Norbert Paska[1]

Zastosowanie modeli ZINB GLMM z efektem losowym agenta w taryfikacji ubezpieczeń majątkowych

Streszczenie

W artykule przestawiono tematykę dotyczącą aktuarialnej analizy częstości szkód przy wykorzystaniu uogólnionych liniowych modeli mieszanych wraz z modelami z nadwyżką zer. Dodatkowo rozważano zastosowanie w owych modelach rozkładu ujemnego dwumianowego w celu wyeliminowania problemu nadmiernego rozproszenia. Badanie oparto na danych panelowych obejmujących dane tych samych pojazdów w kolejnych okresach po zawarciu rocznych polis komunikacyjnych. Szeroko opisano pojęcie efektu losowego oraz wykazano możliwość użycia w tej funkcji agenta ubezpieczeniowego. Wszystkie modele zawierały paręnaście zmiennych występujących jako efekty stałe oraz jedną zmienną jako efekt losowy. Finalnie porównano wyniki zastosowania modeli GLM, ZI GLM, GLMM, ZI GLMM z rozkładem Poissona oraz rozkładem ujemnym dwumianowym oraz dokonano wyboru modelu, który potrafi najadekwatniej do ryzyka odwzorować prawdopodobieństwo wystąpienia szkody.

Słowa kluczowe: uogólnione liniowe modele mieszane, ZINB, agent ubezpieczeniowy, efekty losowe, ubezpieczenia komunikacyjne, taryfikacja

JEL: C01, C23, C25

1. Wstęp

Uogólnione modele liniowe z funkcją łącznikową o rozkładzie Poissona oraz Gamma stanowią powszechnie wykorzystywane w naukach aktuarialnych narzędzie do modelowania m.in. liczby szkód oraz wysokości średniej szkody. Szerokie omówienie zastosowania tego typu modeli w naukach aktuarialnych możemy znaleźć m.in. w pracy S. Habermana i A.E. Renshawa[2].

---

[1] Szkoła Główna Handlowa w Warszawie, Kolegium Analiz Ekonomicznych.

[2] S. Haberman, A.E. Renshaw, *Generalized Linear Models and Actuarial Science*, „Journal of the Royal Statistical Society" (Series D – The Statistician) 1996, vol. 45, no. 4, s. 407–436.

Pomimo znacznej popularności modeli GLM w taryfikacji ubezpieczeń K.K.W. Yau i in.[3] w wątpliwość podali to, czy przy modelowaniu za ich pomocą danych panelowych zawierających obserwacje tego samego kierowcy w kolejnych latach polisowych (dla polis rocznych) spełnione jest jedno z założeń modeli GLM mówiące o niezależności rozkładów zmiennych, i jako efekt losowy grupujący skorelowane ze sobą obserwacje zaproponowali wymiar czasu. W literaturze przedmiotu można znaleźć także inne przykłady efektów losowych, np. wskazany przez A. Wolny-Dominiak[4] wymiar geograficzny czy też markę pojazdu.

W ocenie autora niniejszego artykułu w wątpliwość można podać także niezależność zmiennych w przypadku obserwacji reprezentowanych przez tego samego agenta ubezpieczeniowego. Dane pokazują bowiem, że homogeniczne grupy ryzyka mogą osiągać różne wyniki szkodowe w zależności od pośrednika.

Dodatkowym problemem występującym w analizie szkód w ubezpieczeniach jest zbyt duża liczba zer oraz problem nadmiernej dyspersji[5]. Z tego względu zarówno szeroko opisywane i wykorzystywane w literaturze modele aktuarialne (GLM), jak i ich modyfikacje wprowadzające tylko korektę ze względu na efekt losowy (GLMM) mogą okazać się niewystarczające.

