

KAMIL GALA, KAROLINA KOLAK

Ubezpieczeniowy Fundusz Gwarancyjny

Wykorzystanie informacji kredytowej w procesie oceny ryzyka ubezpieczeniowego w ubezpieczeniach komunikacyjnych

Streszczenie

W pracy przedstawiono teoretyczne i praktyczne aspekty wykorzystania informacji kredytowej do oceny ryzyka ubezpieczeniowego. Omówione zostały potencjalne źródła związku między zjawiskami kredytowymi i ubezpieczeniowymi oraz metody jego modelowania. Dokonano przeglądu rozwiązań stosowanych w praktyce oraz wniosków z badań empirycznych opisanych w literaturze przedmiotu. W pracy przedstawiono również założenia wspólnego projektu badawczego Ubezpieczeniowego Funduszu Gwarancyjnego oraz Biura Informacji Kredytowej SA, mającego na celu weryfikację związku między historią kredytową klienta a związanym z nim ryzykiem ubezpieczeniowym na polskim rynku ubezpieczeń komunikacyjnych. Uzyskane wyniki wskazują na potencjalne korzyści płynące ze współpracy między sektorem ubezpieczeniowym i bankowym w obszarze oceny ryzyka ubezpieczeniowego.

Słowa kluczowe: ryzyko ubezpieczeniowe, taryfikacja *a priori*, dane kredytowe, uogólnione modele liniowe, ubezpieczenia komunikacyjne

1. Wstęp

Zagadnienie oceny ryzyka ubezpieczeniowego i kalkulacji składki jest bardzo istotne w kontekście gospodarki finansowej zakładu ubezpieczeń. Ustalenie składki na zbyt niskim poziomie może zagrozić wypłacalności ubezpieczyciela, natomiast zbyt wysoki poziom składki może obniżyć poziom jego konkurencyjności na rynku¹. W związku z tym zgromadzenie odpowiednich danych statystycznych i wykorzystanie do oceny ryzyka właściwych metod aktuarialnych może dać zakładowi ubezpieczeń przewagę konkurencyjną. Jedną z możliwości

¹ *Metody aktuarialne. Zastosowania matematyki w ubezpieczeniach*, red. W. Ronka-Chmielowiec, Wydawnictwo Naukowe PWN, Warszawa 2013, s. 79.

rozwoju w tym obszarze jest wykorzystanie informacji pochodzącej z rejestrów kredytowych.

Celem niniejszej pracy jest próba odpowiedzi na pytanie, czy wykorzystanie informacji kredytowej może być użyteczne w procesie oceny ryzyka ubezpieczeniowego. Praca swoim zakresem obejmuje omówienie aspektów teoretycznych i praktycznych związanych z tym zagadnieniem, jak również przedstawienie wyników badania empirycznego przeprowadzonego w 2014 r. wspólnie przez Ubezpieczeniowy Fundusz Gwarancyjny i Biuro Informacji Kredytowej SA.

W punkcie drugim pracy omówiono proces oceny ryzyka w ubezpieczeniach komunikacyjnych. W punkcie kolejnym zaprezentowano kwestie dotyczące wykorzystania w tym procesie informacji kredytowej, skupiając się na źródłach tego rodzaju praktyk, wynikach badań empirycznych oraz korzyściach i zagrożeniach z tym związanych. W punkcie czwartym zostały omówione uogólnione modele liniowe, należące do grupy aktuarialnych metod taryfikacji *a priori*. Punkt piąty przedstawia założenia i wnioski ze wspomnianego wcześniej badania. Pracę kończy podsumowanie.

2. Ocena ryzyka w ubezpieczeniach komunikacyjnych

Cechą charakterystyczną działalności ubezpieczeniowej jest tzw. odwrócony cykl produkcyjny, który oznacza, że cena ochrony ubezpieczeniowej (składka) jest ustalana przed zawarciem umowy ubezpieczenia, kiedy całkowity koszt udzielenia tej ochrony (m.in. w postaci wypłaconych świadczeń i kosztów likwidacji szkód) nie jest znany. Konsekwencją tego faktu jest konieczność stosowania w zakładzie ubezpieczeń specyficznych rozwiązań, m.in. w obszarze kalkulacji aktuarialnych. W szczególności istotnym zagadnieniem jest problem odpowiedniego wyznaczenia składki. Zgodnie z art. 18 ustawy z dnia 22 maja 2003 r. o działalności ubezpieczeniowej (Dz. U. z 2003 r. Nr 124, poz. 1151 z późn. zm.), wysokość składki ubezpieczeniowej jest ustalana przez zakład ubezpieczeń po dokonaniu oceny ryzyka ubezpieczeniowego. Składka ta powinna być ustalona w wysokości, która powinna co najmniej zapewnić wykonanie wszystkich zobowiązań wynikających z umów ubezpieczenia i pokrycie kosztów wykonywania działalności ubezpieczeniowej zakładu ubezpieczeń, a zakład ubezpieczeń powinien gromadzić dane statystyczne na potrzeby jej wyznaczenia. Ogólną charakterystykę procesu oceny ryzyka i wyznaczania

składki można znaleźć m.in. w pracach pod redakcją W. Ronki-Chmielowiec² oraz W. Ostasiewicza³.

Niniejsza praca jest poświęcona ocenie ryzyka w ubezpieczeniach komunikacyjnych rozumianych jako ubezpieczenie odpowiedzialności cywilnej posiadaczy pojazdów mechanicznych (OC p.p.m.) oraz autocasco (AC), stanowiących, odpowiednio, podgrupę w ramach grupy 10 i grupę 3 działu II zgodnie z załącznikiem do ustawy o działalności ubezpieczeniowej. Wyznaczenie składki indywidualnej w ubezpieczeniach komunikacyjnych na ogół odbywa się dwustopniowo. W pierwszym kroku (taryfikacja *a priori*) na podstawie obserwowalnych cech ubezpieczonego lub jego pojazdu (np. miejsce zamieszkania czy moc silnika) jest ustalana składka podstawowa. Celem tego etapu jest podział niejednorodnego portfela klientów na mniej zróżnicowane grupy taryfowe, w których liczbę oraz wysokość roszczeń można modelować za pomocą niezależnych zmiennych losowych o tych samych rozkładach⁴. W praktyce jednak utworzone w ten sposób grupy taryfowe cechują się niejednorodnością nawet po uwzględnieniu wielu zmiennych taryfowych, co wynika z asymetrii informacji między ubezpieczonym i ubezpieczycielem⁵. W związku z tym w drugim kroku składka podstawowa jest korygowana na podstawie historii szkodowej ubezpieczonego (taryfikacja *a posteriori*, ang. *experience rating*). Wynika to z przekonania, że wiedza o ukrytych czynnikach ryzyka (takich jak umiejętności kierowcy lub znajomość przepisów) jest odkrywana stopniowo przez liczbę szkód spowodowanych przez ubezpieczonego w kolejnych okresach⁶.

Istotną rolę w procesie oceny ryzyka w ubezpieczeniach komunikacyjnych odgrywają metody aktuarialne. Właściwa selekcja zmiennych taryfowych powinna uwzględniać ich istotność statystyczną, obserwowalność i mierzalność, a także wykluczać wszelkie możliwości manipulacji. Należy także rozważyć, które ze zmiennych są ze sobą powiązane i czy są skorelowane ze zmiennymi dotyczącymi historii ubezpieczenia, a także czy niosą odpowiedni ładunek informacyjny. Do realizacji tych celów zastosowanie znajdują modele ekonometryczne, w tym uogólnione modele liniowe. Aby utworzone w ten sposób grupy taryfowe były odpowiednio duże (by móc stosować centralne twierdzenie graniczne i prawo

² Ibidem.

³ *Składki i ryzyko ubezpieczeniowe. Modelowanie stochastyczne*, red. W. Ostasiewicz, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2004.

⁴ Ibidem, s. 226.

⁵ M. Denuit, X. Maréchal, S. Pitrebois, J. Walhin, *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, Wiley, New York 2007, s. 121.

⁶ Ibidem, s. 121.

wielkich liczb), a także mało zróżnicowane wewnętrznie i zróżnicowane między sobą, stosowane są metody analizy skupień⁷. Z kolei w taryfikacji *a posteriori* są wykorzystywane metody statystyki bayesowskiej oraz teorii zaufania (ang. *credibility theory*), natomiast najbardziej rozpowszechnioną formą taryfikacji *a posteriori* są systemy bonus-malus. Omówienie tych metod można znaleźć w pracy M. Denuita i innych⁸.

