

ZALEŻNOŚĆ STOCHASTYCZNA W AKTUARIALNYCH MODELACH TARYFIKACJI *A POSTERIOR*

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY
Nr 15(21)

Kamil Gala

Ubezpieczeniowy Fundusz Gwarancyjny

Wojciech Bijak

Szkoła Główna Handlowa w Warszawie,

Ubezpieczeniowy Fundusz Gwarancyjny

e-mails: kgala@ufg.pl; wobi@sgh.waw.pl

ISSN 1644-6739
e-ISSN 2449-9765

DOI: 10.15611/sps.2017.15.03

JEL Classification: G22 Insurance; Insurance Companies; Actuarial Studies

Streszczenie: System *bonus-malus* jest jedną z podstawowych metod taryfikacji *a posteriori* stosowanych w ubezpieczeniach komunikacyjnych. W tej metodzie klasa taryfowa ubezpieczonego zależy od jego klasy taryfowej w poprzednim roku oraz liczby szkód zgłoszonych w ciągu roku. Należy jednak zwrócić uwagę, że w praktyce ubezpieczeń komunikacyjnych proces powstania szkód jest w istocie wielowymiarowy. W niniejszej pracy został omówiony przypadek, w którym szkody powodujące odpowiedzialność ubezpieczyciela z tytułu zawartej umowy ubezpieczenia OC p.p.m. zostały podzielone na szkody rzeczowe oraz szkody osobowe. Zbadano zależność stochastyczną między zmiennymi losowymi reprezentującymi szkody obu rodzajów, a następnie opisano system *bonus-malus*, w którym kara za spowodowane szkody zależy od jej rodzaju. Wyniki analizy wskazują, że w przypadku występowania korelacji między liczbą szkód rzeczowych oraz liczbą szkód osobowych zaproponowany system pozwala na lepszą ocenę ryzyka ubezpieczeniowego.

Słowa kluczowe: ubezpieczenia komunikacyjne, zależność stochastyczna, system *bonus-malus*, taryfikacja *a posteriori*, kopula.

1. Wstęp

Ubezpieczenia komunikacyjne ze względu na swój masowy charakter są obszarem, w którym aktuarialne metody kalkulacji składki (zarówno *a priori*, jak i *a posteriori*) znajdują najszersze zastosowanie. W metodach tych najczęściej bierze się pod uwagę wybraną statystykę szkodową (np. liczbę szkód, częstość szkód, wartość wypłat) i na podstawie danych historycznych buduje się model probabilistyczny służący do oceny ryzyka ubezpieczeniowego na podstawie różnych czynników opisujących ubezpieczonego, jego pojazd lub jego historię szko-

dową. Należy jednak zwrócić uwagę na to, że nie zawsze to podejście jest wystarczające.

W ubezpieczeniach komunikacyjnych można wskazać wiele sytuacji, w których proces powstania szkód jest w istocie wielowymiarowy. Wtedy przedmiotem zainteresowania jest nie jedna zmienna losowa (której realizacją jest analizowana statystyka szkodowa), a wiele zmiennych losowych, wzajemnie ze sobą powiązanych. Wśród typowych przypadków należy wymienić:

- liczbę szkód różnych rodzajów obciążających tę samą umowę ubezpieczenia [Vernic 1997],
- liczbę szkód z różnych umów tego samego ubezpieczonego w tym samym okresie [Shi, Valdez 2014],
- liczbę szkód spowodowanych przez tego samego kierowcę w kolejnych latach [Vernic 2000; Boucher, Denuit, Guillén 2008],
- liczbę i wysokość szkód [Pinquet 1997; Sawicka 2013],
- wielu użytkowników pojazdu,
- wiele pojazdów dla jednego ubezpieczonego (floty) [Angers i in. 2006].

W związku z powyższym pojawia się potrzeba stosowania metod analizy pozwalających na weryfikację występowania współzależności między różnymi czynnikami ryzyka oraz uwzględnienie jej w procesie taryfikacji. We wszystkich wskazanych wyżej przypadkach można bowiem argumentować, że przyjęcie założenia o niezależności poszczególnych składowych modelu może nie być zasadne. Przykładowo klasyczny model wiarygodności Poissona (*Poisson credibility model*) zakłada, że dla ustalonego ubezpieczonego zmienne losowe reprezentujące liczbę szkód w kolejnych latach są niezależne. Każdy ubezpieczony cechuje się przy tym indywidualnym parametrem ryzyka, traktowanym jako realizacja pewnej zmiennej losowej, odzwierciedlającym nieobserwowalne charakterystyki ubezpieczonego (np. znajomość przepisów w ruchu drogowego). Na poziomie portfela prowadzi to do pojawienia się korelacji między liczbą szkód w kolejnych latach (zob. np. [Denuit i in. 2007]). Z kolei w przypadku ubezpieczeń wielu pojazdów zależność może wynikać z występowania wspólnego czynnika ryzyka w postaci podejścia przedsiębiorstwa zarządzającego flotą do eksploatacji i dbania o stan techniczny swoich pojazdów [Angers i in. 2006].

Niniejsza praca jest poświęcona zagadnieniu modelowania odpowiedzialności zakładu ubezpieczeń z tytułu zawartych umów ubezpieczenia odpowiedzialności cywilnej posiadaczy pojazdów mechanicznych (dalej: OC p.p.m.). Odpowiedzialność zakładu ubezpieczeń mo-

że mieć różne źródła, związane m.in. z rodzajem szkody spowodowanej przez ubezpieczonego kierowcę. Zagadnienia modelowania przyczyn odpowiedzialności zakładu ubezpieczeń było wielokrotnie, w różnych kontekstach, analizowane w literaturze aktuarialnej. W pracy [Denuit i in. 2007] rozważany jest podział szkód na szkody osobowe oraz rzeczowe, a zmienne losowe reprezentujące liczbę szkód analizowane są za pomocą modelu bayesowskiego, w którym ukryte czynniki ryzyka dla obu rodzajów szkody są ze sobą skorelowane. Model ten oraz metody aktuarialnej teorii wiarygodności (*credibility theory*) zostały wykorzystane przez autorów cytowanej pracy do wyznaczenia parametrów systemu *bonus-malus*, w którym przejścia między klasami zależą od różnych rodzajów zdarzeń (*multi-event system*). W pracy [Shi, Valdez 2014] rozważane są ubezpieczenia z kompleksowym zakresem ochrony (*comprehensive coverage*), w przybliżeniu odpowiadające zakresem ochrony polskim ubezpieczeniom OC p.p.m. i *autocasco*, w których zostały wyróżnione trzy rodzaje szkód: szkody osobowe oraz szkody rzeczowe wyrządzone osobom trzecim, a także uszkodzenia własnego pojazdu. Do ich modelowania wykorzystano trójwymiarowy rozkład ujemny dwumianowy, w którym rozkłady brzegowe modelowane są z wykorzystaniem regresji ujemnej dwumianowej, a struktura zależności określona jest za pomocą kopuli. Z kolei w pracy [Vernic 1997] do modelowania łącznego rozkładu liczby szkód osobowych i liczby szkód rzeczowych w ubezpieczeniu OC p.p.m. wykorzystano dwuwymiarowy uogólniony rozkład Poissona.

