Daniel Sobiecki[1]

Kolegium Analiz Ekonomicznych
Szkoła Główna Handlowa w Warszawie

Predykcja szkód z uwzględnieniem zależności w ubezpieczeniach AC i OC komunikacyjnym

Streszczenie

Przedmiotem opracowania jest predykcja szkód z uwzględnieniem zależności między dwiema liniami biznesowymi w ubezpieczeniach komunikacyjnych. Wykorzystano wielowymiarowe modele wiarygodności do oszacowania korekty składki ze względu na historię szkodową klienta. Prezentowany model jest uogólnieniem modelu Bühlmanna–Strauba i pozwala analizować wpływ nie tylko liczby szkód, ale także momentów ich wystąpienia. Wnioski z analizy rzeczywistego portfela polis mogą być wykorzystane w sprzedaży krzyżowej ubezpieczeń AC i OC komunikacyjnego.

Słowa kluczowe: sprzedaż krzyżowa, taryfikacja *a posteriori*, wielowymiarowe modele wiarygodności

1. Wstęp

W taryfikacji ubezpieczeń komunikacyjnych można wyróżnić klasyfikację ryzyka *a priori* i korektę ryzyka *a posteriori* (ang. *experience rating*), którą prowadzi się na podstawie historii szkodowej kierowców. Wynika to z faktu, że grupy ubezpieczonych po podziale według widocznych charakterystyk ryzyka nadal nie są jednorodne z powodu nieobserwowanych cech. Zalicza się do nich m.in. doświadczenie w kierowaniu autem, szybkość reakcji, znajomość kodeksu drogowego czy skłonność do jazdy po spożyciu alkoholu. Istnieje bogata literatura przedmiotu przedstawiająca zastosowanie m.in. modeli wiarygodności i podejścia bayesowskiego w taryfikacji *a posteriori*[2].

1 sobieckid@gmail.com.

2 Na przykład: E.W. Frees, V.R. Young, Y. Luo, *Case studies using panel data models*, „North American Actuarial Journal" 2001, vol. 5(4), s. 24–42; R. Schnieper, *Robust Bayesian experience rating*, „ASTIN Bulletin" 2004, vol. 34(1), s. 125–150.

Tradycyjne podejście zakłada wykorzystanie historii szkodowej klienta w danej linii biznesowej (produkcie ubezpieczeniowym) do predykcji liczby szkód w dokładnie tej samej linii biznesowej. Dla przykładu, w pracy J. Lemaire'a[3] jest rozważany jeden nieobserwowany parametr ryzyka dla każdego ubezpieczonego, a składka jest obliczana na podstawie informacji indywidualnej i zbiorowej. W praktyce istnieje duża grupa klientów (tzw. klienci pakietowi), którzy wykupują wiele produktów ubezpieczeniowych w tym samym zakładzie ubezpieczeń, i tę dodatkową informację warto wykorzystać w procesie taryfikacji. W takiej sytuacji nieobserwowany parametr ryzyka można modelować jako wielowymiarowy wektor skorelowanych zmiennych losowych. W niniejszej pracy analiza dotyczy klientów, którzy wykupują dwa główne produkty działu II, czyli ubezpieczenie AC i OC komunikacyjne. W analizowanym portfelu współczynnik korelacji między dwoma latentnymi parametrami ryzyka wynosi 0,3. Przedstawione metody mogą być łatwo uogólnione na przypadek większej liczby produktów.

W analizie wykorzystano wielowymiarowe modele wiarygodności, zaprezentowane w pracy M. Englunda i innych[4]. Autorzy uwzględniają w procesie predykcji strukturę czasową zgłaszania szkód, wysuwając postulat malejącego z czasem wpływu szkody na niewidoczny parametr ryzyka. Jeśli dla danego klienta mamy dostępną historię szkodową dla okresów j od 1 do J, to szkody starsze (odleglejsze w czasie od okresu predykcji $J + 1$) mają mniejszą wagę niż obserwacje z późniejszych okresów. Pojęcie wieku szkody, czyli uwzględnienie zależności czasowych w historii szkodowej, zostało wprowadzone do taryfikacji w przypadku jednowymiarowym w pracy J. Pinqueta i innych[5].