3. Wykorzystanie informacji kredytowej do oceny ryzyka w ubezpieczeniach komunikacyjnych

W poprzednim punkcie zostały przedstawione rozważania dotyczące oceny ryzyka w ubezpieczeniach komunikacyjnych. Cechą charakterystyczną tego procesu jest wykorzystanie wielu zmiennych taryfowych w postaci obserwowalnych cech pojazdu, ubezpieczonego lub jego historii szkodowej, a metody aktuarialne stosowane do tego celu sięgają połowy XX w. Z tego punktu widzenia wykorzystanie informacji kredytowej do oceny ryzyka w ubezpieczeniach komunikacyjnych jest praktyką stosunkowo nową, z którą wciąż wiąże się wiele niejasności i kontrowersji. Podstawowe pytania z tym związane to m.in.:

- Czy istnieje korelacja między historią kredytową ubezpieczonego a jego szkodowością?
- Jeśli taka korelacja istnieje, to które informacje kredytowe wskazują na zwiększony lub zmniejszony potencjał szkodowy ubezpieczonego i jakie jest ich znaczenie w porównaniu z pozostałymi danymi dostępnymi dla ubezpieczyciela?
- Czy taka korelacja nie jest pozorna, tj. czy nie wynika ona z korelacji historii kredytowej z innymi zmiennymi taryfowymi, które są już wykorzystywane?
- Czy można wskazać związki przyczynowo-skutkowe między zjawiskami kredytowymi i ubezpieczeniowymi?
- Jakie są uwarunkowania i konsekwencje wykorzystania danych kredytowych do oceny ryzyka ubezpieczeniowego?

W dalszej części tego punktu zostaną omówione wybrane kwestie związane ze wskazanymi wyżej zagadnieniami.

⁷ Składki i ryzyko ubezpieczeniowe..., op.cit., s. 231–234.

⁸ M. Denuit, X. Maréchal, S. Pitrebois, J. Walhin, op.cit.

Wykorzystanie informacji kredytowej do oceny ryzyka ubezpieczeniowego wzięło swój początek w USA w latach 80. XX w. i od tamtej pory jest szeroko stosowane w indywidualnych liniach ubezpieczeń, przede wszystkim w ubezpieczeniach komunikacyjnych, mieszkaniowych i życiowych⁹. Zasadność i skuteczność tego rozwiązania wielokrotnie były przedmiotem dyskusji zarówno w branży, jak i wśród regulatorów, a dyskusje te stanowiły inspirację do podjęcia wielu badań poświęconych identyfikacji oraz opisowi korelacji między historią kredytową klienta i jego szkodowością. Jednym z pierwszych było badanie przeprowadzone przez Tillinghast-Towers Perrin¹⁰. W badaniu tym została przeprowadzona jednokierunkowa (ang. *one-way*) analiza związku między scoringiem kredytowym i współczynnikiem szkodowości dla dziewięciu firm oferujących ubezpieczenia komunikacyjne i mieszkaniowe. Otrzymane wyniki wskazują na wyraźną korelację ujemną, tj. klienci o niskim scoringu kredytowym cechują się większą szkodowością. Istotnym ograniczeniem tego badania jest wspomniana wyżej jednokierunkowość, przez co obserwowana korelacja może wynikać ze związku analizowanych cech ze zmiennymi nieuwzględnionymi w analizie. Repertuar stosowanych metod został rozszerzony w kolejnych badaniach, których wyniki omówiono w pracach J.E. Monaghana¹¹, C.-S. Wu i J.C. Guszczu¹² oraz M. Millera i R.A. Smitha¹³. W pracach tych obok analizy jednokierunkowej zostały zastosowane również metody analizy wielowymiarowej, a w szczególności modele regresji, pozwalające uwzględnić korelacje między zmiennymi objaśniającymi. Ogólnie rzecz biorąc, można stwierdzić, że wyniki tych badań wskazują na to, że informacje kredytowe mogą być użytecznym predyktorem przyszłej szkodowości, także w obecności zmiennych objaśniających pochodzących z innych źródeł. W szczególności M. Miller i R.A. Smith wskazują, że w przypadku analizowanych produktów, w tym ubezpieczeń komunikacyjnych, scoring kredytowy jest jednym z najważniejszych czynników ryzyka, przy czym związek tej

⁹ C.-S. Wu, J.C. Guszczu, *Does Credit Score Really Explain Insurance Losses? – Multivariate Analysis from a Data Mining Point of View*, 2003 CAS Winter Forum, Casualty Actuarial Society, 2003, s. 116.

¹⁰ *Credit Reports and Insurance Underwriting*, NAIC White Papers, National Association of Insurance Commissioners, 1997.

¹¹ J.E. Monaghan, *The Impact of Personal Credit History on Loss Performance in Personal Lines*, 2000 CAS Winter Forum, Casualty Actuarial Society, 2000, s. 79–106.

¹² C.-S. Wu, J.C. Guszczu, op.cit.

¹³ M. Miller, R.A. Smith, *The Relationship of Credit-based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity*, Epic Actuaries LLC, 2003, http://www.ask-epic.com/Publications/Relationship%20of%20Credit%20Scores_062003.pdf (dostęp: 16.08.2014).

zmiennej ze szkodowością wydaje się silniejszy dla częstości szkód niż dla ich wysokości¹⁴. Warto w tym miejscu rozważyć podkreślić fakt, że wykorzystanie danych kredytowych przez ubezpieczycieli było również przedmiotem analiz na potrzeby nadzoru ubezpieczeniowego, np. w Teksasie¹⁵ i Michigan¹⁶. Wyniki tych analiz pozostają w zgodzie z wnioskami przedstawionymi we wspomnianych wcześniej pracach.

Wyniki omawianych badań empirycznych wskazują, że sytuacja kredytowa może być ważnym czynnikiem ryzyka w ubezpieczeniach indywidualnych. W związku z tym naturalne wydaje się pytanie, jakie są źródła korelacji między zjawiskami kredytowymi oraz ubezpieczeniowymi, a w szczególności – czy istnieje między nimi związek przyczynowo-skutkowy. O ile przyczynowość nie jest wymagana przy budowie modeli aktuarialnych, o tyle pozwala ona na lepszą interpretację i zrozumienie otrzymanych wyników¹⁷. W konsekwencji w literaturze przedmiotu można znaleźć różne próby wyjaśnienia związku między historią kredytową oraz szkodowością ubezpieczonego. Teorie te można zasadniczo podzielić na dwie grupy. Pierwszą grupę stanowią wyjaśnienia zakładające bezpośredni związek przyczynowy między sytuacją kredytową i szkodami. Wskazuje się, że klienci o złej sytuacji kredytowej mogą być bardziej skłonni do zgłaszania nawet niewielkich, ale względnie dla nich dotkliwych szkód, do zawyżania ich wysokości lub do zgłaszania szkód, które w rzeczywistości nie miały miejsca¹⁸. Uważa się również, że nadmierne obciążenie finansowe może być także źródłem stresu, który przekłada się potem na zachowanie na drodze, powodując np. obniżenie poziomu koncentracji kierowcy¹⁹. Przedstawiciele drugiej grupy teorii zakładają, że historia kredytowa niesie informację o tych

¹⁴ Ibidem, s. 39.

¹⁵ *Use of Credit Information by Insurers in Texas*, Report to the 79th Legislature, Texas Department of Insurance, 2004, <http://www.tdi.texas.gov/reports/documents/creditall04.pdf> (dostęp: 11.08.2014); *Use of Credit Information by Insurers in Texas: The Multivariate Analysis*, Report to the 79th Legislature, Texas Department of Insurance, 2005, <http://www.tdi.texas.gov/reports/documents/credit05sup.pdf> (dostęp: 11.08.2014).

¹⁶ F.M. Fitzgerald, *The Use of Insurance Credit Scoring In Automobile and Homeowners Insurance*, Office of Financial and Insurance Services, Michigan 2002, www.michigan.gov/documents/cis_ofis_credit_scoring_report_52885_7.pdf (dostęp: 16.08.2014).

¹⁷ *Risk Classification (for All Practice Areas)*, „Actuarial Standard of Practice” 2005, no. 12, Actuarial Standards Board, s. 4, http://www.actuarialstandardsboard.org/pdf/asops/asop012_132.pdf (dostęp: 14.08.2014).

¹⁸ J.E. Monaghan, op.cit., s. 86–88.

