem liniowym z rozkładem gamma (wzór 8). Następnie z poszczególnych modeli cząstkowych usunięte zostały zmienne, dla których żadna z możliwych kategorii nie miała istotnego (przyjmując poziom istotności równy 0,1) wpływu na wystąpienie i wysokość szkody. W efekcie w obu cząstkowych modelu wykorzystywano inne zestawy zmiennych objaśniających: w przypadku szacowania prawdopodobieństwa wystąpienia szkody istotnymi zmiennymi okazały się typ oraz wiek pojazdu, natomiast przy estymacji wysokości szkody statystycznie istotne okazały się wiek pojazdu, a także płeć, wiek oraz miejsce zamieszkania właściciela. Wartości oszacowanych parametrów zestawiono w tabeli 1.

Tabela 1. Oszacowania parametrów modelu

Prawdopodobieństwo szkody				Wysokość szkody				
Zm.	Wartość	$\hat{\gamma}$	p-wart.	Zm.	Wartość	$\hat{\beta}$	p-wart.	exp($\hat{\beta}$)
CONST		–2,29***	0,00	CONST		7,68***	0,00	2164,60
TYPE	CONVT	–21,74	1,00	VEH_AGE	GR_2	0,07	0,33	1,07
	COUPE	–22,54	0,99		GR_3	0,11	0,15	1,12
	HBACK	–21,92	0,97		GR_4	0,19**	0,01	1,21
	HDTOP	–22,25	0,99	SEX	M	0,16***	0,00	1,17
	MCARA	–1,88***	0,00	AREA	B	–0,03	0,67	0,97
	MIBUS	–5,76***	0,00		C	0,05	0,48	1,05
	PANVN	–22,25	0,99		D	0,00	0,97	1,00
	RDSTR	–21,34	1,00		E	0,14	0,16	1,15
	SEDAN	–8,53***	0,00		F	0,35***	0,00	1,42
	STNWG	–5,01***	0,00	AGE	GR_2	–0,20**	0,03	0,82
	TRUCK	–4,00***	0,00		GR_3	–0,29***	0,00	0,75
	UTE	–22,23	0,98		GR_4	–0,3***	0,00	0,74
VEH_AGE	GR_2	0,19	0,80		GR_5	–0,42***	0,00	0,66
	GR_3	–0,52	0,50		GR_6	–0,33***	0,00	0,72
	GR_4	1,81***	0,00					

NULL DEV.	864,37	NULL DEV.	7705,1
RESID. DEV.	593,11	RESID. DEV.	7538,1

Prawdopodobieństwo wystąpienia szkody jest największe dla karawanów (MCARA). Istotne jest również dla ciężarówek (TRUCK), samochodów o nadwoziu kombi (STNWG), minibusów (MIBUS) oraz w mniejszym stopniu, samochodów z nadwoziem sedan (SEDAN). Ponadto, prawdopodobieństwo to w znaczącym stopniu zwiększa duży wiek pojazdu (GR_4). Duża różnica pomiędzy zmiennością modelowanej zmiennej (NULL DEV.) a zmiennością reszt modelu (RESID. DEV.) wskazuje, że przyjęty zestaw zmiennych pozwala na wyjaśnienie znaczącej części różnic w obserwowanej szkodowości polis.

Także w odniesieniu do modelowania wysokości szkody w przypadku najstarszych pojazdów wysokość ta, w porównaniu do pojazdów najmłodszych, jest istotnie wyższa i wynosi około 20%. Wysokość szkody w przypadku mężczyzn jest o około 17% wyższa niż dla kobiet. Duża różnica szkodowości, sięgająca mniej więcej 40%, występuje pomiędzy regionem A oraz F. Najwyższą wartością szkód charakteryzują się kierowcy najmłodsi. Różnica w stosunku do pozostałych grup wiekowych jest w każdym przypadku istotna statystycznie i waha się od 20% do 35%. Należy dodać, że średnia wartość szkody dla polisy odnoszącej się do pojazdu z najniższej grupy wiekowej, którego posiadaczem jest kobieta zamieszkująca region A, z najmłodszej grupy wiekowej wynosi w rozpatrywanym zbiorze 2164 \$. Jakość dopasowania tej części modelu jest jednak znacznie niższa niż poprzedniej. Redukcja zróżnicowania jest bowiem niewielka.