Wielowymiarowe modele wiarygodności znajdują w literaturze wiele rodzajów zastosowania, w których wymiary wektora ryzyka definiuje się inaczej niż przez produkty ubezpieczeniowe czy linie biznesowe. H. Bühlmann i A. Gisler[6] wymieniają przykłady zastosowań do analizy: szkód umiarkowanych i ekstremalnych, rzeczowych i osobowych w ubezpieczeniu OC komunikacyjnym, częstości i wysokości szkody oraz danych z różnych źródeł.

Natomiast literatura dotycząca metod, w których są wykorzystywane dane z różnych linii biznesowych do tego samego ubezpieczonego, nie jest duża.

---

[3] J. Lemaire, *Bonus-Malus Systems in Automobile Insurance*, Kluwer Academic Publishers, New York 1995.

[4] M. Englund, M. Guillén, J. Gustafsson, L.H. Nielsen, J.P. Nielsen, *Multivariate latent risk: A credibility approach*, „ASTIN Bulletin" 2008, vol. 38(1), s. 137–146.

[5] J. Pinquet, M. Guillén, C. Bolancé, *Allowance for the age of claims in Bonus-Malus systems*, „ASTIN Bulletin" 2001, vol. 31, s. 337–348.

[6] H. Bühlmann, A. Gisler, *A course in credibility theory and its applications*, Springer-Verlag, Berlin–Heidelberg–New York 2005.

Jednym z nielicznych wyjątków jest praca D. Desjardins i innych[7], w której system bonus-malus dla flot pojazdów opiera się na historii szkodowej i danych o wykroczeniach związanych z bezpieczeństwem ruchu drogowego.

Prezentowane w artykule podejście do taryfikacji *a posteriori* ma potencjał do poprawiania dokładności oceny ryzyka i może być wykorzystane także w innych sytuacjach, gdy ważne jest rozumienie współzależności między liniami biznesowymi. Za przykład może posłużyć sprzedaż krzyżowa (ang. *cross-selling*), czyli technika marketingowa polegająca na sprzedaży klientowi produktów (lub usług) komplementarnych bądź uzupełniających. Zakład ubezpieczeń może skorzystać z informacji o historii szkodowej klienta w jednej linii biznesowej, by zaoferować mu dodatkowe produkty po konkurencyjnej cenie. Podejście takie wymaga zastosowania przedstawionych modeli wielowymiarowych, które uwzględniają korelację miedzy ryzykiem klienta w poszczególnych produktach. W pracy F. Thuringa[8] można znaleźć zastosowanie wyżej wymienionych metod do wyboru grupy docelowej dla sprzedaży krzyżowej, tak by maksymalizować zysk z podjętej akcji sprzedażowej.

Niniejsza praca jest podzielona w następujący sposób. W punkcie drugim zaprezentowano klasyczny model Bühlmanna–Strauba i jego rozszerzenie na przypadek dwóch wymiarów, które łatwo uogólnić na przypadek *k* wymiarów. W punkcie trzecim przedstawiono model uwzględniający wiek szkody. Zastosowanie omawianych metod do analizy rzeczywistego portfela klientów omówiono w czwartym punkcie. Wnioski z analizy są wymienione w ostatniej części artykułu.

2. Model Bühlmanna–Strauba

2.1. Jednowymiarowy model Bühlmanna–Strauba

W tym podpunkcie zostanie przedstawiony model Bühlmanna–Strauba dla jednego wymiaru i bez wpływu wieku szkody. Modyfikacja w stosunku do

[7] D. Desjardins, G. Dionne, J. Pinquet, *Experience rating schemes for fleets of vehicles*, „ASTIN Bulletin" 2001, vol. 31, s. 81–105.

[8] F. Thuring, *Multivariate credibility with application to cross-selling financial services products*, niepublikowana praca doktorska, City Univeristy London, 2012.

klasycznej wersji, która zakłada równość klientów *a priori*, polega na uwzględnieniu w modelowaniu znanych różnic między klientami.