¹⁹ P. Brackett, L. Golden, *Biological and Psychobehavioral Correlates of Risk Scores and Automobile Insurance Losses: Toward an Explication of Why Credit Scoring Works*, „Journal of Risk and Insurance” 2007, vol. 74, issue 1, s. 23–63.

cechach kierowcy, które mogą mieć istotny wpływ na jego stosunek do ryzyka i zachowanie na drodze. Oznacza to, że osoba dobrze zarządzająca swoimi finansami może również odpowiedzialnie zachowywać się za kierownicą, a także bardziej dbać o stan techniczny swojego pojazdu²⁰.

Należy podkreślić fakt, że przedstawiona lista interpretacji związku między historią kredytową klienta i jego szkodowością nie jest wyczerpująca. W szczególności w literaturze dominują teorie wyjaśniające ujemną korelację między historią kredytową i liczbą szkód. Można jednak wyobrazić sobie sytuację, w której niekorzystna sytuacja kredytowa obniża ryzyko ubezpieczeniowe. Może się okazać, że presja finansowa stanowi w istocie czynnik motywujący do dbania o swój pojazd, gdyż jego naprawa stanowiłaby dodatkowe obciążenie – zarówno finansowe, jak i czasowe. Analogicznie sytuacja kredytowa może wpływać na intensywność użytkowania pojazdu. Osoba szukająca dodatkowych oszczędności może rzadziej używać pojazdu ze względu np. na wysokie koszty paliwa lub miejsca parkingowego. Rzadsze podróże samochodem przekładają się z kolei na mniejszą liczbę szkód komunikacyjnych. Jeśli natomiast niska ocena punktowa jest związana z dostępnością pewnych produktów kredytowych w mniej rozwiniętych regionach (np. terenach wiejskich), to można zaobserwować korelację dodatnią wynikającą z mniej intensywnego ruchu w tych regionach i związaną z tym mniejszego prawdopodobieństwa wypadku.

Powyższe rozważania wskazują, że mechanizm oddziaływania zjawisk kredytowych na ryzyko ubezpieczeniowe nie jest łatwy do interpretacji, a w szczególności jego kierunek może nie być jednoznaczny. Nie ulega jednak wątpliwości, że istnieje wyraźny statystyczny związek między informacją kredytową i ryzykiem ubezpieczeniowym, związku tego nie można wytłumaczyć za pomocą innych danych dostępnych w ocenie ryzyka²¹.

Na początku niniejszego punktu wspomniano, że wykorzystanie informacji kredytowej przez ubezpieczycieli budzi kontrowersje. O ile w świetle badań empirycznych przemawiają za tym względy statystyczne, o tyle istotną kwestią do rozważenia pozostają względy etyczne i społeczne. F.M. Fitzgerald²² wymienia wiele formułowanych przez konsumentów argumentów przeciwko wykorzystaniu informacji kredytowej przez ubezpieczycieli. Wśród zarzutów wskazywane są m.in.: niesprawiedliwość (przez karanie osób, które wpadły w kłopoty finansowe

²⁰ *Credit-Based Insurance Scores: Impacts on Consumers of Automobile Insurance*, A Report to Congress, Federal Trade Commission, 2007, s. 31, http://www.ftc.gov/os/2007/07/P044804FACTA_Report_Credit-Based_Insurance_Scores.pdf (dostęp: 15.08.2014).

²¹ C.-S. Wu, J.C. Guszcza, op.cit., s. 130.

²² F.M. Fitzgerald, op.cit., s. 14–15.

z nie swojej winy, np. z powodów zdrowotnych), niejawnym sposobie nadawania scoringu kredytowego (klient nie wie, co może zrobić, aby go poprawić), trudność w korygowaniu błędów w rejestrach kredytowych, dyskryminacja osób o niższych dochodach (które najbardziej ucierpią przez podniesienie składek), brak jasnego związku między historią kredytową i prawdopodobieństwem zgłoszenia szkody, a także naruszenie prywatności. Wskazuje się również, że klienci mogą rezygnować z ubezpieczenia, na które ich nie stać. Ma to szczególne znaczenie przy ubezpieczeniach obowiązkowych, takich jak OC p.p.m. Zwiększenie liczby nieubezpieczonych ma negatywne konsekwencje dla funkcjonowania systemu, np. przez zwiększenie kosztów świadczeń wypłacanych przez fundusze gwarancyjne (w przypadku Polski jest to UFG), ponoszonych częściowo przez nieubezpieczonych posiadaczy pojazdów, którzy ze względu na trudną sytuację finansową nie są w stanie spłacić roszczeń zwrotnych (regresowych) w całości.

Kontrowersje związane z wykorzystaniem informacji kredytowej wpływają na istotne zróżnicowanie w zakresie funkcjonujących praktyk. Canadian Council of Insurance Regulation podaje, że ok. 90% firm ubezpieczeniowych w Stanach Zjednoczonych wykorzystuje dane kredytowe do ustalania wysokości składki, czerpiąc dane na poziomie klienta bezpośrednio z rejestrów kredytowych²³. Jednocześnie poszczególne stany różnią się od siebie pod względem regulacji w tym zakresie. Około połowa z nich przyjęła regulacje dopuszczające wykorzystanie informacji kredytowej pod pewnymi warunkami (m.in. dotyczącymi informowania klienta i uzasadnienia aktuarialnego), natomiast niektóre stany całkowicie zabroniły tej praktyki w odniesieniu do ubezpieczeń komunikacyjnych. W kilku stanach wprowadzono dodatkowe obostrzenia, dotyczące np. odrębnego traktowania szczególnych przypadków życiowych klienta i możliwości wykorzystania danych kredytowych tylko do nowo zawieranych umów. W Kanadzie natomiast wykorzystanie informacji kredytowej podlega regulacjom dotyczącym ochrony danych osobowych i wymaga uprzedniej zgody klienta, któremu przysługuje również prawo do informacji na temat przetwarzanych danych oraz korekty błędnych danych. W Polsce – podobnie jak w innych krajach Europy – omawiana praktyka nie jest tak znana i popularna, o czym świadczy niewielka liczba publikacji na ten temat. Związane jest to z obowiązującymi przepisami dotyczącymi ochrony danych osobowych, a także regulacjami związanymi z tajemnicą ubezpieczeniową (art. 19 ustawy z dnia 22 maja 2003 r. o działalności

²³ *Use of Credit Scores by Insurers*, Issues Paper, Canadian Council of Insurance Regulation, 2011, [http://www.ccir-ccra.org/en/init/credit_scor/CCIR%20credit%20scores%20issues%20paper\(En\).pdf](http://www.ccir-ccra.org/en/init/credit_scor/CCIR%20credit%20scores%20issues%20paper(En).pdf) (dostęp: 14.08.2014).

ubezpieczeniowej, Dz. U. z 2003 r. Nr 124, poz. 1151) oraz tajemnicą bankową (art. 104 ustawy z dnia 29 sierpnia 1997 r. – Prawo bankowe, Dz. U. z 1997 r. Nr 140, poz. 939 z późn. zm.).

Podsumowując, należy stwierdzić, że wykorzystanie informacji kredytowej do oceny ryzyka ubezpieczeniowego jest interesującym, ale również złożonym zagadnieniem, które ma swoje zalety i wady. Z jednej strony, szerokie stosowanie scoringu kredytowego przez amerykańskich ubezpieczycieli oraz wyniki badań empirycznych wskazują na statystycznie istotną korelację między zjawiskami kredytowymi oraz ubezpieczeniowymi. Z drugiej strony, natura tego związku nie jest łatwa do interpretacji i do zakomunikowania klientowi. Wynika stąd, że wykorzystanie informacji kredytowej wymaga uważnego zaprojektowania procesu oceny ryzyka i zwrócenia w nim uwagi na relacje z klientem oraz wizerunek firmy. Z tego względu preferowaną formą wykorzystania danych kredytowych może być np. oferowanie niższej składki kierowcom z dobrą historią kredytową. Należy również zadbać o to, żeby cały proces był przejrzysty dla klienta – powinien on rozumieć to, jak jego historia kredytowa wpływa na składkę i co może zrobić, żeby ją poprawić.