Wychodząc naprzeciw przedstawionym problemom, w niniejszej pracy zaproponowano wprowadzenie do taryfikacji ubezpieczeń modelu zarówno obejmującego korektę ze względu na efekt losowy (odpornego na zbyt dużą liczbę zer w danych), jak i – poprzez zastosowanie rozkładu ujemnego dwumianowego – eliminującego problem nadmiernej dyspersji. Innowacyjne w pracy jest także wskazanie i zastosowanie jako efektu losowego wymiaru pośrednika ubezpieczeniowego. Badanie oparto na danych panelowych, dzięki czemu uzyskano silne predyktory szkodowości, zakwalifikowane następnie do roli efektu stałego. Umożliwiło to uzyskanie adekwatniej opisujących ryzyko i bardziej stabilnych w czasie oszacowań parametrów modeli.

Warto także zauważyć, że w dostępnej literaturze dotyczącej tematyki aktuarialnej – choć wspomina się o możliwości zastosowania jednocześnie modelu *zero-inflated* oraz GLMM – nie pokazano konkretnych przykładów zastosowania

---

[3] K.K.W. Yau, K.C.H. Yip, H.K. Yuen, *Modelling repeated insurance claim frequency data using the generalized linear mixed model*, „Journal of Applied Statistics" 2003, vol. 30, no. 8, s. 857–865.

[4] A. Wolny-Dominiak, *Taryfikacja w ubezpieczeniach majątkowych z wykorzystaniem modeli mieszanych*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2014.

[5] K.C.H. Yip, K.K.W. Yau, *On modeling claim frequency data in general insurance with extra zeros*, „Insurance Mathematics and Economics" 2005, vol. 36(2), s. 153–163, https://www.sciencedirect.com/science/article/pii/S0167668704001301.

tego typu modelu z wieloma zmiennymi do analizy częstości szkód, co w niniejszej pracy zostanie uwzględnione.

W części drugiej niniejszego artykułu przedstawiono dane oraz metodologię badania. Podsumowanie najważniejszych informacji dotyczących uogólnionych liniowych modeli mieszanych znajduje się w części trzeciej, natomiast czwartą poświęcono tematyce efektu losowego ze szczególnym uwzględnieniem wymiaru pośrednika ubezpieczeniowego. Następnie opisano modele z nadwyżką zer, a w części szóstej przedstawiono wyniki badania oraz porównanie klas modeli GLM, ZI GLM, GLMM, ZI GLMM z rozkładem Poissona oraz rozkładem ujemnym dwumianowym. W części siódmej dokonano podsumowania.

2. Opis zbioru danych oraz metodologia badania

Zbiór danych składa się z 99 702 wierszy – każdy z nich reprezentuje roczną polisę komunikacyjną na pakiet OC i Autocasco, zawartą w okresie od 1 czerwca 2014 r. do 31 maja 2016 r. Polisy te zostały zakupione u 505 dealerów samochodowych. Braki danych uzupełniono:

- za pomocą średniej dla zmiennych liczbowych;
- nadaniem kategorii „other" zmiennym nienumerycznym.

Należy tutaj nadmienić, że braki danych stanowiły maksymalnie 0,17% wartości zmiennej i miały charakter losowy, co zostało uznane za wystarczającą legitymizację użycia średniej do ich imputacji.

Dodatkowo zmienne nienumeryczne zamieniono na zmienne binarne (każdą z kategorii zmiennej), a spośród nich te, które mają udział mniejszy niż 3%, odrzucono z dalszej analizy, jako kategorię małoliczną. Następnie część zmiennych usunięto: zabronioną prawnie (w taryfikacji ubezpieczeń) płeć[6] oraz zmienne silnie skorelowane (współczynnik Pearsona $\geq |0{,}65|$).

Finalnie zbiór danych zawierał 22 zmienne, które mogły wystąpić w kategorii efektu stałego; 14 z nich trafiło do modelu i zostały opisane w tabeli 1, a ich statystyki opisowe w tabeli 2.

[6] Ustawa o działalności ubezpieczeniowej z dnia 22 maja 2003 r. (Dz.U. z 2003 r. Nr 124, poz. 1151).