W kolejnych punktach pracy omówiono zagadnienie modelowania liczby szkód w ubezpieczeniach OC p.p.m. oraz system *bonus-malus* uwzględniający rodzaj powodowanych przez ubezpieczonego szkód. Opisano różne podejścia do modelowania rozkładów brzegowych (np. mieszane modele Poissona) oraz rozkładu łącznego (oparte na kopulach). Rozważania teoretyczne zostały uzupełnione analizą empiryczną, gdzie poruszono kwestię dopasowania rozkładów brzegowych oraz rozkładu łącznego liczby szkód osobowych i rzeczowych. Pracę zamykają podsumowanie i wnioski.

2. Modelowanie liczby szkód w ubezpieczeniach OC p.p.m.

2.1. Klasyfikacja szkód z tytułu OC p.p.m.

Ubezpieczenie odpowiedzialności cywilnej posiadaczy pojazdów mechanicznych (OC p.p.m.) obejmuje szkody wyrządzone osobom trzecim w związku z ruchem tych pojazdów. Tak określony zakres ochro-

ny ubezpieczeniowej przekłada się na dużą różnorodność szkód powodujących odpowiedzialność zakładu ubezpieczeń z tytułu zawartej umowy OC p.p.m. Zakres ochrony obejmuje zarówno szkody w pojeździe (np. urwane lusterko), jak i szkody w mieniu poza pojazdem (np. uszkodzone ogrodzenie) oraz szkody osobowe i ich następstwa (np. koszty leczenia). Szkody te mogą być klasyfikowane na różne sposoby, również w sposób hierarchiczny (np. podział szkód osobowych na związane ze zgonem oraz związane z obrażeniami ciała).

W niniejszej pracy została przyjęta najprostsza klasyfikacja, w której szkody dzieli się na szkody rzeczowe (obejmujące szkody w pojeździe i mieniu poza pojazdem) oraz szkody osobowe. W kolejnym punkcie zostanie przedstawiony model probabilistyczny, umożliwiający analizę tak zdefiniowanego dwuwymiarowego procesu powstawania szkód.

2.2. Probabilistyczny model powstawania szkód

Rozważmy ogólny proces powstawania szkód, w którym każda szkoda jest ustalonego typu, przy czym typy ponumerowane są od 1 do m . Liczba szkód spowodowanych przez i -tego ubezpieczonego w okresie $(t - 1, t]$ może być przedstawiona jako wektor losowy $\mathbf{N}_{it} = (N_{it}^1, \dots, N_{it}^m)$, gdzie N_{it}^j oznacza liczbę szkód j -tego typu. Tak zdefiniowany proces jest więc wielowymiarowy, a jego dalsza analiza wymaga przyjęcia założeń dotyczących struktury probabilistycznej wektora \mathbf{N}_{it} , szczególnie brzegowych rozkładów prawdopodobieństwa oraz rozkładu łącznego. Warto zauważyć, że definiując $N_{it} = \sum_{j=1}^m N_{it}^j$, otrzymuje się klasyczny, jednowymiarowy proces powstawania szkód.

W dalszej części pracy będzie rozważany przypadek dwuwymiarowy ($m = 2$), w którym szkody można podzielić na szkody osobowe ($j = 1$) oraz szkody rzeczowe ($j = 2$). Dla większej czytelności przyjęto dalej oznaczenie literowe „ O ” oraz „ R ”, odpowiednio ($j \in \{O, R\}$). W takim przypadku $\mathbf{N}_{it} = (N_{it}^O, N_{it}^R)$, gdzie N_{it}^O oraz N_{it}^R oznaczają odpowiednio liczbę szkód osobowych oraz liczbę szkód rzeczowych obciążających i -tą umowę ubezpieczenia w okresie $(t - 1, t]$. Warto podkreślić, że alternatywnym kryterium podziału na typy może być np. rodzaj umowy (OC p.p.m., AC) lub wysokość szkody (np. poniżej oraz powyżej ustalonego progu kwotowego). Przedstawiony model można więc uogólnić na większą liczbę wymiarów.

W dalszej części tego punktu przedstawiono założenia, które mogą zostać przyjęte do modelowania rozkładu prawdopodobieństwa wek-

tora (N_{it}^O, N_{it}^R) na potrzeby analizy wielowymiarowego systemu *bonus-malus*. Założenia te zostały zweryfikowane empirycznie na podstawie danych pochodzących z bazy danych Ośrodka Informacji Ubezpieczeniowego Funduszu Gwarancyjnego (dalej: OI UFG)¹. Wykorzystane dane obejmowały informację o umowach OC p.p.m. zawartych w 2013 r. ze wskazaniem liczby szkód rzeczowych i osobowych obciążających te umowy.

Rozkłady brzegowe

Do modelowania rozkładów brzegowych wykorzystywane są rozkłady dyskretne zdefiniowane na zbiorze $\{0, 1, 2, \dots\}$. W analizie zostały uwzględnione następujące rozkłady o wskazanych poniżej funkcjach rozkładu prawdopodobieństwa:

1. Rozkład Poissona:

$$\mathbb{P}(N = k) = \exp(-\lambda) \frac{\lambda^k}{k!}, \lambda > 0.$$

Dalej będzie stosowane oznaczenie $N \sim \text{Poiss}(\lambda)$.

2. Rozkład ujemny dwumianowy (mieszany rozkład Poissona z rozkładem gamma jako rozkładem mieszającym, dalej: NB):

$$\mathbb{P}(N = k) = \int_0^\infty \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} \cdot \frac{a^a}{\Gamma(a)} \theta^{a-1} \exp(-a\theta) d\theta,$$

gdzie $\lambda > 0$ oraz $a > 0$.

3. Mieszany rozkład Poissona z rozkładem logarytmiczno-normalnym jako rozkładem mieszającym (dalej: P-LN):

$$\mathbb{P}(N = k) = \int_0^\infty \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} \cdot \frac{1}{\theta s \sqrt{2\pi}} \exp\left(-\frac{\left(\ln(\theta) + \frac{s^2}{2}\right)^2}{2s^2}\right) d\theta,$$

gdzie $\lambda > 0, s > 0$.

