

Szkoła Główna Handlowa w Warszawie

Kolegium Analiz Ekonomicznych

Autoreferat

do rozprawy doktorskiej mgr. Norberta Paski

**pod tytułem „Zastosowanie uogólnionych liniowych
modeli mieszanych w taryfikacji ubezpieczeń
komunikacyjnych”**

napisanej pod kierunkiem naukowym

prof. dr. hab. Bartosza Witkowskiego

oraz

dr. hab. Michała Bernardello, prof. SGH

Warszawa 2021

Spis treści

1. Uzasadnienie wyboru tematu.....	3
2. Przedmiot oraz zakres.....	4
3. Struktura pracy.....	8
4. Cel pracy.....	11
Bibliografia autoreferatu	14

1. Uzasadnienie wyboru tematu

Wedle informacji opublikowanej przez Komisję Nadzoru Finansowego (2020), w 2019 roku na polskim rynku ubezpieczeniowym w dziale II (pozostałe ubezpieczenia osobowe i majątkowe) działały 33 zakłady ubezpieczeń, dysponujące składką przypisaną w wysokości 31 miliardów złotych. Rynek ten można zatem uznać za konkurencyjny, co z kolei wymusza na firmach ubezpieczeniowych rywalizowanie o docelowego klienta poprzez prowizje, kontrakty premiowe dla sprzedawców, działania marketingowe, akcje promocyjne firmy i produktów, czy też szerokość zakresu ochrony oferowanych produktów. Jedną z ważniejszych ról w procesie pozyskania klienta pełni jednak cena ubezpieczenia. Jej wysokość powinna być wyliczona zarówno z uwzględnieniem aspektu skłonności klienta do jej opłacenia, jak i oceny ryzyka ubezpieczeniowego.

Dla firmy ubezpieczeniowej kluczowa jest zatem możliwie jak najdokładniejsza segmentacja ryzyka, uzyskiwana za pomocą modeli taryfikacyjnych. Dzięki nim zakład ubezpieczeń realizować może obydwa postulaty, które stawia przed nami ekonomia: po pierwsze realizowany jest postulat sprawiedliwości, gdyż ciężar odpowiedzialności za wyrządzone szkody w większym stopniu spoczywa na osobach je wywołujących. Z tego powodu ubezpieczyciel premiuje niższą składką te osoby, które szkód wyrządzają mniej, a „karze” wyższą tych, którzy są częstymi sprawcami szkód. Z drugiej strony firma dba o efektywność, gdyż tylko premiowanie i „karanie” osób, kolejno, mało i bardzo narażonych na szkodę, umożliwia firmie ubezpieczeniowej maksymalizację zysku – zarówno poprzez wyższy wolumen sprzedaży polis odpowiadających jej grupie docelowej, jak i poprzez zmniejszenie strat ponoszonych w konsekwencji wypłacanych odszkodowań.

Jak wskazują Antonio i Valdez (2012), w świetle heterogeniczności posiadanego portfela ryzyk, zakłady ubezpieczeń nie mogą oferować wszystkim ubezpieczonym tych samych stawek za ubezpieczenie, gdyż mogłoby to narazić firmę na zjawisko selekcji negatywnej¹ (ang. *adverse selection*). To z kolei według nich spowodować mogłoby niekorzystne konsekwencje finansowe, a w długim okresie także brak wypłacalności i upadek zakładu ubezpieczeń, a nawet destabilizację całego rynku ubezpieczeniowego. Brak odpowiedniej taryfikacji składki może prowadzić zatem do nieodpowiedniej i nierynkowej relacji

¹ Selekcja negatywna definiowana jest jako sytuacja, w której ubezpieczenia chętniej są kupowane przez osoby obciążone wyższym ryzykiem. Zagadnienie to jest ściśle powiązane ze zjawiskiem asymetrii informacji między ubezpieczającym się, a zakładem ubezpieczeń (Sulewski i Meuwissen, 2014).

posiadanych w portfelu zakładu „złych” i „dobrych” ryzyk². Antonio i Valdez (2012) podkreślają także, że w dłuższym okresie sytuacja ta wymusi podwyższenie średniej składki w zakładzie, aby utrzymać wypłacalności firmy. To z kolei sprawi, że firma stanie się jeszcze bardziej niekonkurencyjna w walce o „dobre” ryzyka, co spowoduje dalszy ich odpływ i utrzymujący się brak rentowności.