Tabela 1. Efekty stałe

Kategoria zmiennej	Zmienna	Opis
Polisa	RENEWAL	czy polisa jest wznowieniem?
Klient	IS_LEASING	czy pojazd wzięty w leasing?
	IS_COMPANY	czy właścicielem pojazdu jest firma?
	CLIENT_AGE	wiek klienta
Historia polisowo-szkodowa klienta	HIST_AC_POL_YEARS	liczba lat polisowych w ostatnich 4 latach
	HIST_AC_NO_CLAIM_4Y	liczba szkód w ostatnich 4 latach
Pojazd	IS_DIESEL	czy pojazd ma silnik diesela?
	CAR_AGE	wiek pojazdu
	CAPACITY_th	pojemność silnika (tys. cm^3)
	CAR_MAKE_AGR	marka pojazdu
Miejsca zamieszkania ubezpieczonego	DIST_MIN	odległość od miasta wojewódzkiego
	D_BIG_CITY	czy miasto wojewódzkie?
	VOIVODESHIP_CENTRALNA_PL	województwa: mazowieckie, kujawsko-pomorskie, wielkopolskie, świętokrzyskie, łódzkie
	VOIVODESHIP_ZACHODNIA_PL	województwa: dolnośląskie, lubuskie, zachodniopomorskie

Źródło: opracowanie własne.

Tabela 2. Statystyki opisowe dla efektów stałych

Zmienna	min.	1Q	me	$\bar{x}$	3Q	maks.
RENEWAL	0	0	1	0,5444	1	1
IS_LEASING	0	0	0	0,1221	0	1
IS_COMPANY	0	0	0	0,1711	0	1
CLIENT_AGE	2	34,84	45,54	42,21	58,29	93,82
HIST_AC_POL_YEARS	0	2	4	3,002	4	4
HIST_AC_NO_CLAIM_4Y	0	0	0	0,614	1	325
IS_DIESEL	0	0	0	0,3381	1	1
CAR_AGE	0	2	4	4,637	7	19
CAPACITY_th	0	1,368	1,596	1,715	1,997	7,011
CAR_MAKE_AGR_FIAT	0	0	0	0,1652	0	1
CAR_MAKE_AGR_MAZDA	0	0	0	0,09338	0	1
CAR_MAKE_AGR_SKODA	0	0	0	0,08409	0	1
DIST_MIN	0	0	20,95	36,52	61,23	249,42

Zmienna	min.	1Q	me	$\overline{x}$	3Q	maks.
D_BIG_CITY	0	0	0	0,3561	1	1
VOIVODESHIP_ CENTRALNA_PL	0	0	0	0,4449	1	1
VOIVODESHIP_ ZACHODNIA_PL	0	0	0	0,1085	0	1

Źródło: opracowanie własne.

Zgodnie z metodologią przyjętą przez autorów *On modeling claim frequency data in general insurance with extra zeros*[7] w celach porównawczych we wszystkich modelach użyto tych samych zmiennych, które zostały wybrane w oparciu o kryterium statystycznej istotności z $\alpha = 0,1$.

3. Uogólnione liniowe modele mieszane

Uogólnione liniowe modele mieszane są rozszerzeniem uogólnionych modeli liniowych o mające niezerowe wariancje efekty losowe, zdefiniowane w liniowym predyktorze, przy założeniu nieobserwowanej wprost heterogeniczności w obrębie niektórych współczynników regresji.

K. Antonio i J. Beirlant[8] wskazują, że dla każdego wymiaru i dostępnych jest n_i obserwacji. Zakłada się, że powtórzone pomiary $Y_{i1}, \ldots, Y_{in_i}$ są niezależne oraz mają funkcję gęstości z rodziny rozkładów wykładniczych. Dla uogólnionych liniowych modeli mieszanych kanoniczną funkcję gęstości rozkładów wykładniczych można zapisać jako:

$$f\left(y_{ij}|b_i,\ \beta,\phi\right) = \exp\left(\frac{y_{ij}\theta_{ij} - \psi\left(\theta_{ij}\right)}{\phi} + c\left(y_{ij},\phi\right)\right),\ j = 1, \ldots, i, \tag{1}$$

$$\mu_{ij} = E\left[Y_{ij}|b_i\right] = \psi'\left(\theta_{ij}\right) \tag{2}$$

oraz

[7] K.C.H. Yip, K.K.W. Yau, op.cit.