¹ Zakres danych gromadzonych w bazie OI UFG określony jest w art. 102 Ustawy z dnia 22 maja 2013 r. o ubezpieczeniach obowiązkowych, Ubezpieczeniowym Funduszu Gwarancyjnym i Polskim Biurze Ubezpieczycieli Komunikacyjnych (Dz.U. 2013 poz. 392, t.j.).

4. Mieszany rozkład Poissona z rozkładem odwrotnym gaussowskim jako rozkładem mieszającym (dalej: P-IG):

$$\mathbb{P}(N = k) = \int_0^{\infty} \exp(-\lambda\theta) \frac{(\lambda\theta)^k}{k!} \cdot \frac{1}{\sqrt{2\pi\tau\theta^3}} \exp\left(-\frac{(\theta-1)^2}{2\tau\theta}\right) d\theta,$$

gdzie $\lambda > 0$ oraz $\tau > 0$.

5. Rozkład Poissona z podwyższoną liczbą zer (dalej: ZIP):

$$\mathbb{P}(N = k) = p \cdot \mathbf{1}(k = 0) + (1 - p) \cdot \exp(-\lambda) \frac{\lambda^k}{k!},$$

gdzie $\lambda > 0$ oraz $p \in (0,1)$, a $\mathbf{1}(\cdot)$ jest funkcją indykatorową, tzn. $\mathbf{1}(k = 0) = 1$, jeśli $k = 0$, i $\mathbf{1}(k = 0) = 0$ w przeciwnym przypadku.

6. Rozkład Neymana typ A (dalej: N-A):

$$N =_p \sum_{i=1}^M K_i,$$

gdzie $=_p$ oznacza równość rozkładów, $M \sim Poiss(\mu)$, $K_i \sim Poiss(\lambda)$, $\mu > 0$, $\lambda > 0$, a zmienne M, K_1, K_2, \dots, K_M są niezależne. Ponadto jeśli $M = 0$, to $N = 0$. Rozkład Neymana jest więc rozkładem złożonym, w którym rozkład pierwotny oraz rozkłady wtórne są rozkładami Poissona.

Należy zwrócić uwagę, że rozkłady 2-4 to mieszane rozkłady Poissona, w których liczba szkód ma warunkowy (przy ustalonej wartości zmiennej losowej $\Theta = \theta$) rozkład Poissona z parametrem $\lambda\theta$, natomiast Θ jest nieujemną zmienną losową, taką że $E\Theta = 1$. Rozkłady te często są wykorzystywane do modelowania populacji niejednorodnej, a także do wprowadzania zależności między uszkodzeniami w różnych okresach [Denuit i in. 2007].

Powyższe rozkłady zostały dopasowane do próbki rzeczywistych danych pochodzących z bazy OI UFG. Próbka liczyła 1 mln obserwacji, a rozkład empiryczny liczby szkód przedstawiony został w tab. 1.

Parametry rozkładów zostały oszacowane za pomocą metody największej wiarygodności, osobno dla szkód rzeczowych oraz szkód osobowych. Następnie przeprowadzono standardowy test zgodności chi-kwadrat, grupując w razie potrzeby komórki, w których teoretyczna liczba szkód była mniejsza niż 5. Uzyskano wartości statystyki testowej przedstawione w tab. 2.

Tabela 1. Empiryczny rozkład liczby szkód rzeczowych i liczby szkód osobowych

Liczba szkód		Osobowe						Suma
		0	1	2	3	4	5	
Rzeczowe	0	970 458	496	62	15	6	3	971 040
	1	25 289	1 018	207	40	12	7	26 573
	2	1 753	181	60	16	5	5	2 020
	3	207	34	11	6	2	1	261
	4	67	4	2	4	0	2	79
	5	22	3	0	2	0	0	27
Suma		997 796	1 736	342	83	25	18	1 000 000

Źródło: opracowanie własne.

Tabela 2. Statystyka testu chi-kwadrat dla dopasowanych rozkładów brzegowych

Rozkład	Statystyka testu chi-kwadrat*	
	szkody rzeczowe	szkody osobowe
Poisson	30 252	149
NB	201	12
P-IG	61	40
P-LN	39	683
ZIP	1 336	106
N-A	740	112

* Pogrubioną czcionką zaznaczono najmniejszą wartość dla danego rodzaju szkody. Wszystkie wartości zaokrąglono do najbliższej liczby całkowitej.

Źródło: opracowanie własne.

Uzyskane wyniki wskazują, że najlepiej dopasowane rozkłady to rozkład Poissona mieszany rozkładem logarytmiczno-normalnym dla szkód rzeczowych oraz rozkład Poissona mieszany rozkładem gamma (rozkład ujemny dwumianowy) dla szkód osobowych. Potwierdza to zasadność stosowania modeli, w których wariancja liczby szkód przewyższa jej wartość oczekiwaną (w odróżnieniu od rozkładu Poissona, w którym te wielkości są sobie równe). Należy mieć jednak na uwadze, że w przypadku szkód osobowych nie ma podstaw do odrzucenia hipotezy zerowej mówiącej o zgodności rozkładu teoretycznego z rozkładem empirycznym dopiero przy poziomie istotności 0,001, natomiast dla szkód rzeczowych poziom ten jest dużo mniejszy. Na potrzeby dalszej analizy można jednak uznać, że modele z klasy mieszanych rozkładów Poissona są wystarczająco dobrze dopasowane do danych. W tabeli 2 podano parametry najlepiej dopasowanych rozkładów.

Tabela 3. Parametry najlepiej dopasowanych rozkładów

Rodzaj szkody	Parametr	Wartość*
Rzeczowe	λ	0,03
	s	1,40
Osobowe	λ	0,003
	a	226

* Wartości podane w zaokrągleniu.

Źródło: opracowanie własne.

Uzyskane wyniki wskazują, że częstość szkód rzeczowych jest ok. 10 razy wyższa niż dla szkód osobowych, natomiast liczba szkód osobowych cechuje się dużo większą zmiennością. Oznacza to, że występowanie szkód osobowych jest zjawiskiem stosunkowo rzadkim i zmiennym, co przekłada się na trudność jego modelowania.

Rozkład łączny

Analiza rozkładu łącznego liczby szkód rzeczowych oraz liczby szkód osobowych wymaga przyjęcia założeń dotyczących struktury zależności między tymi zmiennymi losowymi. Z danych przedstawionych w tab. 1 wynika, że współczynnik korelacji Pearsona wynosi ok. 0,2, współczynnik V -Cramera przyjmuje wartość 0,1, a hipotezę zerową w teście niezależności chi-kwadrat należy odrzucić przy każdym rozsądnym poziomie istotności. Oznacza to, że rozkład łączny musi uwzględniać korelację między rozkładami brzegowymi.