Efekt poprawy modeli taryfikacyjnych można uzyskać na co najmniej dwa sposoby. Pierwszy z nich polega na wprowadzeniu do taryfikacji nowych zmiennych, które odpowiednio determinują ryzyko ubezpieczeniowe. Husnjak, Peraković, Forenbacher i Mumdziev (2015) stwierdzają, że analiza zachowania kierowcy może dać firmie ubezpieczeniowej przewagę konkurencyjną poprzez zaproponowanie różnych stawek ubezpieczenia dla kierowców, którzy nie należą do jednego segmentu ryzyka ubezpieczeniowego. Drugim sposobem jest zwiększenie mocy predykcyjnej modeli taryfikacyjnych poprzez poprawę precyzji oszacowania estymatorów, o czym mowa w niniejszej pracy.

2. Przedmiot oraz zakres

Powszechnym narzędziem wykorzystywanym w naukach aktuarialnych do modelowania liczby szkód oraz wysokości średniej szkody są uogólnione modele liniowe z funkcją łącznikową, kolejno, o rozkładzie Poissona oraz gamma. Pomimo znacznej popularności owej klasy modeli, w literaturze opisane są problemy przy ich zastosowaniu. Po pierwsze pod wątpliwość można poddać fakt, czy modelując za ich pomocą dane panelowe, zawierające obserwacje tego samego kierowcy w kolejnych latach polisowych, spełnione jest jedno z założeń uogólnionych modeli liniowych mówiące o niezależności rozkładów zmiennych. W 1999 roku Frees, Yound i Luo stwierdzili, że dane ubezpieczeniowe ze względu na przedstawiony problem powinny być opisywane za pomocą liniowych modeli mieszanych. Podobny wniosek zaprezentowali w 2007 roku Antonio i Beirlant, podkreślając, że uogólnione modele liniowe (GLM) są nieodpowiednie do analizy tej samej obserwacji w kolejnych latach i właściwszym byłoby wykorzystanie uogólnionych liniowych modeli mieszanych (GLMM). Podobny wniosek, także w kontekście ubezpieczeniowym, przedstawiają w swojej pracy Antonio, Frees i Valdez (2010) tym razem wskazując na możliwość zastosowania w taryfikacji wielowymiarowych modeli GLMM, grupując poprzez efekty losowe obserwacje w ramach

² Złe (dobre) ryzyko definiowane jest w praktyce gospodarczej jako polisa charakteryzująca się ponadprzeciętnie wysokim (niskim) prawdopodobieństwem wystąpienia szkody.

pojazdu zagnieżdżonego w wymiarze właściciela pojazdu (floty), zagnieżdżonych następnie w obrębie firmy ubezpieczeniowej.

W innym wypadku można poddać pod wątpliwość niezależność rozkładów zmiennych dla obserwacji reprezentowanych przez tego samego pośrednika ubezpieczeniowego. Problem ten został przedstawiony przez autora niniejszej rozprawy (Paska, 2018). Uzasadnione jest to przede wszystkim faktem, że agent, jako pośredniczący pomiędzy zakładem ubezpieczeniowym, a klientem, dostarcza pierwszemu z wymienionych niezbędnych do wyliczenia składki informacji i to od jego docieklowości, skrupulatności i prawdomówności w znacznej mierze zależy jakość danych. Nie bez znaczenia pozostaje także kwestia osobistej relacji klient-agent, dzięki której agent może w pewnym stopniu ocenić ryzyko ubezpieczeniowe, a jego relacje z zakładem ubezpieczeń (m.in. w formie kontraktów sprzedażowych) mogą mieć decydujący wpływ na to, do której firmy jaki typ ryzyka agent skieruje.

Innym przykładem są wnioski z publikacji Wolny-Dominiak (2014), gdzie autorka wskazuje, że możliwość kolizji dwóch pojazdów poruszających się po tym samym regionie powoduje niespełnienie założenia niezależności m.in. ze względu na fakt, że pojazdy te są narażone na to samo prawdopodobieństwo złych warunków atmosferycznych, co z kolei zwiększa prawdopodobieństwo kolizji. Portfel ryzyk ma wtedy strukturę klastrową, gdzie klastery są zbiorami polis z tego samego regionu geograficznego. W opinii autorki to z kolei wymusza modelowanie zależności m.in. poprzez stosowanie w modelach czynników losowych ryzyka poprzez przyjęcie wymiaru geograficznego jako efektu losowego.