[8] K. Antonio, J. Beirlant, *Actuarial statistics with generalized linear mixed models*, „Insurance Mathematics and Economics" 2007, vol. 40, issue 1, s. 58–76.

$$Var\left[Y_{ij}|b_i\right] = \phi\psi''\left(\theta_{ij}\right) = \phi V\left(\mu_{ij}\right), \tag{3}$$

gdzie $V(.)$ jest funkcją wariancji, ψ (.) oraz c (.) znanymi funkcjami, a θ oraz ϕ parametrami rozkładu.

Można przedstawić formalny zapis modelu GLMM jako kombinację liniową zmiennej zależnej ze zmiennymi objaśniającymi oraz efektami losowymi, które wpływają na zmienną zależną poprzez funkcję łącznikową, tj.:

$$g\left(\mu_{ij}\right) = x'_{ij}\beta + z'_{ij}b_i, \tag{4}$$

gdzie $g(.)$ jest funkcją łącznikową, β symbolizuje wektor parametrów efektów stałych, a b_i wektor efektów losowych[9].

4. Efekty losowe

Specyfikację modelu GLMM uzupełnia założenie, że efekty losowe b_i są względem siebie niezależne oraz mają funkcję gęstości $f(b_i \mid \alpha)$, gdzie α symbolizuje nieznane parametry gęstości. Tradycyjnie przyjmuje się, że efekty losowe mają rozkład normalny z wartością oczekiwaną równą 0 oraz macierzą kowariancji G:

$$b_i \sim N(0, G).$$

Korelacje pomiędzy obserwacjami wewnątrz tego samego wymiaru wynikają ze współdzielenia tego samego efektu losowego b_i[10].

Jak wspomniano we wstępie do niniejszej pracy, w literaturze najczęściej podnosi się brak niezależnych rozkładów zmiennych dla wymiaru czasu – przykłady można znaleźć m.in. w dwóch wspomnianych już pracach[11]. Warto jednak uwzględnić w kategorii efektu losowego także inne wymiary agregujące obserwację, tj. sprawiające, że rozkłady zmiennych mogą być od siebie zależne – tak jak np. wymiar pośrednika ubezpieczeniowego. Po pierwsze, dlatego że agent pośredniczy między zakładem ubezpieczeniowym i klientem (m.in. poprzez

[9] Ibidem.
[10] Ibidem.
[11] K.K.W. Yau, K.C.H. Yip, H.K. Yuen, op.cit.; K. Antonio, J. Beirlant, op.cit.

dostarczanie ubezpieczycielowi informacji potrzebnych do wyliczenia składki) i to od jego dociekliwości, skrupulatności i prawdomówności w znacznej mierze zależy jakość danych. Po drugie, agent, często znając osobiście klienta, może – wprawdzie w pewnym stopniu subiektywnie, lecz często też dosyć celnie – ocenić jego ryzyko ubezpieczeniowe. Dodatkowo z obserwacji wynika, że często relacje pośrednika z ubezpieczalniami oraz podpisane kontrakty sprzedażowe decydują o tym, do której firmy jaki typ ryzyka agent skieruje.

Ma to także odzwierciedlenie w danych – w tabeli 3 przedstawiono analizę jednoczynnikową, w której pośredników ubezpieczeniowych podzielono na pięć grup. Oznaczenie (std) dotyczy odchylenia tych wartości od ich wartości średniej. Na potrzeby analizy założono, że składka AC zaproponowana przez zakład ubezpieczeń jest możliwie najbardziej adekwatna do ryzyka (ze względu na możliwe parametry taryfowe).