4. Aktuarialne metody taryfikacji *a priori*

W poprzednich punktach zostało podkreślone znaczenie właściwego ustalania składki dla gospodarki finansowej zakładu ubezpieczeń i jego konkurencyjności. Nic więc dziwnego, że w literaturze aktuarialnej można znaleźć wiele metod służących do oceny ryzyka na podstawie obserwowalnych cech ubezpieczonego lub jego pojazdu. W tym punkcie zostaną omówione uogólnione modele liniowe, stanowiące ważną grupę aktuarialnych modeli statystycznych.

Uogólnione modele liniowe (ang. *Generalized Linear Models* – GLM) zostały wprowadzone w 1972 r. przez J.A. Neldera i R.M. Wedderburna²⁴. Jest to szeroka klasa modeli, obejmująca m.in. regresję liniową, regresję logistyczną i regresję Poissona. W modelach tych zostaje uchylone założenie, że zmienna zależna ma rozkład normalny, co przy zachowaniu ich łatwej interpretacji powoduje, że są one szczególnie użyteczne w analizie danych ubezpieczeniowych.

²⁴ J.A. Nelder, R.M. Wedderburn, *Generalized Linear Models*, „Journal of the Royal Statistical Society” 1972, vol. 135, no. 3, s. 370–384.

Uogólnione modele liniowe składają się z trzech podstawowych elementów: składnika losowego, składnika systematycznego oraz funkcji wiążącej (ang. *link function*)²⁵. Składnik losowy określa warunkowy rozkład zmiennej objaśnianej Y przy ustalonych wartościach zmiennych objaśniających. Zakłada się, że jest to rozkład z tzw. rodziny wykładniczej rozkładów, czyli rozkładów, których gęstość lub funkcję prawdopodobieństwa można zapisać w postaci

$$f_Y(y; \theta; \psi) = \exp\left(\frac{y\theta - b(\theta)}{\psi} + c(y; \psi)\right) \quad \text{dla } y \in D_\psi, \quad (1)$$

gdzie θ i ψ są parametrami, b oraz c ustalonymi funkcjami, a D_ψ jest nośnikiem rozkładu, być może zależnym od wartości parametru ψ . Taka postać rozkładu jest wygodna z teoretycznego i obliczeniowego punktu widzenia. Można np. pokazać, że

$$\mu = \mathbb{E}Y = b'(\theta), \quad \text{Var}(Y) = \psi \cdot b''(\theta).$$

Parametr θ jest nazywany parametrem kanonicznym i może być funkcją standardowych parametrów rozkładu. W szczególności w najważniejszych przypadkach jest on funkcją wartości oczekiwanej, tj. $\theta = g(\mu)$, co oznacza, że $\text{Var}(Y) = V(\mu)$ dla pewnej funkcji V , nazywanej funkcją wariancji. Dzięki temu GLM są użyteczne do modelowania populacji, dla której wariancja nie jest stała (tzw. heteroskedastyczność).

Do omawianej rodziny należy wiele popularnych rozkładów prawdopodobieństwa, m.in. rozkład normalny, rozkład Poissona (stosowany w ubezpieczeniach do modelowania liczby szkód) oraz rozkłady gamma i odwrotny gaussowski (wykorzystywane do modelowania wysokości szkody). W tabeli 1 zostały przedstawione podstawowe rozkłady z rodziny wykładniczej wraz z odpowiadającymi im funkcjami prawdopodobieństwa oraz parametrami θ i ψ .

Drugim elementem GLM jest składnik systematyczny, czyli liniowa kombinacja zmiennych objaśniających następującej postaci

$$\eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}.$$

Należy zwrócić uwagę na fakt, że tak jak w klasycznym modelu regresji liniowej istotna jest tutaj liniowość względem parametrów. Oznacza to, że zmienne X_{ij}

²⁵ Zob. np. J. Fox, *Applied Regression Analysis and Generalized Linear Models*, Sage, Los Angeles 2008, s. 379–381.

mogą być funkcjami oryginalnych zmiennych i zawierać np. interakcje między czynnikami.

Tabela 1. Wybrane rozkłady z rodziny wykładniczej i ich charakterystyki

Rozkład	Funkcja prawdopodobieństwa (gęstości), $f(x)$	D_ψ	θ	$b(\theta)$	ψ
Normalny	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$(-\infty, +\infty)$	μ	$\frac{\theta^2}{2}$	σ^2
Poissona	$\exp(-\lambda) \frac{\lambda^x}{x!}$	$\{0, 1, 2, \dots\}$	$\ln(\lambda)$	$\exp(\theta)$	1
Dwumianowy	$\binom{n}{x} p^x (1-p)^{n-x}$ $n \in \mathbb{N}$ – ustalone	$\{0, 1, \dots, n\}$	$\ln\left(\frac{p}{1-p}\right)$	$n \ln(1 + \exp(\theta))$	1
Gamma	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$	$(0, +\infty)$	$-\frac{\beta}{\alpha}$	$-\ln(-\theta)$	$\frac{1}{\alpha}$
Odwrotny gaussowski	$\frac{\alpha y^{-\frac{3}{2}}}{\sqrt{2\pi\beta}} \exp\left(-\frac{(\alpha - \beta y)^2}{2\beta y}\right)$	$(0, +\infty)$	$-\frac{\beta^2}{2\alpha^2}$	$-\sqrt{-2\theta}$	$\frac{\beta}{\alpha^2}$

Źródło: opracowanie własne na podstawie: R. Kaas, M.J. Goovaerts, J. Dhaene, M. Denuit, *Modern Actuarial Risk Theory*, Kluwer Academic Publishers, Dordrecht 2001.

Ostatnim elementem jest funkcja wiążąca, którą będziemy oznaczać przez g . Zakładamy, że funkcja ta jest różniczkowalna i odwracalna. Rolą funkcji wiążącej jest przekształcenie wartości oczekiwanej na predyktor liniowy według wzoru

$$g(\mu_i) = \eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}. \quad (2)$$

Powyższy wzór pokazuje, w jakim sensie GLM uogólniają klasyczny model regresji liniowej, który otrzymujemy, przyjmując rozkład normalny dla składnika losowego oraz $g(y) = y$ dla każdego y . Klasa uogólnionych modeli liniowych zawiera również m.in. regresję logistyczną oraz regresję Poissona²⁶. Jeśli funkcja g

²⁶ Zob. np. A.C. Cameron, P. Trivedi, *Microeconometrics: Methods and Applications*, Cambridge University Press, New York 2005.

jest tak dobrana, że zachodzi $\theta = \eta = g(\mu)$, to nazywamy ją kanoniczną funkcją wiążącą.

Parametry $\alpha, \beta_1, \dots, \beta_k$ są szacowane najczęściej metodą największej wiarygodności. Dzięki temu jest możliwe stosowanie formalnych metod wnioskowania statystycznego. W szczególności dzięki asymptotycznej normalności tych estymatorów możliwe jest konstruowanie przedziałów ufności dla parametrów i przeprowadzanie parametrycznych testów istotności poszczególnych zmiennych.

Z powyższych rozważań wynika, że uogólnione modele liniowe z jednej strony cechują się prostotą wynikającą z podobieństwa do dobrze znanych modeli liniowych, a z drugiej strony – elastycznością, umożliwiającą modelowanie zmiennych z szerokiej rodziny rozkładów, np. zmiennych dyskretnych. Sprawia to, że są one szczególnie użyteczne w ubezpieczeniach komunikacyjnych, w których podstawowe zmienne (liczba szkód i wysokość odszkodowań) nie mają rozkładu normalnego, a ponadto zachodzi potrzeba jednoczesnego uwzględnienia wielu zmiennych taryfowych. Warto zwrócić uwagę na to, że uogólnione modele liniowe z jednej strony pozwalają weryfikować, czy dana zmienna taryfowa istotnie różnicuje szkodowość w populacji, a z drugiej strony umożliwiają łatwe wyznaczenie składki. Przykładowo, przyjmując logarytmiczną funkcję wiążącą, otrzymujemy

$$\mu_i = \exp\left(\alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}\right) = \exp(\alpha) \prod_{j=1}^k \exp(\beta_j X_{ij}), \quad (3)$$

co prowadzi do multiplikatywnej taryfy składek.