Kolejnym przykładem, zaproponowanym nowatorsko w pracy, jest wymiar menadżera agenta ubezpieczeniowego, zagnieżdżonego z wymiarem agenta. Uzasadnieniem wyboru owego wymiaru jako efektu losowego jest po pierwsze fakt, że dobry menadżer jest w stanie odpowiednio zarządzać siecią sprzedaży, minimalizując przytoczone wcześniej ryzyko wpływu agenta na szkodowość. Po drugie, ponieważ menadżerowie agentów działają lokalnie (na przykład na terenie danego powiatu) wymiar ten realizuje podobne cele, co zaproponowany przez Wolny-Dominiak (2014) wymiar geograficzny.

Kolejnym efektem losowym zaproponowanym w pracy jest marko-model ubezpieczonego pojazdu. Motywowane jest to faktem, po pierwsze, że marki i modele są często powiązane z charakterami wybierających ich osób, a co za tym idzie ich stylem jazdy i liczbą

wypadków. Po drugie, poszczególne marki i modele są w różnej częstotliwości kradzione, co wpływa na ryzyko Autocasco (por. Wolny-Dominiak, 2014). Po trzecie, niektóre marki kładą duży nacisk na systemy bezpieczeństwa w swoich pojazdach, a zatem poruszanie się nimi powinno przekładać się na niższą szansę spowodowania kolizji (ze względu na systemy wspierające bezpieczną jazdę) lub niższą szkodę w sytuacji, gdy wypadek się wydarzy (ze względu na lepsze systemy zabezpieczeń wypadkowych). Jednocześnie używanie tego wymiaru w charakterze efektu stałego może być utrudnione zważywszy na fakt niskiej liczebności poszczególnych kategorii zmiennej.

Ostatnim poddanym w pracy analizie przykładem efektu losowego jest wymiar klienta, opisany przez Antonio i Valdeza (2012). Autorzy zauważają, że nieobserwowalne charakterystyki kierowcy lub właściciela pojazdu mogą wpływać na wyniki szkodowe pojazdu w różnych latach. Przykładem takich nieobserwowalnych, a mających wpływ na liczbę szkód charakterystyk są: liczba godzin pracy pojazdów oraz czas przewozu, przeglądy techniczne pojazdów czy też instrukcje załadunków dla pojazdów transportowych.

Rozwiązaniem przytoczonych problemów są uogólnione liniowe modele mieszane, które rozszerzają klasyczne uogólnione modele liniowe o mające niezerowe wariancje efekty losowe, zdefiniowane w liniowym predyktorze, przy założeniu nieobserwowalnej wprost heterogeniczności w obrębie niektórych współczynników regresji. Zastosowania w taryfikacji ubezpieczeń modeli GLMM znaleźć można w pracach Garrido i Zhou (2009), Antonio i Valdeza (2012), czy też Baumgartnera, Grubera i Czado (2015). Modele te są także szeroko stosowane w naukach biologicznych (Bolker et al. 2009 za: Nakagawa i Schielzeth, 2013) oraz medycznych (Gelman i Hill, 2007, Congdon, 2010, Snijders i Bosker, 2011 za: Nakagawa i Schielzeth, 2013).

Kolejnym problemem występującym wspólnie w szeroko stosowanych w aktuariacie modelach GLM z rozkładem Poissona jest problem nadmiernego rozproszenia. Zauważone zostało to już w 1988 roku przez Dionne i Vanasse, a w 1989 roku opisane także przez McCullagh oraz Nelder, którzy stwierdzili wręcz, że zjawisko to występuje bardzo często i to ono de facto powinno być uważane za normalne, natomiast jego brak za sytuację wyjątkową. Frees w 2010 roku opisał, że z powodu koncentracji obserwacji zmiennej objaśnianej w zerze (tj. obserwacji bezszkodowych), wariancja musi przewyższać wartość oczekiwaną. W 2015 roku David i Jemna wykazali na przykładzie francuskiego rynku

ubezpieczeń, że parametr dyspersji w modelu częstości szkód jest istotny statystycznie, co wskazuje na słuszność wykorzystania rozkładu ujemnego dwumianowego. Podobne wnioski przedstawili w swoich pracach także Antonio i Valdez (2012) oraz Shi i Valdez (2014).