Tabela 3. Szkodowość według grup dystrybutorów

Grupa	Liczebność grupy	Liczba polis	Częstość (std)	Śr. szkoda (std)	Iloczyn	Składka AC (std)
I	63	10 072	101,1%	104,5%	105,6%	97,2%
II	92	18 638	85,3%	109,6%	93,4%	99,8%
III	136	31 147	100,8%	98,0%	98,8%	98,9%
IV	95	21 079	107,4%	100,1%	107,5%	101,6%
V	119	18 766	104,4%	93,0%	97,1%	101,8%
Suma	**505**	**99 702**	**100,0%**	**100,0%**	**100,0%**	**100,0%**

Źródło: opracowanie własne.

Zauważmy, że pomimo zbliżonej wartości składki AC (pochodna czynników ryzyka) wyniki szkodowe w owych pięciu grupach odchylają się od siebie. Na przykład grupa II, pomimo zbliżonej wartości składki względem średniej, charakteryzuje się o ~6,6% lepszą szkodowością niż średnia, natomiast grupa IV o 7,5% gorszą.

5. Modele z nadwyżką zer

Składkę czystą (tj. za ryzyko) otrzymujemy jako wynik iloczynu prawdopodobieństwa wystąpienia szkody i oczekiwanej wartości szkody. Modele prawdopodobieństwa szkody najczęściej przyjmują postać modelu z rozkładem Poissona.

Podejście to jest jednak często niepoprawne, m.in. ze względu na nadmierny udział obserwacji bezszkodowych (w kontekście danych ubezpieczeniowych) lub – bardziej ogólnie – nadmierny udział zer w próbie.

Rozkład liczebności szkód AC w zbiorze danych opisanym szczegółowo w części 2 przedstawiono w tabeli 4. Potwierdza to nasze przypuszczenia dotyczące dużego udziału zer w próbie.

Tabela 4. Rozkład liczby szkód AC

Liczba szkód	0	1	2	3 lub więcej
Liczebność	86 482	10 798	733	86

Źródło: opracowanie własne.

Do analizowania danych tego typu najczęściej stosuje się *hurdle model* lub *zero-inflated model*, przyjmując, że obserwacje, dla których zachodzi $y = 0$, stanowią znacząco różny zbiór w porównaniu z pozostałymi jednostkami obserwacji[12].

Zaproponowany przez J. Mullahy'ego[13] *hurdle model* składa się z dwóch, niezależnych części: modelu z binarną zmienną zależną (gdzie jeden z poziomów oznacza zero, tj. brak szkód) oraz modelu uciętego dla wartości zero, modelującego dodatnie realizacje zmiennej zależnej (liczba szkód). Model ten możemy formalnie zapisać w postaci:

$$\begin{cases} P\left(y_i = 0\right) = f_1(0) \\ P\left(y_i = j\right) = f_2(j)\dfrac{1-f_1(0)}{1-f_2(0)},\ j = 1,2,3,\ldots \end{cases} \tag{5}$$

Alternatywnie można rozważyć dyskretną, nieujemną zmienną losową Y (jak np. liczba szkód) z nadwyżką liczby zer (ang. *zero-inflated distribution*) z parametrami $\varnothing$ $(0 < \varnothing < 1)$ oraz λ reprezentującymi kolejno udział zer oraz parametr średniej w rozkładzie Poissona. Modele *zero-inflated* można ogółem zapisać jako:

$$\begin{cases} P\left(Y = 0\right) = \varnothing + \left(1-\varnothing\right) * P\left(K = 0\right) \\ P\left(Y = y\right) = \left(1-\varnothing\right) * P\left(K = y\right),\ y = 1,2,3,\ldots \end{cases} \tag{6}$$

[12] P. Fiszeder, M. Polasik, *Modelowanie liczby transakcji dokonywanych przy użyciu gotówki i kart płatniczych na rynku polskim*, „Acta Universitatis Nicolai Copernici Ekonomia XXXIX", z. 389, Toruń 2009, s. 93–104.

[13] J. Mullahy, *Specification and Testing in Some Modified Count Data Models*, „Journal of Econometrics" 1986, vol. 33, no. 3, s. 341–365.