W kontekście przedmiotu niniejszej pracy na uwagę zasługuje fakt, że przy stosowaniu modelu statystycznego łatwo jest uwzględnić zmienne zbudowane na podstawie danych kredytowych. W takim przypadku składnik systematyczny będzie postaci

$$\eta_i = f_1(\text{dane ubezpieczeniowe}) + f_2(\text{dane kredytowe}), \quad (4)$$

gdzie pierwszy składnik odpowiada liniowej funkcji danych, którymi dysponuje ubezpieczyciel, a drugi składnik zawiera zmienne kredytowe. Taka postać modelu umożliwia jego weryfikację statystyczną, w tym ocenę istotności statystycznej zarówno każdej zmiennej kredytowej z osobna, jak i ich łącznej istotności. Warto podkreślić to, że uogólnione modele liniowe pozwalają wykorzystać nie tylko zmienne kategoryczne (zakodowane najczęściej za pomocą zmiennych binarnych), ale także zmienne numeryczne, co rozszerza możliwości ich stosowania.

Uogólnione modele liniowe zostały zastosowane w badaniu UFG i BIK, które zostanie omówione w kolejnym punkcie.

5. Badanie UFG i BIK

W poprzednich punktach zostały omówione podstawowe zagadnienia związane z oceną ryzyka w ubezpieczeniach komunikacyjnych oraz przedstawione wybrane aspekty wykorzystania w tym procesie informacji z rejestrów kredytowych. Przytoczone wyniki badań amerykańskich wskazują na występowanie korelacji między zjawiskami ubezpieczeniowymi i zjawiskami bankowymi. Pojawia się w związku z tym pytanie, czy podobne zależności można zaobserwować w przypadku Polski. W niniejszym punkcie zostaną przedstawione założenia i rezultaty projektu badawczego zrealizowanego wspólnie przez Ubezpieczeniowy Fundusz Gwarancyjny i Biuro Informacji Kredytowej SA, mającego na celu udzielenie odpowiedzi na to pytanie.

5.1. Cel badania i jego założenia

Celem projektu badawczego realizowanego wspólnie przez UFG i BIK było sprawdzenie zdolności predykcyjnych zagregowanych danych statystycznych pochodzących z baz UFG i BIK w obszarze ryzyka ubezpieczeniowego odnośnie do ubezpieczeń komunikacyjnych. Szczególnie skupiono się na sprawdzeniu użyteczności danych kredytowych jako dodatkowego źródła wiedzy przy prognozowaniu przyszłej szkodowości klienta. Na gruncie wyników badań empirycznych przedstawionych w punkcie trzecim została postawiona robocza hipoteza, że negatywna historia kredytowa klienta jest powiązana z wyższym ryzykiem po stronie ubezpieczyciela.

W związku z charakterem bazy danych Ośrodka Informacji UFG, prowadzonej na podstawie przepisów art. 102 ustawy z dnia 22 maja 2003 r. o ubezpieczeniach obowiązkowych, UFG i PBUK (Dz. U. z 2003 r. Nr 124, poz. 1152), badanie obejmowało ubezpieczenia komunikacyjne AC oraz OC p.p.m. W obszarze rozważań znajdowały się wyłącznie umowy ubezpieczenia zawarte przez osoby fizyczne, przy czym – zgodnie z obowiązującymi przepisami dotyczącymi ochrony danych osobowych – wykluczono połączenie danych ubezpieczeniowych i kredytowych na poziomie klienta. Z tego względu zdecydowano o zagregowaniu danych na poziomie kodu pocztowego, miesiąca zawarcia umowy

ubezpieczenia, przedziału wiekowego i płci (tzw. wymiary łączące), a także w podziale ze względu na rodzaj umowy (OC/AC). Oprócz analizy danych UFG i BIK połączonych na poziomie zagregowanym, po stronie UFG możliwe było przeprowadzenie analizy indywidualnych danych ubezpieczeniowych, do których zostały dołączone zagregowane dane kredytowe.

Za okres obserwacji przyjęto lata 2011–2012. W konsekwencji dla każdego z badanych miesięcy otrzymano informacje dotyczące umów zawartych w danym miesiącu dla danej kombinacji poziomów wymiarów łączących oraz informacje z BIK według stanu na początek tego samego miesiąca oraz na tym samym poziomie agregacji. Ten sposób łączenia danych jest podyktowany specyfiką metod taryfikacji *a priori*, omówioną w punkcie drugim – zakład ubezpieczeń ocenia ryzyko związane z umową ubezpieczenia na podstawie danych dostępnych w momencie jej zawierania.

Na potrzeby badania ryzyko związane z umową ubezpieczenia było mierzone za pomocą częstości zdarzeń, tzn. ilorazu liczby zdarzeń ubezpieczeniowych zaszłych w okresie obowiązywania umowy oraz ekspozycji na ryzyko, przy czym ekspozycja umowy ubezpieczenia na ryzyko została wyznaczona jako okres jej obowiązywania wyrażony w latach. Dodatkowo, modelowane było prawdopodobieństwo, że w okresie ochrony wystąpi co najmniej jedno zdarzenie. W części badania skoncentrowanej na danych zagregowanych wielkości te zostały zastąpione odpowiednio przez częstość zdarzeń na poziomie agregatu oraz odsetek zawartych umów, dla których wystąpiło co najmniej jedno zdarzenie ubezpieczeniowe. W procesie modelowania zostały wykorzystane opisane w punkcie czwartym uogólnione modele liniowe – regresja Poissona oraz regresja ujemna dwumianowa w przypadku liczby zdarzeń i regresja logistyczna dla prawdopodobieństwa wystąpienia szkody. We wszystkich modelach składnik systematyczny przyjmował postać daną wzorem (4).

5.2. Przygotowanie danych

Na potrzeby badania obie uczestniczące w nim strony przygotowały zestaw zmiennych, które – przy uwzględnieniu odpowiednio wiedzy z zakresu analizy ryzyka kredytowego i ubezpieczeniowego każdej z nich, a także wyników badań amerykańskich – mogły być użyteczne do realizacji celu badania. Dane kredytowe wykorzystane w badaniu obejmowały następujące obszary informacyjne:

- poziom zadłużenia;
- nadmierne zadłużenie;

- obciążenia klienta;
- opóźnienia w spłacie;
- podejrzenie wyłudzenia;
- ocena punktowa.

Dane przygotowane przez UFG obejmowały natomiast:

- dane o umowach ubezpieczenia i obciążających je szkodach;
- cechy demograficzne ubezpieczonego (wiek i płeć);
- region zamieszkania ubezpieczonego;
- liczbę szkód obciążających poprzednie umowy ubezpieczonego;
- informację, czy ubezpieczony pojazd ma historię ubezpieczenia w bazie OI UFG;
- rodzaj pojazdu.

Po połączeniu obu zbiorów danych powstał zbiór liczący ok. 6,5 mln obserwacji. W celu zapewnienia odpowiedniej liczebności agregatów konieczna była dalsza agregacja danych względem wymiarów łączących. Na podstawie analiz jednokierunkowych zmiennych objaśniających z częstością zdarzeń (w tym metod analizy skupień) zdecydowano o połączeniu niektórych grup wiekowych, a także dokonano agregacji kodów pocztowych, do czego zastosowano dwa podejścia. W pierwszym z nich pogrupowano kody pocztowe według pierwszych trzech cyfr, odrzucając grupy o zbyt małej wielkości mierzonej liczbą zdarzeń oraz ekspozycją na ryzyko. W drugim podejściu (agregacja mieszana) grupy kodowe o wystarczającej wielkości zostały włączone do analizy jako samodzielne obserwacje, natomiast pozostałe grupy kodowe połączono według trzech pierwszych cyfr. Utworzone w ten sposób zbiory liczyły odpowiednio 14,5 tys. oraz 18 tys. obserwacji. Na potrzeby drugiej części badania dla każdego rodzaju umowy została wylosowana próba licząca 100 tys. obserwacji, do których dołączono dane kredytowe na poziomie agregacji ustalonym w pierwszym kroku. W celu weryfikacji zdolności prognostycznych budowanych modeli estymacja parametrów odbywała się na podstawie zbioru uczącego (umowy zawarte w 2011 r.), a weryfikacja modelu była przeprowadzana na zbiorze walidacyjnym (2012 r.). Należy zwrócić uwagę na fakt, że ze względu na brak historii ubezpieczenia w przypadku wielu ubezpieczonych nie można wyznaczyć częstości szkód obciążających poprzednie umowy ubezpieczonego. Jednocześnie jednak wykorzystanie liczby, a nie częstości zdarzeń mogłoby zaburzyć wyniki analizy dla podmiotów, które miały w przeszłości wiele umów. Z tego względu w tej części analizy wykorzystano syntetyczny wskaźnik opisujący przeszłą szkodowość skonstruowany jako predyktor *credibility* w modelu

Bühlmann–Strauba²⁷. Ubezpieczonym bez historii ubezpieczenia jest przypisywana rynkowa częstość zdarzeń, a dla ubezpieczonych z historią wartość tej zmiennej jest średnią ważoną częstości rynkowej i częstości indywidualnej, przy czym waga częstości indywidualnej jest tym większa, im większa była ekspozycja na ryzyko poprzednich umów ubezpieczonego.