Dodatkowym problemem charakteryzującym dane ubezpieczeniowe jest, w przypadku analizowania częstości szkód, nadmierna liczba zer w rozkładzie zmiennej zależnej. Z tego względu zarówno szeroko opisywane i wykorzystywane w literaturze aktuarialnej uogólnione modele liniowe oraz ich modyfikacje wprowadzające tylko korektę ze względu na efekt losowy, mogą okazać się niewystarczające. Yip i Yau (2005), wprowadzając modele zero-inflated, podkreślili, że modele GLM mogą być modyfikowane w sytuacji występowania zbyt dużej liczby zer w rozkładzie zmiennej objaśnianej, aby uniknąć efektu nadmiernego rozproszenia. Antonio i Valdez (2012) udowodnili, że modele te, tj. zaproponowane przez Yip i Yau, są lepsze względem modeli „nie ZI”.

Grize, Fischer oraz Luetzelschwab (2020) zauważają, że obecnie modele GLM oraz GLMM próbuje się zastępować innymi technikami uczenia maszynowego, takimi jak sieci neuronowe, lasy losowe czy też modele XGB (ang. *extreme gradient boosting*). Jak podkreślają jednak Akinyemi oraz Leiser (2020) metody te są wciąż rzadziej wykorzystywane ze względu na brak transparentności oraz interpretowalności tych metod względem klasycznych modeli GLM oraz ich modyfikacji w postaci modeli GLMM. Także Goldburd, Khare, Tevet oraz Guller (2020) podkreślają, że modele GLM i ich rozszerzenia dają równie satysfakcjonujące wyniki odzwierciedlenia ryzyka, jak i inne zaawansowane modele statystyczne. Ich przewagą jest natomiast wysoka interpretowalność wyników wymagana często m.in. przez państwowe organy nadzoru ubezpieczycieli. Wniosek ten potwierdzają w swojej pracy także Henckaerts, Antonio, Clijsters i Verbelen (2018).

Przedmiotem pracy jest porównanie stosowanych w ramach taryfikacji ubezpieczeń majątkowych podejść do predykcji częstości szkód AC. Zagadnienie to jest kluczowym elementem działalności podstawowej zakładu ubezpieczeń, gdyż według informacji Komisji Nadzoru Finansowego (2021) produkt Autocasco odpowiada za około 26% wyniku technicznego zakładów ubezpieczeń majątkowych. Wynik techniczny tworzony jest natomiast w głównej mierze na podstawie stosowanej taryfy składek, a przyjęta metoda szacowania częstości szkód przyczynić może się do adekwatniejszego oszacowania składki ubezpieczeniowej i w konsekwencji zwiększenia zysku zakładu oraz ochrony

przed antyselekcją ryzyka. W pracy przedstawiono przegląd literatury aktuarialnej dotyczącej metod wykorzystywanych do tego celu.

3. Struktura pracy

Praca składa się z pięciu rozdziałów: w pierwszym przedstawiony jest zarys historii ubezpieczeń z podkreśleniem kontekstu taryfikacyjnego. Historia dotyczy czasów od starożytnych aż do współczesnych i opisuje między innymi historię pożyczki morskiej, koasekuracji, reasekuracji, ubezpieczeń morskich, czy też pośrednictwa ubezpieczeniowego. Przede wszystkim jednak opisana jest historia ubezpieczeń komunikacyjnych – ryzyka odpowiedzialności cywilnej pojazdów i kwestii prawnych obowiązkowości tego ubezpieczenia, a także ubezpieczeń następstw nieszczęśliwych wypadków oraz Autocasco. Opisano również historię międzynarodowych, ubezpieczeniowych porozumień tworzących warunki do swobodnego przemieszczania się pojazdów lądowych pomiędzy krajami – od pierwszego Poolu Nordyckiego aż do powstania systemu Zielonej Karty. W dalszej kolejności szeroko opisano historię taryfikacji ubezpieczeń – od prac dotyczących wyliczania prawdopodobieństwa zajścia zdarzenia, poprzez modele liniowe aż do rozwoju uogólnionych modeli liniowych i ich modyfikacji. Następnie opisano ograniczenia uogólnionych modeli liniowych w ubezpieczeniach (m.in. problem nadmiernego rozproszenia, czy też zbyt dużej liczby zerowych realizacji zmiennej zależnej). Zwrócono również uwagę na występujący w ubezpieczeniach panelowy charakter danych oraz na wymiary grupujące obserwacje. Dalej omówiono literaturę dotyczącą uogólnionych liniowych modeli mieszanych celem przedstawienia metod rozwiązania przytoczonych problemów. Podsumowano konsekwencje niestosowania uogólnionych liniowych modeli mieszanych, a także zalety tej klasy modeli względem innych technik uczenia maszynowego. W rozdziale tym opisano też spotykane w literaturze determinanty częstości szkód komunikacyjnych oraz przedstawiono dyskusję dotyczącą zasadności stosowania w taryfikacji niektórych cech właściciela przedmiotu ubezpieczenia – przede wszystkim płci. Wskazano również na obecny stan prawny dotyczącego owego zagadnienia. Dalej poruszono kwestię kluczowości taryfikacji ubezpieczeń w kontekście transferu ryzyka, heterogeniczności posiadanego portfela ryzyk, selekcji negatywnej oraz skutków finansowych dla zakładu ubezpieczeń.