Zmienna losowa K może przyjmować różne formy uogólnienia rozkładu Poissona, np. gdy wśród danych obserwujemy dodatkowo problem nadmiernego rozproszenia, możemy zastosować rozkład ujemny dwumianowy. Kolejno powstałyby zatem modele ZIP (*zero-inflated Poisson*) oraz ZINB (*zero-inflated negative binomial*), których funkcję prawdopodobieństwa, wartość oczekiwaną, wariancję oraz parametry kształtu przedstawiono w tabeli 5[14].

Tabela 5. Model ZIP i ZINB

Model	P(K = y)	E(Y)	Var(Y)	Parametr kształtu
ZIP	$\frac{\varepsilon^{-\lambda}\lambda^{y}}{y!}$	$(1-\varnothing)\ \lambda$	$\mathrm{E(Y)}(1+\lambda-\mathrm{E(Y)})$	brak
ZINB	$\binom{r+k-1}{r} t^{k}(1-t)^{r}$	$(1-\varnothing)\ \lambda$	$\mathrm{E(Y)}(1+\lambda(1+\delta)-\mathrm{E(Y)})$	$\delta = \frac{1}{K}$

Źródło: K.C.H. Yip, K.K.W. Yau, *On modeling claim frequency data in general insurance with extra zeros*, „Insurance Mathematics and Economics" 2005, vol. 36(2), s. 153–163, https://www.sciencedirect.com/science/article/pii/S0167668704001301.

Warto zauważyć, że dla $\phi = 0$ model ZIP jest redukowany do modelu Poissona, natomiast dla $\delta = 0$ model ZINB jest redukowany do modelu ZIP.

6. Modele częstości szkód AC – wybór optymalnej klasy modeli

Przeanalizowano model częstości szkód AC przy użyciu modeli: GLM, ZI GLM, GLMM, ZI GLMM z rozkładem Poissona oraz rozkładem ujemnym dwumianowym. Porównanie wyników uzyskanych kryteriów informacyjnych AIC oraz BIC przedstawiono w tabeli 6.

Najlepszym modelem ze względu na minimalizację kryteriów informacyjnych okazał się model *zero-inflated negative binomial generalized linear mixed model*. Jego użycie daje dokładniejsze oszacowania parametrów modelu częstości szkód, co finalnie przedkłada się na lepsze pod względem ryzyka oszacowanie składki czystej. Owa relacja zachodzi także w porównaniu par: modeli

[14] K.C.H. Yip, K.K.W. Yau, op.cit.

zero-inflated z modelami nieuwzględniającymi nadwyżki zer oraz modeli GLMM zestawionymi z modelami GLM.

Tabela 6. Kryteria informacyjne dla modeli częstości szkód AC

Model	AIC	BIC
GLM Poisson	77 952,2	78 113,9
ZI GLM Poisson	77 394,0	77 565,2
GLMM Poisson	77 935,1	78 106,2
ZI GLMM Poisson	77 385,4	77 566,1
ZI GLMM NB	77 301,4	77 491,6

Źródło: opracowanie własne.

Wyniki oszacowania parametrów modeli z tabeli 6 przedstawiono w tabeli 7.

Tabela 7. Oszacowania efektów stałych modelu częstości szkód AC

	GLM Poisson	ZI GLM Poisson	GLMM Poisson	ZI GLMM Poisson	ZI GLMM NB
Intercept	–1,85293	–1,22648	–1,85826	–1,23555	–1,86202
RENEWAL	0,04282	0,04121	0,04536	0,04329	0,04165
IS_DIESEL	0,11420	0,11422	0,11900	0,11805	0,11783
HIST_AC_POL_YEARS	–0,05887	–0,05914	–0,05974	–0,05975	–0,05923
HIST_AC_NO_CLAIM_4Y	0,00438	0,00576	0,00441	0,00574	0,00517
CAR_AGE	–0,04731	–0,04680	–0,04710	–0,04665	–0,04714
IS_LEASING	0,12737	0,12899	0,12131	0,12398	0,12422
CLIENT_AGE	–0,00442	–0,00437	–0,00442	–0,00437	–0,00437
DIST_MIN	0,00069	0,00069	0,00065	0,00066	0,00067
CAPACITY_th	0,14113	0,14245	0,13784	0,13968	0,14080
CAR_MAKE_AGR_FIAT	–0,12952	–0,13014	–0,13319	–0,13284	–0,13280
CAR_MAKE_AGR_MAZDA	0,13121	0,12427	0,13885	0,13140	0,13484
CAR_MAKE_AGR_SKODA	–0,15006	–0,15091	–0,15683	–0,15632	–0,15662
IS_COMPANY	–0,15632	–0,15614	–0,15507	–0,15500	–0,15334
D_BIG_CITY	0,12505	0,12538	0,12457	0,12518	0,12429
VOIVODESHIP_ CENTRALNA_PL	–0,13204	–0,13339	–0,11640	–0,12096	–0,12271
VOIVODESHIP_ ZACHODNIA_PL	0,18611	0,18458	0,17675	0,17819	0,18121