Niech N_{ij} oznacza liczbę szkód zgłoszonych przez i-tego klienta w j-ym okresie ubezpieczeniowym, przeważnie roku polisowym, z danego produktu ubezpieczeniowego i niech $\boldsymbol{N}_i = (N_{ij})_{j=1,\ldots,J}$. Każdy klient ma swój nieobserwowany, indywidualny profil ryzyka θ_i (inaczej: efekt losowy; w tym modelu niezależny od czasu), który jest realizacją zmiennej losowej $\boldsymbol{\Theta}_i$ (inaczej: parametr niejednorodności portfela). Dodatkowo dysponujemy wiedzą *a priori* o różnicach między klientami w poszczególnych okresach, wyrażoną przez λ_{ij}, która to wielkość jest zmienną objaśnianą w regresji Poissona (lub innym modelu zm. licznikowej) $\lambda_{ij} = \omega_{ij}\exp\left(x_{ij}a\right)$, gdzie ω_{ij} oznacza długość okresu ekspozycji na ryzyko, x_{ij} jest k-wymiarowym wektorem zmiennych objaśniających, a $a \in R^k$ jest wektorem parametrów. Przegląd zagadnień związanych z zastosowaniem modeli regresji w ubezpieczeniach można znaleźć w pracy E. Ohlssona i B. Johanssona[9]. Warto również wspomnieć o kwestii podziału w ubezpieczeniu OC komunikacyjnym zgłoszonych szkód na umiarkowane i ekstremalne, które modeluje się oddzielnie, uzyskując poprawę dokładności taryfikacji[10].

Wykorzystując wprowadzoną notację, załóżmy, że

$$N_{ij} \mid \theta_i \sim Pois(\lambda_{ij}\theta_i), \tag{1}$$

$\mathrm{E}(\theta_i) = 1$ i $\mathrm{V}(\theta_i) = \sigma_{11}^2$. Dodatkowo, niech $(\boldsymbol{\Theta}_1, \boldsymbol{N}_1)$, $(\boldsymbol{\Theta}_2, \boldsymbol{N}_2)$,… będą niezależnymi wektorami losowymi, gdzie $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$, … są niezależne i o tych samych rozkładach.

Zdefiniujmy zmienne $X_{ij} = N_{ij} / \lambda_{ij}$, $\bar{X}_i = \frac{1}{\lambda_{i.}}\sum_j \lambda_{ij} X_{ij}$, gdzie $\lambda_{i.} = \sum_j \lambda_{ij}$ (zakładamy, że $\lambda_{ij} > 0$ dla każdego klienta i okresu). Wtedy zachodzi $\mathrm{E}(X_{ij} \mid \theta_i) = \theta_i$ oraz $\mathrm{V}(X_{ij} \mid \theta_i) = \theta_i / \lambda_{ij}$.

Najlepszy liniowy predyktor (ang. *Best Linear Predictor* – BLP) dla N_{iJ+1} na podstawie historii szkodowej $N_{i1}, \ldots, N_{iJ}$ jest wyznaczony jednoznacznie przez najlepszy liniowy predyktor dla θ_i na podstawie $\bar{X}_i$. Ponieważ X_{ij} spełnia warunki modelu Bühlmanna–Strauba, najlepszy liniowy predyktor dla θ_i, który oznaczmy przez BS_i w przypadku modelu jednowymiarowego, ma postać:

$$BS_i = \mathrm{E}(\theta_i) + COV\left(\theta_i, \bar{X}_i\right) V(\bar{X}_i)^{-1}\left(\bar{X}_i - E\left(\bar{X}_i\right)\right) \tag{2}$$

[9] E. Ohlsson, B. Johansson, *Non-life insurance pricing with generalized linear models*, Springer-Verlag, Berlin–Heidelberg 2010.

[10] Zob. D. Sobiecki, *Dwustopniowe modelowanie składki za ubezpieczenie komunikacyjne OC*, „Prace Naukowe" UE we Wrocławiu, z. 312, Wrocław 2013, s. 116–134.

$$= 1 + z_i \left(\bar{X}_i - 1 \right) = 1 + z_i \left(\frac{N_{i.}}{\lambda_{i.}} - 1 \right), \tag{3}$$

gdzie $N_{i.} = \sum_j N_{ij}$ i $z_i = \lambda_{i.} / \left(\lambda_{i.} + \sigma_{11}^{-2} \right)$. Stąd wynika postać najlepszego liniowego predyktora dla N_{iJ+1} :

$$BLP\left(N_{iJ+1} | N_{i1}, \ldots, N_{iJ} \right) = BS_i \cdot \lambda_{iJ+1}.$$