Mając na uwadze wyniki analizy empirycznej dla rozkładów brzegowych, zdecydowano o zastosowaniu modelu będącego wielowymiarowym rozszerzeniem standardowego Poissonowskiego modelu wiarygodności. Założenia tego modelu są następujące:

1. i -ta umowa ubezpieczenia ($i = 1, \dots, N$) cechuje się dwuwymiarowym parametrem ryzyka (θ_i^O, θ_i^R) będącym realizacją dwuwymiarowej zmiennej losowej (Θ_i^O, Θ_i^R) .

2. Przy ustalonej wartości $(\Theta_i^O, \Theta_i^R) = (\theta_i^O, \theta_i^R)$ zmienne losowe $(N_{i1}^R, N_{i1}^O), \dots, (N_{it}^R, N_{it}^O)$ są niezależne i mają ten sam rozkład prawdopodobieństwa; zmienne N_{it}^j warunkowo mają rozkład Poissona z parametrem $\lambda^j \theta_i^j$ ($j \in \{O, R\}$).

3. Zmienne $(\Theta_1^O, \Theta_1^R), \dots, (\Theta_N^O, \Theta_N^R)$ są niezależne i mają ten sam rozkład prawdopodobieństwa.

4. Dla każdego i zmienna losowa (Θ_i^O, Θ_i^R) ma dwuwymiarowy rozkład prawdopodobieństwa dany dystrybuantą H . Rozkład ten jest

absolutnie ciągły i na mocy twierdzenia Sklara [Nelsen 2006] można dystrybuantę tego rozkładu przedstawić jako:

$$H(\theta^O, \theta^R) = C(F_O(\theta^O), F_R(\theta^R)),$$

gdzie F_j jest dystrybuantą rozkładu zmiennej losowej Θ_i^j , a C jest kopulą dwuwymiarową, tzn. dystrybuantą rozkładu dwuwymiarowego, którego rozkłady brzegowe są jednostajne na odcinku $[0,1]$. Jeśli kopula C oraz rozkłady zmiennych Θ_i^j są absolutnie ciągłe, to gęstość wektora (Θ_i^O, Θ_i^R) można wyrazić wzorem:

$$h(\theta^O, \theta^R) = c(F_O(\theta^O), F_R(\theta^R)) \cdot f_O(\theta^O) \cdot f_R(\theta^R),$$

gdzie $c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v)$, a f_j oznacza gęstość zmiennej Θ_i^j dla $j \in \{O, R\}$.

Przy powyższych założeniach łączny rozkład wektora (N_{it}^O, N_{it}^R) można przedstawić wzorem

$$\begin{aligned} p(k, l) &= \mathbb{P}(N_{it}^O = k, N_{it}^R = l) = \\ &= \int_{(0, +\infty) \times (0, +\infty)} \exp(-\lambda^O \theta^O) \frac{(\lambda^O \theta^O)^k}{k!} \exp(-\lambda^R \theta^R) \frac{(\lambda^R \theta^R)^l}{l!} h(\theta^O, \theta^R) d\theta^O d\theta^R. \end{aligned} \quad (1)$$

Powyższy wzór może zostać wykorzystany do estymacji parametrów rozkładu łącznego metodą największej wiarygodności (MNW), co może być trudne ze względu skomplikowaną postać wzoru (1) oraz dużą liczbę parametrów do oszacowania (λ^O, λ^R , parametry rozkładów brzegowych oraz parametry kopuli C). W związku z tym zastosowano dwuetapową metodę największej wiarygodności, w której:

- w pierwszym etapie parametry rozkładów brzegowych są szacowane za pomocą MNW,
- w drugim etapie buduje się funkcję wiarygodności, korzystając ze wzoru (1), przy czym wartości parametrów rozkładów brzegowych uzyskane w pierwszym etapie traktuje się jako dane – funkcja wiarygodności jest więc funkcją wyłącznie parametrów kopuli.

Powyższa procedura zbliżona jest do tej opisaney w pracy [McNeil, Frey, Embrechts 2005] i pozwala istotnie zmniejszyć złożoność problemu.

W niniejszej pracy rozważono trzy kopule często spotykane w literaturze przedmiotu:

- kopula Clayтона:

$$C_\alpha(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{\frac{1}{\alpha}}, (u, v) \in [0,1]^2, \alpha > 0,$$

- kopula Gumbela:

$$C_\alpha(u, v) = \exp\left(-\left(-\ln(u)\right)^\alpha + \left(-\ln(v)\right)^\alpha\right)^{\frac{1}{\alpha}}, (u, v) \in [0,1]^2, \alpha > 1,$$

- kopula Gaussa:

$$C_\alpha(u, v) = \Phi_\alpha(\Phi^{-1}(u), \Phi^{-1}(v)), (u, v) \in [0,1]^2, \alpha \in (-1,1),$$

gdzie Φ_α jest dystrybuantą standardowego dwuwymiarowego rozkładu normalnego ze współczynnikiem korelacji α , natomiast Φ^{-1} to funkcje odwrotne do dystrybuanty standardowego jednowymiarowego rozkładu normalnego (tzn. ze średnią równą 0 i wariancją równą 1).

Powyższe kopule mają jeden parametr, co oznacza, że w drugim etapie procedury estymacji maksymalizowana jest funkcja jednej zmiennej, α .

W tabeli 4 przedstawiono wyniki estymacji wraz ze statystyką testu dopasowania chi-kwadrat dla poszczególnych kopuli. Oprócz oceny parametru α podano również wartość współczynnika τ Kendalla, który dla rozważanych kopuli jest rosnącą funkcją parametru α . Najlepsze dopasowanie do danych zostało uzyskane dla kopuli Gumbela, która została wyróżniona w tabeli pogrubieniem.

Tabela 4. Wyniki estymacji parametru kopuli

Kopula	α	τ	Chi-kwadrat
Clayton	64,7	0,97	2421
Gumbel	20,0	0,95	1097
Gauss	0,97	0,84	1098

Zródło: opracowanie własne.

Uzyskane wyniki wskazują, że niezależnie od wyboru kopuli najlepsze dopasowanie otrzymuje się przy parametrze wskazującym na silną korelację między ukrytymi parametrami ryzyka, (Θ^O, Θ^R) , nie jest to jednak zależność funkcyjna. Oznacza to, że modelowanie łącznego rozkładu wektora (N_{it}^O, N_{it}^R) za pomocą modelu ze wspólnym parametrem ryzyka dla obu rodzajów szkód nie jest wystarczające.