Źródło: opracowanie własne.

Wnioski wyciągnięte z interpretacji ocen parametrów w poszczególnych modelach są zbliżone, tj. zmienne są stymulantami lub destymulantami niezależnie od modelu, co jest zgodne z oczekiwaniami. Hipoteza badania brzmiała bowiem: efektem zastosowania kolejnych klas modeli będzie tylko dokładniejsze oszacowanie parametrów, a nie ich całkowicie inny charakter.

Szczegóły oszacowania najlepszego ze względu na minimalizację kryteriów informacyjnych modelu przedstawiono w tabeli 8.

Tabela 8. Wyniki estymacji modelu ZI GLMM NB

	Parametr	Odch. stand.	Stat. z	Pr (>|z|)	Iloraz szans
(Intercept)	–1,8620	0,0647	–28,7820	< 2e–16***	–
RENEWAL	0,0417	0,0223	1,8680	0,06172	1,043
IS_DIESEL	0,1178	0,0226	5,2180	1,81e–07***	1,125
HIST_AC_POL_YEARS	–0,0592	0,0070	–8,5040	< 2e–16***	0,942
HIST_AC_NO_CLAIM_4Y	0,0052	0,0016	3,2550	0,001135**	1,005
CAR_AGE	–0,0471	0,0037	–12,6210	< 2e–16***	0,954
IS_LEASING	0,1242	0,0324	3,8300	0,000128***	1,132
CLIENT_AGE	–0,0044	0,0009	–4,7290	2,25e–06***	0,996
DIST_MIN	0,0007	0,0003	2,1160	0,034373*	1,001
CAPACITY_th	0,1408	0,0203	6,9450	3,80e–12***	1,151
CAR_MAKE_AGR_FIAT	–0,1328	0,0331	–4,0130	5,99e–05***	0,876
CAR_MAKE_AGR_MAZDA	0,1348	0,0355	3,7950	0,000148***	1,144
CAR_MAKE_AGR_SKODA	–0,1566	0,0435	–3,6010	0,000317***	0,855
IS_COMPANY	–0,1533	0,0518	–2,9580	0,003094**	0,858
D_BIG_CITY	0,1243	0,0284	4,3720	1,23e–05***	1,132
VOIVODESHIP_ CENTRALNA_PL	–0,1227	0,0237	–5,1740	2,29e–07***	0,885
VOIVODESHIP_ ZACHODNIA_PL	0,1812	0,0331	5,4690	4,52e–08***	1,199
Zero-inflated model					
(Intercept)	–15,7	537	–0,029	0,977	
Wariancja efektu losowego					
AGENT_GROUP_DWH (Intercept)	0,00549				

Źródło: opracowanie własne.

Na poziomie istotności 0,1 należy odrzucić hipotezę zerową, mówiącą o braku istotności każdej z użytych zmiennych. Wszystkie zmienne użyte w modelu w kategorii efektu stałego są istotne, a ich charakter potwierdza intuicję, że m.in.:

- pojazdy zasilane silnikiem diesla co do zasady jeżdżą częściej, co powoduje wyższą ekspozycję na ryzyko i wyższą szkodowość;
- im więcej lat polisowych ma klient oraz im jest starszy, tym wzrasta jego doświadczenie i spada szkodowość;
- w dużych miastach wypadki zdarzają się częściej.