2.2. Dwuwymiarowy model Bühlmanna–Strauba

W tym podpunkcie zostanie przedstawiony dwuwymiarowy model Bühlmanna–Strauba, ponieważ w analizie empirycznej jest badana zależność między dwoma produktami ubezpieczeniowymi – AC i OC komunikacyjnym. Przedstawione rozważania można analogicznie przeprowadzić dla przypadku *k* wymiarów. W prezentowanym modelu każda linia biznesowa ma swój nieobserwowany parametr ryzyka, co przekłada się na formułowanie poniższych założeń.

Niech N_{ikj} oznacza liczbę szkód, a λ_{ikj} oznacza oczekiwaną liczbę szkód zgłoszonych przez *i*-tego klienta w *j*-ym roku (*j* = 1, …, *J*) z *k*-tej linii biznesowej. Załóżmy, że

$$N_{ij} = (N_{i1j}, N_{i2j})' \mid \left(\theta_{i1}, \theta_{i2} \right) \sim \otimes_{k=1}^{2} Pois\left(\lambda_{ikj} \theta_{ik} \right)$$

oraz niech $\theta_i = (\theta_{i1}, \theta_{i2})'$. Załóżmy także, że

$$E(\theta_i) = 1, \;\; V(\theta_i) = A = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{pmatrix}.$$

Poszukujemy ponownie najlepszego liniowego predyktora N_{iJ+1} na podstawie zaobserwowanej historii szkodowej. Wprowadźmy oznaczenie analogiczne do przypadku modelu jednowymiarowego: $X_{ikj} = N_{ikj} / \lambda_{ikj}$, $\bar{X}_{ik} = \frac{1}{\lambda_{ik.}} \sum_j \lambda_{ikj} X_{kij}$, gdzie $\lambda_{ik.} = \sum_j \lambda_{ikj}$. Zastosowanie wielowymiarowego modelu Bühlmanna–Strauba prowadzi do najlepszego liniowego predyktora dla θ_{ik} (k = 1, 2) na podstawie $\{\bar{X}_{i1}, \bar{X}_{i2}\}$ postaci:

$$MBS_{ik} = \mathrm{E}(\theta_i) + COV\left(\theta_{ik},(\bar{X}_{i1},\ \bar{X}_{i2}\ )\right)V(\begin{pmatrix} \bar{X}_{i1} \\ \bar{X}_{i2} \end{pmatrix})^{-1}\left(\begin{pmatrix} \bar{X}_{i1} \\ \bar{X}_{i2} \end{pmatrix} - E\begin{pmatrix} \bar{X}_{i1} \\ \bar{X}_{i2} \end{pmatrix}\right) \tag{4}$$

dla $k = 1, 2$. Ze względu na symetrię dalsze przekształcenia będą zaprezentowane dla $k = 1$. Postać MBS_{i2} jest analogiczna do poniższej:

$$MBS_{i1} = 1 + \alpha_{i1}\left(\bar{X}_{i1} - 1\right) + \alpha_{i2}\left(\bar{X}_{i2} - 1\right), \tag{5}$$

gdzie współczynniki zaufania przyjmują postać:

$$\alpha_{i1} = \frac{\dfrac{\sigma_{11}^2}{\lambda_{i2.}} + \det\left(A\right)}{\det\left(V\left(\bar{X}_i\right)\right)},$$

$$\alpha_{i2} = \frac{\dfrac{\sigma_{12}^2}{\lambda_{i1.}}}{\det\left(V\left(\bar{X}_i\right)\right)},$$

przy czym $\det\left(V\left(\bar{X}_i\right)\right) = \left(\dfrac{1}{\lambda_{i1.}} + \sigma_{11}^2\right)\left(\dfrac{1}{\lambda_{i2.}} + \sigma_{22}^2\right) - (\sigma_{12}^2)^2$. Wyrażenie (5), które przedstawia najlepszy liniowy predyktor w dwóch wymiarach, ma taką samą postać jak prosty estymator Bühlmanna–Strauba (3). Można pokazać, że estymator dwuwymiarowy wygładza predykcję dokonaną na podstawie estymatora jednowymiarowego, a poziom wygładzenia jest proporcjonalny do korelacji między dwiema liniami biznesowymi. Dla zerowego współczynnika korelacji mamy $\alpha_{i2} = 0$ i $MBS_{i1} = BS_{i1}$, tzn. dwuwymiarowy estymator ulega redukcji do prostego predyktora Bühlmanna–Strauba. Podobnie zachowuje się estymator *MBS*, gdy obserwowana ekspozycja w jednej linii biznesowej dąży do nieskończoności.