W rozdziale drugim przedstawiono przegląd literatury aktuarialnej dotyczącej analizy częstości szkód wraz z podkreśleniem problemów występujących w tej analizie, a także propozycji autora rozprawy na ich rozwiązanie. Opisano miary opisujące zyskowność, strukturę i wielkość portfela firmy ubezpieczeniowej, takie jak składkę przypisaną, zarobioną oraz za ryzyko (inaczej: netto / czystą), częstość szkód, a także współczynnik szkodowy, kosztowy oraz mieszany. Zdefiniowano również pojęcie wyniku technicznego oraz komponentów szkodowych IBNR oraz IBNER. Miary te wykorzystano w pracy celem opisu skutków finansowych wprowadzenia proponowanych przez autora pracy modyfikacji modeli taryfikacyjnych. W rozdziale tym opisano też rozkłady prawdopodobieństw wykorzystywanych do analizowania częstości szkód – Poissona oraz ujemnego dwumianowego. Przedstawiono funkcje prawdopodobieństwa rozkładów, a także ich wartości oczekiwane, wariancje, cechy oraz przykładowe kształtowanie się w zależności od przyjętych arbitralnie wartości parametrów. W dalszej części przeanalizowano wykorzystywane w taryfikacji ubezpieczeń uogólnione modele liniowe – ich założenia, ograniczenia, problemy, metody estymacji, a także rozszerzenia, takie jak modele zero-inflated, hurdle models, uogólnione liniowe modele mieszane (zarówno hierarchiczne, jak i jedno i wielowymiarowe, zwane również zagnieżdżonymi). Wprowadzono pojęcie efektu stałego oraz losowego wraz ze wskazaniem aktuarialnych przykładów tych efektów – czasu, geografii oraz klienta. Ponadto w badaniu empirycznym oceniono zaproponowane przez autora efekty losowe: agenta ubezpieczeniowego, menadżera agenta oraz marko-modelu ubezpieczonego pojazdu. Jako wartość dodaną w kontekście zastosowań aktuarialnych należy uznać także użycie dwóch zagnieżdżonych efektów losowych – zarówno agenta ubezpieczeniowego jak i jego menadżera. W szczególności przeanalizowano jakościowo oraz empirycznie przedstawiony przez autora uogólniony liniowy model mieszany zero-inflated z rozkładem ujemnym dwumianowym.

W rozdziale trzecim przedstawiono charakterystykę danych wykorzystanych w badaniu empirycznym. Dane te opisują 99,7 tysięcy rocznych polis komunikacyjnych – informacje dotyczą zarówno przedmiotu ubezpieczenia (tj. cech pojazdu) oraz cech osoby będącej jego właścicielem, jak i zdarzeń szkodowych (między innymi ich liczby oraz wysokości wypłat). Wykorzystanie danych z polskiej firmy ubezpieczeniowej pozwoliło również na wykazanie w pracy przewagi proponowanych metod na przykładzie polskiego rynku ubezpieczeniowego. W rozdziale tym przedstawiono histogramy oraz statystyki opisowe zmiennych. Podkreślono