Należy zwrócić przy tym uwagę, że również model z dwuwymiarowym parametrem ryzyka nie jest dobrze dopasowany do danych. Duże wartości statystyki chi-kwadrat wynikają przede wszystkim z dużej różnicy między obserwowanym rozkładem warunkowym szkód osobowych dla umów z jedną szkodą rzeczową a odpowiadającym mu rozkładem teoretycznym. Sugeruje to, że proces powstawania szkód osobowych różni się w zależności od tego, czy jest to „czysta” szkoda

osobowa (np. wypadek z udziałem pieszego), czy szkoda osobowa towarzysząca szkodom rzeczowym (np. w wyniku zderzenia dwóch pojazdów). Rozwiązaniem tego problemu może być zastosowanie bardziej złożonych rozkładów łącznych, np. rozkładów mieszanych. Zagadnienie to może być przedmiotem dalszych badań, natomiast przedstawiony model stanowi pewien kompromis między złożonością modelu a jego dopasowaniem do rzeczywistych danych.

3. System *bonus-malus* uwzględniający rodzaj szkody

3.1. Opis systemu *bonus-malus*

Wyniki analizy przedstawione w punkcie 2 wskazują, że występuje korelacja między zmiennymi losowymi reprezentującymi liczbę szkód rzeczowych oraz liczbę szkód osobowych. Oznacza to, że można podjąć próbę odpowiedzi na pytanie, czy uwzględnienie zależności między analizowanymi rodzajami szkód zwiększy efektywność systemu *bonus-malus* (dalej: SBM). Najpierw zostanie rozważony ogólny model, uwzględniający m rodzajów szkód, natomiast szczegółowa analiza symulacyjna zostanie przeprowadzona dla przypadku dwuwymiarowego.

W niniejszej pracy rozważony zostanie wielowymiarowy SBM, rozumiany jako indywidualny system oceny ryzyka, na który składają się następujące elementy:

1) $s + 1$ klas taryfowych tworzących zbiór $\mathcal{L} = \{0, \dots, s\}$, przy czym zgłoszenie szkody skutkuje przeniesieniem ubezpieczonego do klasy o wyższym numerze;

2) wektor stawek składki $\mathbf{r} = (r_0, \dots, r_s)$, nazywany taryfą *bonus-malus*;

3) klasę startową, l_0 , najczęściej taką, że $r_{l_0} = 1$;

4) reguły przejścia, które określają obecną klasę taryfową kierowcy na podstawie jego klasy taryfowej w poprzednim okresie oraz liczby i rodzaju zgłoszonych w tym czasie szkód. W przypadku wielowymiarowym reguły przejścia można reprezentować za pomocą macierzy

$$\mathbf{T}(k_1, k_2, \dots, k_m) = [t_{ij}(k_1, k_2, \dots, k_m)]_{(s+1) \times (s+1)},$$

gdzie $t_{ij}(k_1, k_2, \dots, k_m) = 1$, jeśli po zgłoszeniu k_j ($k_j = 0, 1, 2, \dots$) szkód typu j ubezpieczony przechodzi z klasy i do klasy j oraz 0 w przeciwnym przypadku. W konsekwencji macierze $\mathbf{T}(k_1, k_2, \dots, k_m)$

dla dowolnego wektora $(k_1, k_2, \dots, k_m) \in \{0, 1, \dots\}^m$ są macierzami zero-jedynkowymi o dokładnie jednym niezerowym elemencie w każdym wierszu.

Klasyczny, jednowymiarowy SBM można otrzymać, definiując reguły przejścia w taki sposób, że $\mathbf{T}(k_1, k_2, \dots, k_m) = \mathbf{T}(k'_1, k'_2, \dots, k'_m)$, o ile $\sum_j k_j = \sum_j k'_j$.

Kluczowym elementem systemu *bonus-malus* są przyjęte reguły przejścia, które określają jego własności. W odniesieniu do jednowymiarowego SBM można spotkać w literaturze postulaty, które definiują SBM możliwy do zastosowania w praktyce. Na przykład w pracy [Podgórska, Kryszeń, Niemiec 2008] zdefiniowano SBM sprawiedliwy według składki oraz SBM sprawiedliwy według klas. Wymagania względem sprawiedliwych SBM polegają na tym, aby kierowcy należący do klas taryfowych związanych z większym ryzykiem nie płacili mniejszych składek od kierowców z klas o mniejszym ryzyku oraz aby kierowcy uznawani za bardziej ryzykownych (o większej liczbie szkód) nie byli zaliczani do klas o niższej składce niż kierowcy mniej szkodowi. W kontekście wielowymiarowego SBM zagadnienie to wydaje się jeszcze bardziej złożone, ponieważ w przestrzeniach wielowymiarowych najczęściej nie istnieją naturalne liniowe porządki.

W niniejszej pracy przyjęto, że spowodowanie szkody typu j niezależnie od zajmowanej klasy wiąże się z przesunięciem kierowcy o c_j klas w górę (tj. do klasy o wyższym numerze), natomiast brak szkód wiąże się z przesunięciem o b klas w dół (tj. do klasy o niższym numerze). Łączna zmiana klasy z jest więc dana wzorem

$$z(k_1, \dots, k_m) = -b \cdot \mathbf{1}\left(\sum_{j=1}^m k_j = 0\right) + \left(\sum_{j=1}^m c_j k_j\right) \cdot \mathbf{1}\left(\sum_{j=1}^m k_j > 0\right),$$

gdzie: $\mathbf{1}(\cdot)$ oznacza funkcję charakterystyczną zbioru.

Ubezpieczony znajdujący się w klasie l , który spowodował k_j szkód rodzaju j ($j = 1, \dots, m$), znajdzie się w kolejnym okresie w klasie

$$l' = \min(\max(0, l + z(k_1, \dots, k_m)), s + 1).$$

Tak skonstruowany SBM oznaczymy $-b/c_1/c_2/\dots/c_m$. Można ten system uogólnić, stosując jako z funkcję inną niż suma, np. maksimum lub minimum. W ogólnym przypadku taki system oznaczymy $-b/c_1/c_2/\dots/c_m//z(\mathbf{k}, \mathbf{b}, \mathbf{c})$, gdzie $z(\mathbf{k}, \mathbf{b}, \mathbf{c})$ oznacza pewną funkcję zmiennych $k_1, \dots, k_m, b, c_1, \dots, c_m$.