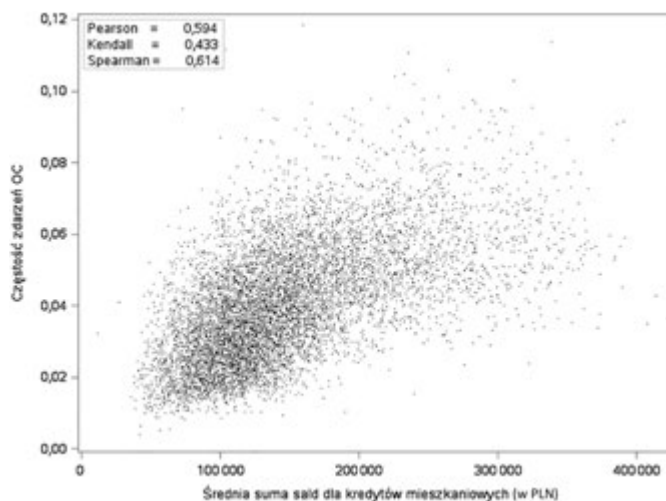
Warto w tym miejscu rozważyć zwrócić uwagę na ograniczenia przeprowadzonego badania. Łączenie danych zagregowanych powoduje, że nie można zagwarantować odpowiedniości między zbiorem ubezpieczonych oraz zbiorem klientów banków. Analiza statystyczna jest więc prowadzona przy założeniu, że pewne relacje między zjawiskami bankowymi i ubezpieczeniowymi są możliwe do zidentyfikowania na przyjętym poziomie agregacji. W związku z tym uzyskane wyniki mają charakter przybliżony, a bardziej precyzyjna weryfikacja postawionej hipotezy wymaga analizy danych połączonych na poziomie klienta, która w obecnym porządku prawnym nie jest możliwa do wykonania.

5.3. Eksploracja danych

Wyniki badań empirycznych opisanych w punkcie trzecim wskazują na użyteczność analiz jednokierunkowych w celu wstępnej oceny użyteczności danych kredytowych do oceny ryzyka. Wstępna analiza danych objęła więc zbadanie rozkładów poszczególnych zmiennych oraz siły związku potencjalnych zmiennych objaśniających z modelowanymi wskaźnikami. Analizę siły związku między zmiennymi objaśniającymi a modelowanymi wskaźnikami prowadzono przy wykorzystaniu współczynników korelacji Pearsona, Kendalla i Spearmana. Wykorzystanie rangowych miar korelacji było podyktowane ich użytecznością przy wykrywaniu monotonicznej, a nie tylko liniowej zależności. Dla przykładu, na rysunku 1 przedstawiono wykres rozrzutu punktów dla częstości zdarzeń OC i średniej sumy sald kredytów hipotecznych wraz z miarami korelacji.

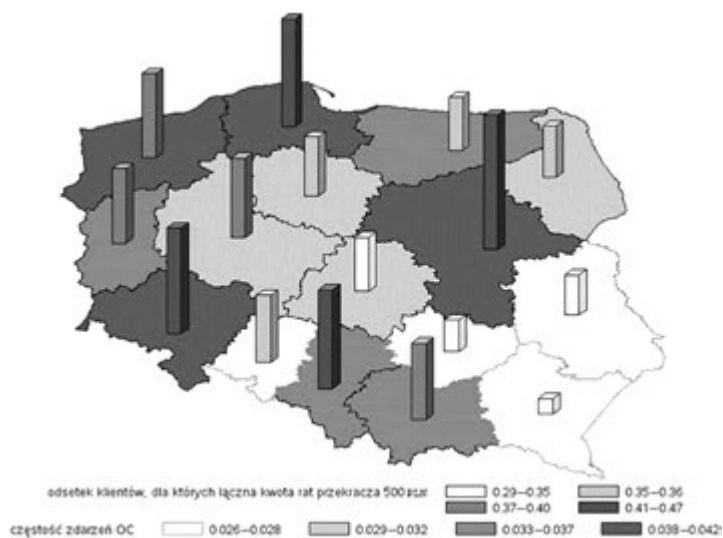
Poniższy wykres wskazuje, że występuje istotna dodatnia korelacja między badanymi zmiennymi. Analiza korelacji pozwoliła również na wstępny wybór zmiennych objaśniających, a także na identyfikację grupy zmiennych silnie skorelowanych ze sobą.

²⁷ Zob. H. Bühlmann, A. Gisler, *A Course In Credibility Theory and Its Applications*, Springer, Berlin–Heidelberg 2005.



Rysunek 1. Przykładowy wykres rozrzutu wraz z miarami korelacji

Źródło: opracowanie własne.



Rysunek 2. Relacja między częstością zdarzeń OC oraz łączną kwotą rat w ujęciu przestrzennym

Źródło: opracowanie własne.

Struktura przestrzenna dostępnych danych pozwala na ich wizualizację za pomocą map. Na rysunku 2 odcień obszaru reprezentuje częstość zdarzeń z umów OC p.p.m., natomiast wykres słupkowy przedstawia odsetek klientów, dla

których łączna kwota rat przekracza 500 PLN (w odniesieniu do liczby klientów, dla których łączna kwota rat zawiera się w przedziale 100–10 tys. PLN). Należy podkreślić to, że wysokość słupków odzwierciedla uporządkowanie wartości badanej zmiennej, a nie różnice między nimi.

Powyższa mapa wskazuje na występowanie obszarów o podwyższonej lub obniżonej (w stosunku do całego rynku) wartości obu badanych zmiennych. Analiza przestrzenna może stanowić więc uzupełnienie standardowej analizy korelacji, a także posłużyć np. do regionalnego zróżnicowania składki.

Jako podstawowe wnioski z analizy opisowej należy wskazać silną korelację między przeszłą i obecną liczbą szkód, a także wysoką korelację części zmiennych kredytowych ze zmiennymi objaśnianymi. Tego rodzaju analiza wstępnie wskazuje na przydatność zmiennych kredytowych do oceny ryzyka ubezpieczeniowego, choć dokładniejsza weryfikacja tej hipotezy wymaga zastosowania metod statystyki wielowymiarowej.

5.4. Omówienie wyników i wnioski

5.4.1. Dane zagregowane

W ramach modelowania dla danych zagregowanych (dane zagregowane UFG połączone z danymi zagregowanymi BIK) zostały sporządzone modele analityczne w podziale na grupy wyznaczone przez:

- 1) rodzaj umowy ubezpieczenia – AC/OC,
- 2) zmienną objaśnianą – liczba zdarzeń/liczba umów szkodowych,
- 3) rodzaj agregacji – agregacja na trzech cyfrach kodu pocztowego/agregacja mieszana.

Do modelowania liczby zdarzeń wykorzystano modele regresji Poissona i regresji ujemnej dwumianowej, natomiast do modelowania liczby umów szkodowych – model regresji dwumianowej (por. punkt czwarty). Zastosowano dwa podejścia do wyboru zmiennych objaśniających – opisowe i postępujące. W podejściu opisowym zbudowano szereg modeli, w których znalazła się jedna zmienna z każdej grupy wyodrębnionej podczas analizy korelacji, natomiast w metodzie postępującej punktem wyjścia był model uwzględniający wyłącznie dane UFG, do których kolejno były dołączane zmienne kredytowe. W obu podejściach zmienne objaśniające oceniano pod kątem istotności statystycznej oraz wpływu na wartość bayesowskiego kryterium informacyjnego (BIC), co miało na celu uniknięcie efektu „przeuczenia” modelu. Dla każdego podejścia i każdej grupy zostało zbudowanych od trzech do pięciu modeli, spośród których wybrano model ostateczny.