specyfikę i konsekwencje wynikające z faktu, że polisy zawarte w zbiorze zostały zakupione w salonach sprzedaży samochodów – przede wszystkim duże nasycenie produktem AC oraz niższy rok eksploatacji ubezpieczanego pojazdu względem średniej w populacji. Następnie w tym samym rozdziale przedstawiono metodę badawczą, stosowaną w badaniu analizy częstości szkód AC w rozdziale czwartym – opisano w szczególności analizowane w pracy modele, determinanty częstości szkód AC, proponowane efekty losowe oraz miary za pomocą których został wybrany optymalny model taryfikacyjny. Wskazano też testowane w badaniu empirycznym subportfele, których wyodrębnienie ma na celu wskazanie optymalnych obszarów dla wykorzystania nowych technik i metod. Opisano również metodę i założenia symulacji finansowej wykorzystanej w pracy, mającej na celu zobrazowanie skutków finansowych wdrożenia proponowanych rozwiązań.

W rozdziale czwartym przedstawiono i porównano dla ryzyka AC wyniki oszacowania parametrów oraz kryteriów informacyjnych dla:

- 1) uogólnionych modeli liniowych (GLM)
 - a) z rozkładem Poissona oraz ujemnym dwumianowym,
 - b) zero-inflated z rozkładem Poissona oraz ujemnym dwumianowym,
- 2) uogólnionych liniowych modeli mieszanych (GLMM) z efektem losowym agenta oraz
 - a) z rozkładem Poissona i ujemnym dwumianowym,
 - b) zero-inflated z rozkładem Poissona i ujemnym dwumianowym.

Na podstawie kryteriów informacyjnych AIC oraz BIC uznano zaproponowany przez autora pracy uogólniony liniowy model mieszany zero-inflated z założonym rozkładem ujemnym dwumianowym (ZINB GLMM AGENT) za najlepszy. W kolejnym kroku model ten porównano z modelami będącym przedstawicielem tej samej klasy metod, ale mającymi inne efekty losowe (tj. czasu oraz marko-modelu). Celem był wybór optymalnego z nich, a okazał się nim ostatecznie zaproponowany przez autora wymiar agenta ubezpieczeniowego. Następnie porównano model jednowymiarowy z zagnieżdżonym modelem GLMM zarówno z efektem losowym agenta ubezpieczeniowego, jak i menadżera agenta. Przedstawiono także wyniki symulacji finansowej, obrazującej kluczowość stosowania zaproponowanego przez autora modelu w taryfikacji ubezpieczeń majątkowych. Następnie przeprowadzono analizę wrażliwości – sprawdzono, jak zmiana kosztów stałych i zmiennych uwzględnionych w składce wpłynie na konkurencyjność oferty zakładu ubezpieczeń i wyniki finansowe, które

zakład ubezpieczeń generuje. Celem dodatkowej weryfikacji modelu porównano wyniki dopasowania krzywych optymalnego modelu częstości szkód z wielkościami empirycznymi, a następnie błędów MSE oraz RMSE na zbiorze walidacyjnym. W kolejnym kroku dokonano analizy zastosowania optymalnego modelu częstości szkód na 12 subportfelach, a wyniki porównano za pomocą współczynnika determinacji R^2 Nakagawa i Schielzeth (2013) dla modeli mieszanych:

- subportfel ryzyk nowo biznesowych,
- subportfel ryzyk wznawieniowych,
- subportfel pojazdów do 4 roku eksploatacji włącznie,
- subportfel pojazdów powyżej 4 roku eksploatacji,
- subportfel pojazdów zarejestrowanych w obszarach miejskich,
- subportfel pojazdów zarejestrowanych w obszarach wiejskich oraz miejsko-wiejskich,
- subportfel pojazdów zarejestrowanych w miastach wojewódzkich,
- subportfel pojazdów zarejestrowanych w innych miejscach niż miasta wojewódzkie,
- subportfel klientów mających w czasie 4 lat poprzedzających zawarcie ubezpieczenia szkodę w ryzyku AC,
- subportfel klientów bezszkodowych w czasie 4 lat poprzedzających zawarcie ubezpieczenia w ryzyku AC,
- subportfel pojazdów firmowych oraz leasingowych,
- subportfel pojazdów zarówno niebędących własnością firmy jak i leasingu.

Wykazano, że dzięki użyciu modelu ZINB GLMM AGENT dla portfela pojazdów zarejestrowanych w mieście zakład ubezpieczeń jeszcze bardziej może zwiększać zyski z wdrożenia proponowanych w pracy rozwiązań.