Wyznaczanie mnożników składki

Istotnym elementem w SBM jest również wektor mnożników składki \mathbf{r} , który może zostać ustalony różnymi metodami, z uwzględnieniem różnych kryteriów optymalności. W niniejszej pracy wykorzystane zostanie metoda Norberga [1976], w której mnożniki składki wyznaczone są tak, aby minimalizować średniokwadratowy błąd oceny ubezpieczonych po osiągnięciu przez SBM stanu stacjonarnego. W tym celu zdefiniowana zostanie zmienna losowa S reprezentująca całkowitą wartość szkód spowodowanych przez ubezpieczonego w ciągu roku oraz zmienne S^j określające wysokość szkód typu j . Zachodzi więc $S = \sum_j S^j$. Przyjęte zostanie założenie, że zmienne S^j mają rozkłady złożone i zachodzi:

$$S^j = \sum_{i=1}^{N^j} X_i^j,$$

gdzie N^j jest liczbą szkód typu j powodowanych w ciągu roku, natomiast X_i^j jest wartością i -tej szkody typu j o wartości oczekiwanej μ^j , przy czym liczba i wysokość poszczególnych szkód są niezależnymi zmiennymi losowymi. Dodatkowo założymy, że N^j mają mieszane rozkłady Poissona z parametrem $\lambda^j \theta^j$, gdzie θ^j jest realizacją zmiennej losowej Θ^j .

Wprowadźmy teraz zmienną losową L , określającą numer klasy SBM, w której znajduje się ubezpieczony po osiągnięciu przez system stanu stacjonarnego. Z własności warunkowej wartości oczekiwanej wynika, że wyrażenie $\mathbb{E}(S - f(L))^2$ jest w klasie funkcji całkowalnych z kwadratem minimalizowane przez $f(L) = \mathbb{E}(S|L)$. Wynika stąd, że optymalne mnożniki składki dane są wzorem

$$r_l = \frac{\mathbb{E}(S|L = l)}{\mathbb{E}(S)}.$$

Można pokazać, że dla dowolnego $m \geq 2$ zachodzi

$$r_l = \sum_{j=1}^m w^j \cdot r_l^j,$$

gdzie $w^j = \frac{\lambda^j \mu^j}{\sum_j \lambda^j \mu^j}$ jest udziałem oczekiwanej wysokości szkód typu j w oczekiwanej wartości wszystkich szkód, natomiast $r_l^j = \mathbb{E}(\Theta^j|L)$

jest cząstkowym mnożnikiem składki związanym ze szkodami typu j . Oznacza to, że optymalne mnożniki składki mogą być wyznaczone za pomocą metod numerycznych jako ważona średnia współczynników r_l^j .

3.2. Analiza symulacyjna

W tej części pracy zbadane zostaną własności omawianego SBM uwzględniającego rodzaj szkody, a szczególnie wrażliwość wybranych miar opisujących SBM na zmianę założeń dotyczących siły zależności między ukrytymi parametrami ryzyka. Ze względu na duży stopień złożoności obliczeniowej posłużono się metodą Monte Carlo, przyjmując pewne uproszczenia dotyczące rozkładów brzegowych i rozkładu łącznego w stosunku do wyników przedstawionych w punkcie 2.2. – rozkład logarytmiczny został zastąpiony rozkładem gamma, a kopuła Gumbela została zastąpiona kopułą Claytona.

Na potrzeby analizy symulacyjnej przyjęto następujące założenia:

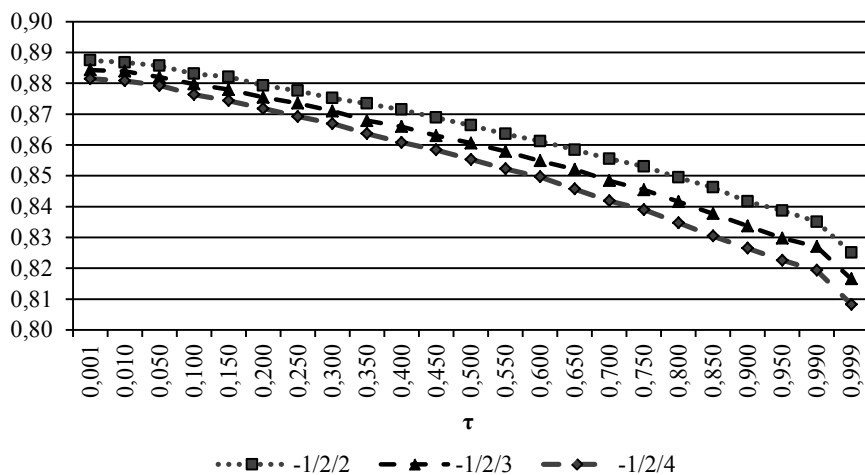
- Rozważany jest podział na szkody rzeczowe i szkody osobowe ($m = 2$).
- Rozkłady brzegowe liczby szkód to mieszane rozkłady Poissona z rozkładem gamma jako rozkładem mieszającym, o funkcji prawdopodobieństwa opisanej w punkcie 2.2.
- Parametry rozkładów brzegowych to $\lambda^R = 0,05$, $\lambda^O = 0,005$, $a^R = 1$, $a^O = 0,5$.
- Na potrzeby obliczeń przyjęto $w^R = 0,8$ oraz $w^O = 0,2$, co w przybliżeniu odpowiada rzeczywistym wartościom dla ubezpieczeń OC p.p.m. dla umów zawartych w 2013 r.
- Rozkład łączny wektora (θ^O, θ^R) dany jest kopułą Claytona.
- Rozważany jest 10-klasowy ($\mathcal{L} = \{0,1, \dots, 9\}$) SBM z klasą startową 4. Nagroda za bezszkodowy rok wynosi 1, kara za szkodę rzeczową 2, natomiast kara za szkodę osobową rozważana jest w trzech wariantach:
 - wariant I – 2 klasy (system -1/2/2),
 - wariant II – 3 klasy (system -1/2/3),
 - wariant III – 4 klasy (system -1/2/4).

We wszystkich wariantach funkcją agregującą jest suma.

- Wygenerowano losowo 30-letnią historię szkodową dla 5 mln umów ubezpieczenia, a następnie wyznaczono historię zmian klasy w SBM. Przyjęto, że po 30 latach SBM osiągnął stan stacjonarny, co potwierdzono, dokonując częściowego porównania z wyni-

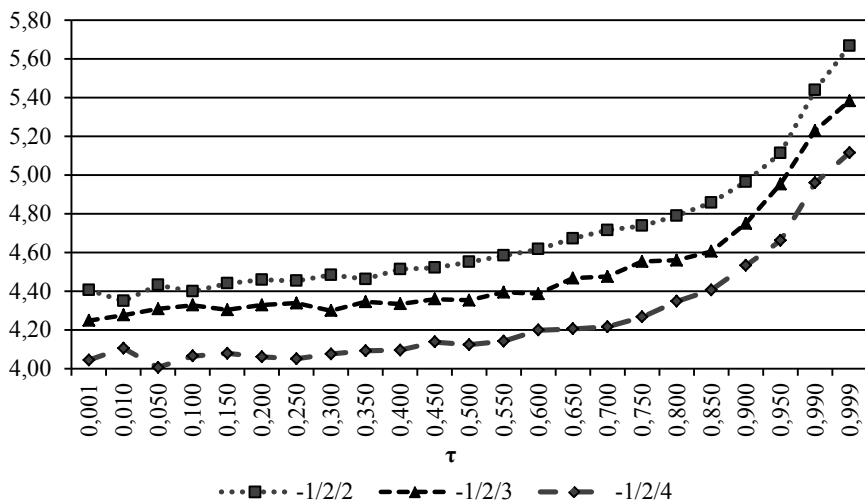
kami dokładnymi, uzyskanymi na drodze całkowania numerycznego.