We wszystkich analizowanych modelach zmienne kredytowe okazywały się statystycznie istotne, przy czym ich liczba była większa w przypadku umów OC. Jednocześnie dane kredytowe wydają się odgrywać większą rolę w modelowaniu liczby zdarzeń niż w modelowaniu liczby umów szkodowych. Wśród zmiennych kredytowych występujących w wielu modelach należy wymienić zmienne dotyczące: wykorzystania limitów kredytowych, udziału salda należności wymagalnych w sumie sald, łącznej kwoty rat oraz sumy sald bez kredytów mieszkaniowych. W większości przypadków znak parametru przy tych zmiennych jest zgodny z hipotezą badawczą, tj. gorsza sytuacja kredytowa jest związana z większą częstością zdarzeń, natomiast w przypadku niektórych zmiennych znak ten jest przeciwny. Sugeruje to, że zjawiska kredytowe mogą oddziaływać na zjawiska ubezpieczeniowe w różny sposób, niepodlegający jednoznacznej interpretacji. Przykładowe wyniki estymacji parametrów modelu regresji ujemnej dwumianowej dla liczby zdarzeń z umów OC p.p.m. zostały przedstawione w tabeli 2. Wyniki te wskazują, że oprócz wieku, płci i historii szkodowej ubezpieczonego statystycznie istotne są informacje dotyczące historii ubezpieczenia pojazdu, a także kilka zmiennych kredytowych.

Tabela 2. Przykładowe wyniki estymacji dla OC p.p.m.

Zmienna		Ocena	P-value
Wyraz wolny		-4,1582	<,0001
Interakcja: płeć * wiek	K <= 30 lat	-0,0059	0,6895
	K (30,35]	-0,097	<,0001
	K (35,40]	-0,0786	<,0001
	K (40,50]	-0,0173	0,132
	K (50,55]	-0,0103	0,4041
	K (55,60]	0,0273	0,04
	K > 60 lat	0,1507	<,0001
	M <= 30 lat	-0,0685	<,0001
	M (30,35]	-0,2394	<,0001
	M (35,40]	-0,283	<,0001
	M (40,50]	-0,1913	<,0001
	M (50,55]	-0,1676	<,0001
	M (55,60]	-0,1099	<,0001
	M > 60 lat	-	-
Odsetek klientów, dla których suma sald bez kredytów mieszkaniowych jest większa niż 10 tys. PLN (w odniesieniu do liczby klientów, dla których suma sald jest większa niż 500 PLN)		1,3953	<,0001

Zmienna	Ocena	P-value
Odsetek klientów, dla których suma sald dla kredytów mieszkaniowych jest w przedziale 200 tys.–1 mln PLN	0,5938	<,0001
Odsetek klientów, dla których łączna kwota rat przekracza 3 tys. PLN (w odniesieniu do liczby klientów, dla których suma sald dla kredytów ratałnych przekracza 500 PLN)	-0,6413	0,0004
Częstość zdarzeń związanych z poprzednimi umowami ubezpieczonych w okresie 5 lat	15,655	<,0001
Odsetek klientów, którzy w ciągu ostatnich 12 miesięcy pytali w więcej niż 3 bankach	0,4504	0,0017
Odsetek zawartych umów, w których pojazd występuje po raz pierwszy w bazie OI UFG	1,438	<,0001
Odsetek klientów, dla których udział sumy salda należności wymagalnych w sumie sald przekracza 10% (w odniesieniu do liczby klientów, dla których suma sald przekracza 500 PLN)	1,0616	<,0001
Odsetek klientów, dla których suma sald bez kredytów mieszkaniowych przekracza średni limit ponad pięciokrotnie	-0,3538	<,0001
Odsetek klientów, dla których stosunek sumy sald dla limitów do sumy limitów przekracza 77%	-0,8549	<,0001

Źródło: opracowanie własne.

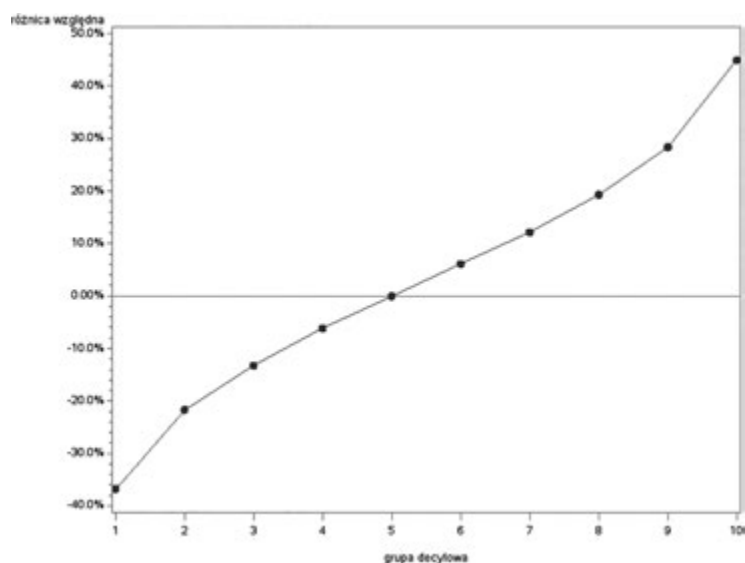
W celu oceny zdolności predykcyjnych utworzonych modeli zostały wyznaczone błędy dopasowania na zbiorze testowym oraz błędy prognozy na zbiorze walidacyjnym. Za miary błędów przyjęto błąd średniokwadratowy (MSE) i średni błąd bezwzględny (MAE). Wartości tych miar zostały wyznaczone dla modeli zbudowanych na podstawie danych UFG, a następnie porównane z błędami obliczonymi na podstawie modeli zawierających dane kredytowe. Uzyskane wyniki wskazują, że zmniejszenie błędów prognozy wynikające z uwzględnienia zmiennych kredytowych może sięgać nawet 10–15%. Największa korzyść z wykorzystania danych kredytowych była widoczna przy modelowaniu częstości zdarzeń z umów OC, podczas gdy przy umowach AC efekt ten był mniej zauważalny.

5.4.2. Dane indywidualne

W drugiej części analizy zostało przeprowadzone modelowanie na zbiorze danych, który powstał przez połączenie danych indywidualnych UFG z danymi zagregowanymi BIK. Łączenie odbywało się po kodzie pocztowym (według agregacji mieszanej), roku i wieku ubezpieczonego. Tego rodzaju analiza miała na celu sprawdzenie, czy dane kredytowe mogą być użyteczne również przy indywidualnej ocenie ryzyka, co ma duże znaczenie praktyczne.

W ramach analizy danych indywidualnych budowano modele dla liczby zdarzeń (model regresji Poissona) oraz prawdopodobieństwa wystąpienia zdarzenia szkodowego (regresja logistyczna). Selekcja zmiennych była przeprowadzona metodą krokową, analogicznie do analizy danych zagregowanych.

Analiza danych indywidualnych prowadzi do podobnych wniosków co analiza danych zagregowanych. Zmienne kredytowe okazały się statystycznie istotne w każdym z budowanych modeli, przy czym ponownie ich znaczenie wydaje się większe w przypadku umów OC. Jako najważniejsze należy wskazać: zmienne dotyczące udziału sumy salda należności wymagalnych bez kredytów mieszkaniowych w sumie salda bez kredytów mieszkaniowych, łączną kwotę rat oraz sumę sald bez kredytów mieszkaniowych. Kierunek oddziaływania tych zmiennych jest zgodny z kierunkiem wskazanym w hipotezie badawczej. Dla przykładu, na rysunku 3 została przedstawiona wynikająca z modelu relacja między odsetkiem klientów, dla których łączna kwota rat przekracza 500 PLN, oraz częstością zdarzeń szkodowych. Obliczenia przeprowadzono dla grup decylowych określonych na podstawie rozkładu wartości badanego wskaźnika kredytowego w raporcie kredytowym. Jako grupa odniesienia została wybrana piąta grupa decylowa.



Rysunek 3. Względna różnica częstości zdarzeń dla łącznej kwoty rat

Źródło: opracowanie własne.

Z wykresu wynika, że częstość zdarzeń w zależności od wartości analizowanej zmiennej waha się od -40% do 50% częstości zdarzeń dla grupy bazowej. Oznacza to, że zmienna ta istotnie różnicuje szkodowość w badanej próbie ubezpieczonych. Jeśli chodzi o zmienne ubezpieczeniowe, to należy podkreślić duży wpływ historii szkodowej ubezpieczonego – również w przypadku danych indywidualnych była to najważniejsza zmienna w każdym budowanym modelu. Innymi istotnymi zmiennymi były: interakcja między wiekiem a płcią, rodzaj pojazdu oraz informacja, czy w bazie UFG występuje historia ubezpieczeniowa pojazdu lub podmiotu. Otrzymane wyniki są więc w dużej mierze spójne z wnioskami płynącymi z analizy danych zagregowanych.