Podsumowanie

W pracy zaproponowano model regresyjny do szacowania wartości pojedynczej szkody z uwzględnieniem polis bezszkodowych. Wykazano, iż model ten dekomponuje się na dwa modele GLM, co znacznie upraszcza jego praktyczne zastosowanie. Omawiany model wykorzystano do szacowania czynników oddziałujących na wysokość szkody portfela polis obowiązkowego ubezpieczenia OC posiadaczy pojazdów mechanicznych. Wśród zmiennych najsilniej wpływających na prawdopodobieństwo zgłoszenia szkody zidentyfikowano kilka typów pojazdów oraz jego wiek. Na wysokość szkody wpływ mają wiek pojazdu, płeć oraz wiek posiadacza, a także region zamieszkania. W omawianym przypadku model lepiej wyjaśnia prawdopodobieństwa zgłoszenia szkody niż jej wysokość.

Literatura

Antonio K. i Beirlant J. (2006), *Risk Classification in Nonlife Insurance* [w:] E.L. Melnick, B.S. Everitt (eds.), *Encyclopedia of Quantitative Risk Analysis and Assessment*, Vol. 1, Wiley, Chichester.

Antonio K. i Valdez E.A. (2012), *Statistical Concepts of a Priori and a Posteriori Risk Classification in Insurance*, „AStA Advances in Statistical Analysis", Vol. 96(2), s. 187-224.

Dirac P.A.M. (1982), *The Principles of Quantum Mechanics*, International Series of Monographs on Physics, Oxford University Press, Oxford.

Greene W.H. (2003), *Econometric Analysis*, 5/e, Pearson Education India, Delhi.

Jong P. de i Heller G.Z. (2008), *Generalized Linear Models for Insurance Data*, Cambridge University Press, Cambridge.

MacCullagh P. i Nelder J.A. (1989), *Generalized Linear Models, CRC Monographs on Statistics & Applied Probability*, Springer Verlag, New York,

Nelder J.A. i Wedderburn R.W. (1972), *Generalized Linear Models*, „Journal of the Royal Statistical Society. Series A (General)", s. 370-384.

Ohlsson, E. i Johansson B. (2010), *Non-Life Insurance Pricing With Generalized Linear Models*, Springer Verlag, New York.

Ostasiewicz, W. i Dębicka J. (2004), *Składki i ryzyko ubezpieczeniowe: modelowanie stochastyczne*, Wydawnictwo Akademii Ekonomicznej, Wrocław.

Pobłocka A. (2011), *Rezerwa IBNR w ubezpieczeniach majątkowych – praktyczne metody jej szacowania*, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław.

R_Core_Team (2013), R: *A language and Environment for Statistical Computing, R Foundation for Statistical Computing*, Vienna, Austria.

Rao J.N. (2003), *Small Area Estimation*, Wiley, Hoboken.

Ronka-Chmielowiec W. i Heilpern S. (2000), *Zarządzanie ryzykiem w ubezpieczeniach*, Wydawnictwo Akademii Ekonomicznej, Wrocław.

Szkutnik W. (2003), *Wybrane modele w zarządzaniu ryzykiem ubezpieczeniowym w ujęciu probabilistycznym*, Wydawnictwo Akademii Ekonomicznej, Katowice.

Taylor G.C. (2000), *Loss Reserving: An Actuarial Perspective*, Kluwer Academic, Dordrecht.

Wanat S. (2008), *Wybrane metody uwzględniania zależności w modelowaniu ryzyka ubezpieczyciela* [w:] T. Trzaskalik (red.), *Modelowanie preferencji a ryzyko'07*, Wydawnictwo Akademii Ekonomicznej, Katowice, s. 377-391.

Wolny A. (2003), *Analiza wrażliwosci wyniku technicznego zakladu ubezpieczen na zmiane stanu rezerwy na nie wyplacone odszkodowania i swiadczenia*, Wydawnictwo Akademii Ekonomicznej, Wroclaw.

Wolny-Dominiak A. (2009), *Szacowanie poziomu zmiennych taryfikacyjnych w ubezpieczeniach typu non-life*, Studia Ekonomiczne, nr 57, Wydawnictwo Akademii Ekonomicznej, Katowice, s. 201-209.

Żądło T. (2008), *Elementy statystyki małych obszarów z programem R*, Wydawnictwo Akademii Ekonomicznej, Katowice.

[www 1] http://www.businessandeconomics.mq.edu.au/our_departments/Applied_Finance_and_Actuarial_Studies/research/books/GLMsforInsuranceData/data_sets (dostęp: 10.12.2013).

REGRESSION MODEL OF A SINGLE CLAIM VALUE WITH NO-CLAIMS POLICIES

Summary: In the paper, we introduce probability distribution which augments any distribution from the exponential family with zero value (ZA-distribution). The new distribution is then employed in a regression model applied for forecasting value of a single claim in a large insurance portfolio (motor or property).

Keywords: ZA-distribution, GLM, insurance portfolio.