Na kolejnych rysunkach przedstawiono obliczone optymalne mnożniki składki dla najniższej oraz najwyższej klasy taryfowej w zależności wartości współczynnika korelacji τ Kendalla.



Rys. 1. Optymalne mnożniki składki dla klasy 0

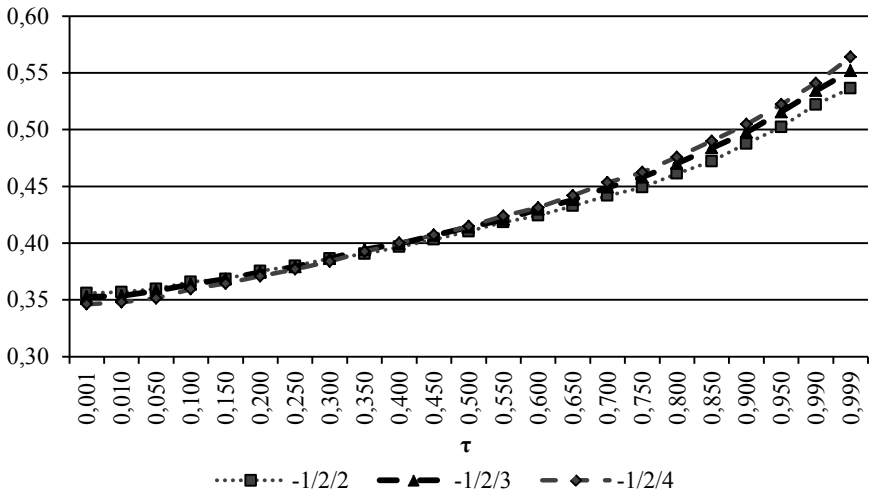
Źródło: opracowanie własne.



Rys. 2. Optymalne mnożniki składki dla klasy 9

Źródło: opracowanie własne.

Uzyskane wyniki wskazują, że dla najlepszej klasy taryfowej optymalny mnożnik składki jest malejącą funkcją parametru τ , natomiast dla najgorszej – rosnącą. Oznacza to, że wraz ze wzrostem siły zależności rośnie rozpiętość skali w systemie SBM, co przekłada się na większą niepewność co do wysokości płaconej przez ubezpieczonego składki. Obserwację tę potwierdza wykres obrazujący odchylenie standardowe składki płaconej w stanie stacjonarnym przez ubezpieczonego wybranego losowo z populacji (rys. 3).



Rys. 3. Odchylenie standardowe mnożnika składki w stanie stacjonarnym

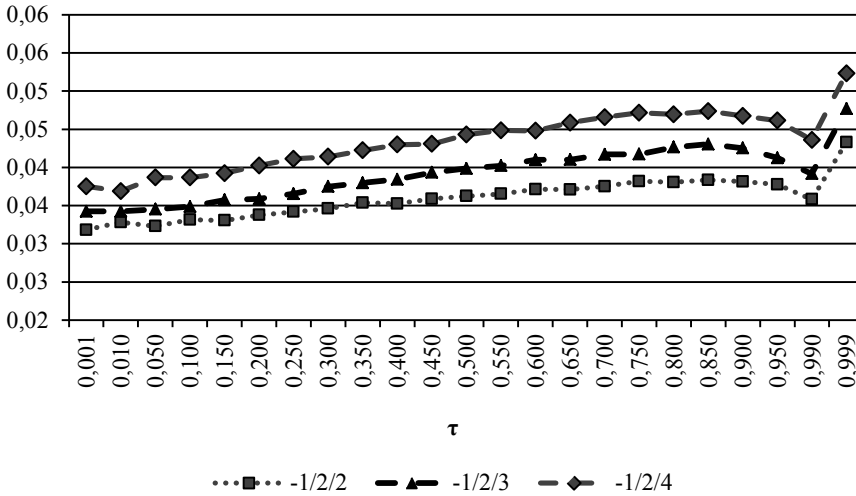
Źródło: opracowanie własne.

Odchylenie standardowe mnożnika składki w stanie stacjonarnym rośnie wraz ze wzrostem parametru τ , osiągając dla silnej zależności wartości o 50-60% wyższe niż w przypadku niezależności, przy czym wyniki są zbliżone dla wszystkich wariantów.

Wyznaczony został również wariant miary nazywanej względnym stacjonarnym oczekiwanym poziomem składki (*Relative Stationary Average Level* – RSAL):

$$RSAL(\tau) = \frac{\bar{r}(\tau) - \min(r_s(\tau))}{\max(r_s(\tau)) - \min(r_s(\tau))},$$

gdzie $\bar{r}(\tau)$ to oczekiwany mnożnik składki w stanie stacjonarnym dla parametru τ , a $r_s(\tau)$ jest optymalnym mnożnikiem składki w klasie s wyznaczonym przy współczynniku korelacji τ .



Rys. 4. Wartości wskaźnika RSAL w zależności od parametru τ

Źródło: opracowanie własne.

Wskaźnik RSAL dla badanego systemu jest niski i nieznacznie rośnie wraz ze wzrostem siły zależności. Wynika to z faktu, że wraz ze wzrostem siły zależności rośnie koncentracja ubezpieczonych w najlepszej i najgorszej klasie taryfowej, przy czym wzrost udziału ubezpieczonych w najgorszej klasie jest względnie dużo większy.

Na zakończenie przeanalizowana zostanie efektywność opisywanego SBM. Efektywność została tutaj zdefiniowana jako błąd średniokwadratowy aproksymacji rzeczywistego względnego poziomu ryzyka za pomocą mnożnika składki. Zbadano zatem wielkość:

$$\mathbb{E}(r(\theta^O, \theta^R) - \tilde{r}(\theta^O, \theta^R))^2,$$

gdzie:

- $r(\theta^O, \theta^R) = \frac{\mathbb{E}(S|\theta^O, \theta^R)}{\mathbb{E}(S)} = w^O \cdot \theta^O + w^R \cdot \theta^R$ jest rzeczywistym względnym poziomem ryzyka,
- $\tilde{r}(\theta^O, \theta^R) = \sum_{l=0}^s r_l \cdot \mathbb{P}(L = l|\theta^O, \theta^R)$ jest wartością oczekiwaną mnożnika składki stacjonarnej płaconej przez ubezpieczonego o wartości parametru ryzyka (θ^O, θ^R) .