5.4.3. Wnioski

Podsumowując wyniki przeprowadzonej analizy, można sformułować następujące wnioski:

- 1) najważniejszą z analizowanych zmiennych w obszarze oceny ryzyka wydaje się liczba przeszłych zdarzeń komunikacyjnych ubezpieczonego; jest to w zgodzie z obowiązującą praktyką rynkową, w której historia szkodowości jest wykorzystywana do taryfikacji, np. przez system bonus-malus;
- 2) zmienne kredytowe są statystycznie istotne w większości zbudowanych modeli; ich wpływ na zmienną objaśnianą jest mniejszy niż w przypadku historii szkodowej, choć otrzymane wyniki wskazują, że wykorzystanie informacji kredytowej może poprawić dopasowanie modelu do danych oraz zmniejszyć błąd prognozy na zbiorze walidacyjnym; najlepsze wyniki zostały uzyskane dla częstości zdarzeń dla umów OC p.p.m. na poziomie zagregowanym;
- 3) wpływ zmiennych kredytowych wydaje się większy w przypadku umów OC; może to wynikać z faktu, że z umów AC są likwidowane również szkody niewynikające z winy ubezpieczonego (np. wandalizm, działanie żywiołów), niemające związku z jego sytuacją kredytową;
- 4) znak parametru przy niektórych zmiennych kredytowych nie jest zgodny ze znakiem określonym w hipotezie badawczej; pokazuje to, że zjawiska bankowe mogą wpływać na zjawiska ubezpieczeniowe w różny sposób, trudny do analizy na podstawie danych zagregowanych; sytuacja kredytowa może np. wpływać na sposób i intensywność użytkowania pojazdu, a także na sam fakt posiadania pojazdu lub korzystania z niego; w związku z tym efekt na poziomie zagregowanym może być trudny do jednoznacznego określenia.

6. Podsumowanie

W niniejszej pracy zostały omówione wybrane kwestie związane z wykorzystaniem informacji kredytowej w procesie oceny ryzyka w ubezpieczeniach komunikacyjnych. Rozważania rozpoczęto od przedstawienia procesu oceny ryzyka ubezpieczeniowego. Następnie zasygnalizowano możliwość wykorzystania w tym procesie informacji kredytowej, a także dokonano przeglądu literatury przedmiotu oraz funkcjonujących w praktyce rozwiązań. W kolejnym punkcie zostały przedstawione modele aktuarialne użyteczne do oceny ryzyka oraz taryfikacji. Ostatnią część pracy stanowi omówienie wyników badania statystycznego przeprowadzonego przez UFG i BIK.

Mimo wielu ograniczeń związanych z przygotowaniem danych można powiedzieć, że uzyskane wyniki wskazują na użyteczność informacji kredytowej w ocenie ryzyka ubezpieczeniowego, przynajmniej w wybranych obszarach, np. przy ubezpieczeniu OC. Wynika stąd, że wykorzystanie informacji kredytowej w taryfikacji może być w przyszłości interesującym kierunkiem działań ubezpieczycieli, a wnioski z przedstawionego badania mogą posłużyć jako punkt wyjścia do dalszej dyskusji dotyczącej tego zagadnienia.

Bibliografia

- Brackett P., Golden L., *Biological and Psychobehavioral Correlates of Risk Scores and Automobile Insurance Losses: Toward an Explication of Why Credit Scoring Works*, „Journal of Risk and Insurance” 2007, vol. 74, issue 1, s. 23–63.
- Bühlmann H., Gisler A., *A Course in Credibility Theory and its Applications*, Springer, Berlin–Heidelberg 2005.
- Cameron A.C., Trivedi P., *Microeconometrics: Methods and Applications*, Cambridge University Press, New York 2005.
- Credit Reports and Insurance Underwriting*, NAIC White Papers, National Association of Insurance Commissioners, 1997.
- Denuit M., Maréchal X., Pitrebois S., Walhin J., *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, Wiley, New York 2007.
- Fox J., *Applied Regression Analysis and Generalized Linear Models*, Sage, Los Angeles 2008.
- Kaas R., Goovaerts M.J., Dhaene J., Denuit M., *Modern Actuarial Risk Theory*, Kluwer Academic Publishers, Dordrecht 2001.

- Metody aktuarialne. Zastosowania matematyki w ubezpieczeniach*, red. W. Ronka-Chmielowiec, Wydawnictwo Naukowe PWN, Warszawa 2013.
- Monaghan J.E., *The Impact of Personal Credit History on Loss Performance in Personal Lines*, 2000 CAS Winter Forum, Casualty Actuarial Society, 2000, s. 79–106.
- Nelder J.A., Wedderburn R.M., *Generalized Linear Models*, „Journal of the Royal Statistical Society” 1972, vol. 135, no. 3, s. 370–384.
- Składki i ryzyko ubezpieczeniowe. Modelowanie stochastyczne*, red. W. Ostasiewicz, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 2004.
- Ustawa z dnia 29 sierpnia 1997r. – Prawo bankowe (Dz. U. z 2002 r. Nr 72, poz. 665).
- Ustawa z dnia 22 maja 2003 r. o działalności ubezpieczeniowej (Dz. U. z 2003 r. Nr 124, poz. 1151).
- Ustawa z dnia 22 maja 2003 r. o ubezpieczeniach obowiązkowych, Ubezpieczeniowym Funduszu Gwarancyjnym i Polskim Biurze Ubezpieczycieli Komunikacyjnych (Dz. U. z 2003 r. Nr 124, poz. 1152).
- Wu C.-S., Guszczka J.C., *Does Credit Score Really Explain Insurance Losses? Multivariate Analysis from a Data Mining Point of View*, 2003 CAS Winter Forum, Casualty Actuarial Society, 2003, s. 113–138.

Źródła sieciowe

- Credit-Based Insurance Scores: Impacts on Consumers of Automobile Insurance*, A Report to Congress, Federal Trade Commission, 2007, http://www.ftc.gov/os/2007/07/P044804FACTA_Report_Credit-Based_Insurance_Scores.pdf (dostęp: 15.08.2014).
- Fitzgerald F.M., *The Use of Insurance Credit Scoring In Automobile and Homeowners Insurance*, Office of Financial and Insurance Services, Michigan 2002, www.michigan.gov/documents/cis_ofis_credit_scoring_report_52885_7.pdf (dostęp: 16.08.2014).
- Miller M., Smith R.A., *The Relationship of Credit-based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity*, Epic Actuaries LLC, 2003, http://www.ask-epic.com/Publications/Relationship%20of%20Credit%20Scores_062003.pdf (dostęp: 16.08.2014).
- Risk Classification (for All Practice Areas)*, „Actuarial Standard of Practice” 2005, no. 12, Actuarial Standards Board, http://www.actuarialstandardsboard.org/pdf/asops/asop012_132.pdf (dostęp: 14.08.2014).
- Use of Credit Information by Insurers in Texas*, Report to the 79th Legislature, Texas Department of Insurance, 2004, <http://www.tdi.texas.gov/reports/documents/creditall04.pdf> (dostęp: 11.08.2014).

Use of Credit Information by Insurers in Texas: The Multivariate Analysis, Report to the 79th Legislature, Texas Department of Insurance, 2005, <http://www.tdi.texas.gov/reports/documents/credit05sup.pdf> (dostęp: 11.08.2014).

Use of Credit Scores by Insurers, Canadian Council of Insurance Regulation, Issues Paper, 2011, [http://www.ccir-ccrra.org/en/init/credit_scor/CCIR%20credit%20scores%20issues%20paper\(En\).pdf](http://www.ccir-ccrra.org/en/init/credit_scor/CCIR%20credit%20scores%20issues%20paper(En).pdf) (dostęp: 14.08.2014).

* * *

The use of credit data for risk classification in automobile insurance

Summary

This paper presents theoretical and practical aspects of using credit data in insurance risk classification. The correlation between credit history and insurance losses is discussed, as well as its possible sources and methods of analysis. Furthermore, the literature on the subject is reviewed, with a focus on the results of empirical studies and practical implementations. Finally, the paper describes the setting of the research study conducted jointly by the Polish Insurance Guarantee Fund and Credit Information Bureau to identify the correlation between credit history and insurance losses in the Polish automobile insurance market. The results of the study indicate that the cooperation between insurers and banks in this field may be beneficial.

Keywords: insurance risk, risk classification, credit data, generalised linear models, automobile insurance