7. Podsumowanie

W artykule przestawiono tematykę aktuarialnej analizy częstości szkód w oparciu o uogólnione liniowe modele mieszane oraz modele *zero-inflated*. Szeroko opisano pojęcie efektu losowego oraz wykazano możliwość użycia w tej funkcji pośrednika ubezpieczeniowego. Badanie oparto na danych panelowych, dzięki czemu uzyskano silne predyktory szkodowości zakwalifikowane następnie do roli efektu stałego. Umożliwiło to uzyskanie lepiej dopasowanych do ryzyka i bardziej stabilnych w czasie oszacowań parametrów modeli.

Wychodząc naprzeciw problemom w modelowaniu szkód ubezpieczeniowych, w niniejszej pracy zaproponowano wprowadzenie do taryfikacji ubezpieczeń modelu zarówno obejmującego korektę ze względu na efekt losowy (odpornego na zbyt dużą liczbę zer w danych), jak i eliminującego problem nadmiernej dyspersji, który w porównaniu z innymi modelami okazał się ze względu na minimalizację kryteriów informacyjnych najlepszy. Wnioski wyciągnięte z interpretacji ocen parametrów tego modelu są zbliżone do ocen parametrów innych modeli, a dodatkowo oszacowania parametrów modelu są dokładniejsze, co finalnie przedkłada się na bardziej dopasowane do ryzyka oszacowanie składki czystej.

Bibliografia

Antonio K., Beirlant J., *Actuarial statistics with generalized linear mixed models*, „Insurance Mathematics and Economics” 2007, vol. 40, issue 1, s. 58–76.

Fiszeder P., Polasik M., *Modelowanie liczby transakcji dokonywanych przy użyciu gotówki i kart płatniczych na rynku polskim*, „Acta Universitatis Nicolai Copernici Ekonomia XXXIX”, z. 389, Toruń 2009, s. 93–104.

Haberman S., Renshaw A.E., *Generalized Linear Models and Actuarial Science*, „Journal of the Royal Statistical Society" (Series D – The Statistician) 1996, vol. 45, no. 4, s. 407–436.

Mullahy J., *Specification and Testing in Some Modified Count Data Models*, „Journal of Econometrics" 1986, vol. 33, no. 3, s. 341–365.

Ustawa o działalności ubezpieczeniowej z dnia 22 maja 2003 r. (Dz.U. z 2003 r. Nr 124, poz. 1151).

Wolny-Dominiak A., *Taryfikacja w ubezpieczeniach majątkowych z wykorzystaniem modeli mieszanych*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2014.

Yau K.K.W., Yip K.C.H., Yuen H.K., *Modelling repeated insurance claim frequency data using the generalized linear mixed model*, „Journal of Applied Statistics" 2003, vol. 30, no. 8, s. 857–865.

Źródła sieciowe

Yip K.C.H., Yau K.K.W., *On modeling claim frequency data in general insurance with extra zeros*, „Insurance Mathematics and Economics" 2005, vol. 36(2), s. 153–163, https://www.sciencedirect.com/science/article/pii/S0167668704001301.

* * *

Application of ZINB GLMM with the agent's random effect in non-life insurance ratemaking

Summary

The article presents an actuarial claim frequency analysis using generalized linear mixed models together with zero-inflated models. In addition, the use of binomial negative distribution in these models was considered to eliminate overdispersion. The study was based on annual motor policies panel data covering information about the same vehicles in subsequent periods. Random effects were widely described and the possibility of using an insurance agent in this role has been demonstrated. All models contained dozens of fixed effects and one variable as a random effect. Finally, the results of GLM, ZI GLM, GLMM, ZI GLMM with Poisson and negative binomial distribution were compared. The choice of the most risk-reflected model was made.

Keywords: generalized linear mixed models, zero-inflated negative-binomial, insurance agent, random effects, motor-insurance, ratemaking