W tabeli 5 przedstawiono porównanie tak zdefiniowanej efektywności SBM w przypadku bliskim niezależności ($\tau = 0,001$), a także w przypadku, gdy $\tau = 0,95$ (wartość zbliżona do otrzymanej w punkcie 2) oraz $\tau = 0,99$ (zależność bliska funkcyjnej).

Tabela 5. Porównanie efektywności SBM

τ	-1/2/2	-1/2/3	-1/2/4	Różnica względna
0,001	0,72	0,72	0,71	-1,56%
0,950	0,65	0,63	0,62	-3,87%
0,990	0,64	0,63	0,61	-3,89%

Źródło: opracowanie własne.

Uzyskane wyniki wskazują, że wraz ze wzrostem siły korelacji rośnie względna efektywność systemu najsurowiej karzącego spowodowanie szkody osobowej. Oznacza to, że w przypadku występowania istotnej korelacji między liczbą szkód różnych rodzajów uwzględnienie rodzaju szkody w regułach przejścia może zwiększyć efektywność systemu *bonus-malus* w porównaniu z SBM, który wszystkie szkody traktuje jednakowo, niezależnie od ich rodzaju.

Podsumowując wyniki przeprowadzonej analizy, można stwierdzić, że:

- wraz ze wzrostem siły zależności rośnie zmienność optymalnej skali SBM,
- wzrost siły zależności zwiększa koncentrację ubezpieczonych w najlepszej i najgorszej klasie taryfowej,
- względna efektywność SBM uwzględniającego rodzaj szkody rośnie wraz ze wzrostem korelacji między liczbą szkód różnych rodzajów.

4. Zakończenie

W niniejszej pracy przeanalizowana została zależność stochastyczna między zmiennymi losowymi reprezentującymi liczbę szkód rzeczowych i liczbę szkód osobowych powodujących odpowiedzialność ubezpieczyciela z tytułu zawartej umowy ubezpieczenia odpowiedzialności cywilnej posiadaczy pojazdów mechanicznych.

Analiza empiryczna danych pochodzących z bazy danych Ośrodka Informacji Ubezpieczeniowego Funduszu Gwarancyjnego pozwoliła zidentyfikować występowanie korelacji między tymi zmiennymi losowymi. Jednocześnie uzyskane wyniki wskazują, że wykorzystanie mieszanych rozkładów Poissona, w których parametry rozkładu warunkowego są zależnymi zmiennymi losowymi o strukturze zależności danej kopułą, pozwala w zadowalający sposób opisać proces powstawania szkód.

W drugiej części pracy przedstawiono system *bonus-malus*, w którym reguły przejścia uwzględniają rodzaj szkody spowodowanej przez ubezpieczonego. Analiza własności takiego systemu wskazuje, że w przypadku występowania silnej korelacji między ukrytymi parametrami ryzyka (zidentyfikowanej w pierwszej części pracy) pozwala on na lepszą ocenę ryzyka spowodowania szkody przez ubezpieczonego.

Przedstawiona analiza może być rozszerzona na kilka sposobów. Po pierwsze, może być pożądana budowa bardziej złożonego łącznego modelu liczby szkód różnych rodzajów. Po drugie, możliwe jest zastosowanie innych kryteriów agregacji szkód niż zaproponowana kombinacja liniowa. W końcu możliwa jest konstrukcja zupełnie innych reguł przejścia, pozwalających zwiększyć efektywność stosowanego systemu *bonus-malus*.

Literatura

- Angers J.F., Desjardins D., Dionne G., Guertin F., 2006, *Vehicle and fleet random effect in a model of insurance rating for fleets of vehicles*, Astin Bulletin, vol. 36, no. 1, s. 25-77.
- Boucher J.P., Denuit M., Guillén M., 2008, *Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions*, Variance, vol. 2, no. 1, s. 135-162.
- Denuit M., Maréchal X., Pitrebois S., Walhin J., 2007, *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, Wiley, New York.
- McNeil A.J., Frey R., Embrechts P., 2005, *Quantitative Risk Management – Concepts, Techniques and Tools*, Preston University Press.
- Nelsen R.B., 2006, *An Introduction to Copulas*, Springer, New York.
- Norberg R., 1976, *A credibility theory for automobile bonus systems*, Scandinavian Actuarial Journal, vol. 36, Issue 2.
- Pinquet J., 1997, *Allowance for cost of claims in bonus-malus systems*, Astin Bulletin, vol. 27, no. 1, s. 33-57.
- Podgórska M., Kryszewski B., Niemiec M., 2008, *Fair Bonus-Malus Systems*, Roczniki Kolegium Analiz Ekonomicznych, Szkoła Główna Handlowa w Warszawie, nr 18/2008, s. 11-24.
- Sawicka J., 2013, *Model stochastycznej zależności liczby szkód i wartości pojedynczej szkody*, Roczniki Kolegium Analiz Ekonomicznych, Szkoła Główna Handlowa w Warszawie, nr 31, s. 157-183.
- Shi P., Valdez E., 2014, *Multivariate negative binomial models for insurance claim counts*, Insurance: Mathematics and Economics, vol. 55, s. 18-29.
- Ustawa z dnia 22 maja 2003 r. o ubezpieczeniach obowiązkowych, Ubezpieczeniowym Funduszu Gwarancyjnym i Polskim Biurze Ubezpieczycieli Komunikacyjnych, Dz.U.2013, poz. 392, t.j.
- Vernic R., 1997, *On the bivariate generalized Poisson distribution*, Astin Bulletin, vol. 27, no. 1, s. 23-31.
- Vernic R., 2000, *A multivariate generalization of the generalized Poisson distribution*, Astin Bulletin, vol. 30, no. 1, s. 57-67.

**STOCHASTIC DEPENDENCE IN ACTUARIAL MODELS
FOR A POSTERIORI RISK CLASSIFICATION**

Summary: The purpose of this paper is to analyze the effects of stochastic dependence in actuarial models for a posteriori risk classification in automobile insurance. It is shown that in many practical situations the claims generating the process is multidimensional and the traditional assumption of independence between its components is not justified. The main aim of this paper is to analyze the correlation between different types of claims in motor third-party liability insurance (bodily injury and property damage) and to describe a bonus-malus system in which this distinction is taken into account in the transition rules. It is shown that in the case of a strong correlation such a bonus-malus system can be more efficient than a standard system which relies only on the total number of claims.

Keywords: automobile insurance, stochastic dependence, bonus-malus system, a posteriori risk classification, copula.