Rozdział piąty stanowi podsumowanie pracy: weryfikacji poddane są hipotezy badawcze, a także przedstawione są wnioski z badania oraz dalsze kierunki rozwoju.

4. Cel pracy

Znalezienie optymalnej klasy modeli jest ważne ze względu na fakt, że oszacowania modeli częstości szkód są składową wykorzystywaną przez zakłady ubezpieczeń do oszacowania składki za ryzyko ubezpieczeniowe. Celem pracy było zaproponowanie

metody szacowania częstości szkód rozwiązującej spotykane w literaturze problemy analizy danych ubezpieczeniowych, co z kolei przekłada się na lepsze dopasowanie oszacowań parametrów modelu do danych empirycznych. Wpływa to następnie na możliwość maksymalizacji zysku osiąganego przez zakład ubezpieczeń z jego działalności podstawowej.

W pracy weryfikowano następujące hipotezy badawcze:

1. Wartości błędów predykcji częstości szkód AC wyliczone na podstawie uogólnionych liniowych modeli mieszanych (ang. *generalized linear mixed model*, GLMM) są co do wartości niższe niż wyliczone na podstawie stosowanych obecnie w taryfikacji ubezpieczeń modeli GLM.
2. Ze względu na charakterystykę danych ubezpieczeniowych, tj. duży udział obserwacji bezszkodowych oraz nadmierną dyspersję, wprowadzony przez autora do taryfikacji ubezpieczeń model ZINB GLMM (ang. *zero-inflated negative binomial generalized linear mixed model*) jest lepiej dopasowany do empirycznych danych szkodowych niż inne modele GLMM.
3. Wprowadzenia do taryfikacji ubezpieczeń modeli z efektem losowym agenta ubezpieczeniowego oraz, osobno, efektem losowym marko-modelu umożliwia lepsze dopasowanie parametrów modeli do empirycznych danych szkodowych w porównaniu z szeroko opisywanymi w literaturze modelami z efektem losowym czasu.
4. Model GLMM z zagnieżdżonym efektem losowym (inaczej model wielopoziomowy) osiągnie niższe wartości błędów w porównaniu z modelem jednopoziomowym

Hipoteza nr 1 została potwierdzona – wykazano bowiem, że wartości kryteriów informacyjnych AIC oraz BIC są niższe dla uogólnionych liniowych modeli mieszanych względem uogólnionych modeli liniowych.

Hipoteza nr 2 również potwierdzona została za pomocą wartości kryteriów informacyjnych AIC oraz BIC. Miara ta bowiem była najniższa dla modelu ZINB GLMM. Sprawdzono także, że model ZINB GLMM AGENT nie jest przeuczony oraz że jest dobrze dopasowany do empirycznych danych. Model ten osiąga najwyższą wartość współczynnika determinacji R^2 opisując w największym stopniu kształtowanie się wariancji zmiennej zależnej. Udowodniono także, że jego stosowanie przekłada się na wyższe wyniki finansowe względem stosowania innych modeli taryfikacyjnych.

Hipoteza nr 3 również została potwierdzona i także, zgodnie z metodą przyjętą w literaturze, w oparciu o kryteria informacyjne AIC oraz BIC. Model z efektem losowym agenta osiągnął najniższe wartości kryteriów informacyjnych, a drugi w kolejności był model z efektem losowym marko-modelu ubezpieczonego pojazdu. Najwyższe wartości AIC oraz BIC osiągnął w badaniu empirycznym model z szeroko opisanym w literaturze aktuarialnej efektem losowym czasu.

Hipoteza nr 4 nie została potwierdzona – jednopoziomowy uogólniony liniowy model mieszany lepiej dopasowuje się do danych dla analizy częstości szkód Autocasco niż model wielopoziomowy z zagnieżdżonym efektem losowym agenta oraz menadżera agenta.

Podsumowując, zaproponowane przez autora pracy rozwiązania są elementem rozwoju opisanej literatury. Cel pracy, jakim było zaproponowanie metody szacowania częstości szkód rozwiązującej spotykane w literaturze problemy analizy danych ubezpieczeniowych oraz polepszający dopasowanie oszacowań parametrów modelu do danych empirycznych, został zrealizowany. Dzięki temu oszacować można adekwatniejszą składkę netto (względem ryzyka), co pozwoli zakładowi ubezpieczeń osiągnąć przewagę na rynku poprzez zaproponowanie składki maksymalizującej jego zyski oraz wolumeny sprzedaży.

Bibliografia autoreferatu

- Akinyemi, K., Leiser, B. (2020). The Use of Advanced Predictive Analytics for Rate Making in Insurance, *Actuarial technology today*, s. 1-4.
- Antonio, K., Beirlant, J. (2007). Actuarial statistics with generalized linear mixed models, *Insurance Mathematics and Economics*, vol. 40, s. 58-76.
- Antonio, K., Valdez, E. A. (2012). Statistical concepts of a priori and a posteriori risk classification in insurance, *ASTA Advances in Statistical Analysis*, vol. 96, s. 187-224.
- Antonio, K., Frees, E. W., Valdez E. A. (2010). A multilevel analysis of intercompany claim counts, *Astin Bulletin*, vol. 40(1), s. 151-177.
- Baumgartner, C., Gruber, L. F., Czado, C. (2015). Bayesian total loss estimation using shared random effects, *Insurance: Mathematics and Economics*, vol. 62, s. 194-201.
- David, M., Jemna, D. (2015). Modeling the frequency of auto insurance claims by means of Poisson and negative binomial models, *Annals of the Alexandru Ioan Cuza University-Economics*, vol. 62(2), s. 151-168.
- Dionne, G., Vanesse C. (1988). A generalization of automobile insurance rating models: the negative binomial distribution with a regression component, *Astin Bylletin*, vol. 19(2), s. 199-212.
- Frees, E. W. (2010). *Regression modeling with actuarial and financial application*. Cambridge: Cambridge University Press.
- Frees, E. W., Yound V. R., Luo Y. (1999). A longitudinal data analysis interpretation of credibility models, *Insurance: Mathematics and Economics*, vol. 24, s. 229-247.
- Garrido, J., Zhou, J. (2009). Full credibility with generalized linear and mixed models, *Astin Bulletin*, vol. 39(1), s. 61-80.
- Goldburd, M., Khare, A., Tevet, D., Guller, D. (2020). *Generalized linear models for insurance rating*. Virginia: Casualty Actuarial Society.
- Grize, Y., Fischer, W., Lueztelschwab, C. (2020). Machine learning applications in nonlife insurance, *Appl Stochastic Models Bus Ind.*, vol. 36, s. 523-537.

- Henckaerts, R., Antonio, K., Clijsters, M., Verbelen, R. (2018). A data driven binning strategy for the construction of insurance tariff classes, *Scandinavian Actuarial Journal*, vol. 2018(8), s. 681-705.
- Husnjak, S., Peraković, D., Forenbacher, I., Mumdziev, M. (2015). Telematics System in Usage Based Motor Insurance, *Procedia Engineering*, vol. 100, s. 816-825.
- Komisja Nadzoru Finansowego. (2020). *Raport o sektora ubezpieczeń po III kwartałach 2019 roku*. Warszawa: Urząd Komisji Nadzoru Finansowego.
- Komisja Nadzoru Finansowego. (2021). *Raport o sektora ubezpieczeń po III kwartałach 2020 roku*. Warszawa: Urząd Komisji Nadzoru Finansowego.
- McCullagh, P., Nelder, J. A. (1989). *Generalized Linear Models*. 2nd ed. Londyn: Chapman and Hall.
- Nakagawa, S., Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models, *Methods in ecology and evolution*, vol. 4(2), s. 133-142.
- Paska, N. (2018). Zastosowanie modeli ZINB GLMM z efektem losowym agenta w taryfikacji ubezpieczeń majątkowych, *Roczniki Kolegium Analiz Ekonomicznych SGH*, vol. 53, s. 63-76.
- Shi, P., Valdez, E. A. (2014). Multivariate negative binomial models for insurance claim counts, *Insurance: Mathematics and Economics*, vol. 55, s. 18-29.
- Sulewski, P., Meuwissen, M. (2014). Fundusze ubezpieczeń wzajemnych jako forma ograniczania ryzyka w rolnictwie, *Zagadnienia ekonomiki rolnej*, vol. 339(2), s. 127-143.
- Wolny-Dominiak, A. (2014). *Taryfikacja w ubezpieczeniach majątkowych z wykorzystaniem modeli mieszanych*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.
- Yip, K C. H., Yau, K. K. W. (2005). On modeling claim frequency data in general insurance with extra zeros, *Insurance Mathematics and Economics*, vol. 36, s. 153-163.