3. Wielowymiarowy model z zależnością czasową

W tym punkcie do zaprezentowanych modeli zostanie wprowadzona zależność czasowa przez umożliwienie zmienności nieobserwowanego, indywidualnego profilu ryzyka w czasie. Oznacza to w przypadku wielowymiarowym, że poza korelacją między liniami biznesu uwzględniamy także autokorelację każdej z nich. W tym przypadku niech $\boldsymbol{\Theta}_{ij}$ będzie wielowymiarowym wektorem indywidualnych profili ryzyka klientów w *j*-ym okresie. Zakłada się, że $\boldsymbol{\Theta}_{ij}$ są niezależne dla *i*-tego klienta oraz $E(\boldsymbol{\Theta}_{ij}) = 1$.

Wiek szkody dla przypadku jednowymiarowego został wprowadzony w pracy J. Pinqueta i innych[11]. Autorzy wyprowadzają zgodne i asymptotycznie normalne estymatory wariancji i współczynniki autokorelacji dla opóźnień *h*, gdzie $h = 1,\ldots, H$:

$$\widehat{var_{11}} = \frac{\sum_{i,j}\left[\left(N_{i1j} - \lambda_{i1j}\right)^2 - N_{i1j}\right]}{\sum_{i,j}\lambda_{i1j}^2},$$

$$\widehat{\rho_{11}(h)} = \frac{1}{\widehat{var_{11}}}\frac{\sum_{i|J\rangle h}\sum_{J\geq j>h}\left(N_{i1j} - \lambda_{i1j}\right)\left(N_{i1j-h} - \lambda_{i1j-h}\right)}{\sum_{i|J\rangle h}\sum_{J\geq j>h}\lambda_{i1j}\lambda_{i1j-h}}.$$

W pracy M. Englunda i innych[12] zależność czasowa jest uogólniona na przypadek wielowymiarowy. Wprowadźmy *hist*-operator, który w przypadku dwuwymiarowym przyjmuje następującą postać:

$$hist\left(X_{iJ}\right) = (X_{i11},\ldots,X_{i1J},X_{i21},\ldots,X_{i2J})', \tag{6}$$

i który zwraca pełną historię szkodową przed okresem predykcji $J + 1$. Proces $\{\boldsymbol{\Theta}_{ij}\}$ jest stacjonarny (bardziej szczegółowo – autoregresyjny) w takim sensie, że dla $i = e$ zachodzi:

$$COV\left(\theta_{ikr},\theta_{els}\right) = \sigma_{kl}^2(\rho_{kl})^{|s-r|},$$

gdzie $s = 1, \ldots, J$ i $r = 1, \ldots, J$, a dodatkowo zachodzi symetria $\sigma_{12}^2 = \sigma_{21}^2$ i $\rho_{12} = \rho_{21}$ (dla $i \neq e$, czyli różnych klientów wszystkie kowariancje są zerowe). Przy tych założeniach najlepszy liniowy predyktor dla θ_{iJ+1} na podstawie historii szkodowej ma postać:

$$BLP\left(\theta_{iJ+1}|hist\left(X_{iJ}\right)\right) = \mathrm{E}(\theta_{iJ+1}) + COV\left(\theta_{iJ+1},hist\left(X_{iJ}\right)'\right)V(hist\left(X_{iJ}\right))^{-1}\times$$
$$\times\left(hist\left(X_{iJ}\right) - E\left(hist\left(X_{iJ}\right)\right)\right) = 1 + A\left(B + S_i\right)^{-1}\left(hist\left(X_{iJ}\right) - 1\right), \tag{7}$$

[11] J. Pinquet, M. Guillén, C. Bolancé, op.cit.
[12] M. Englund, M. Guillén, J. Gustafsson, L.H. Nielsen, J.P. Nielsen, op.cit.

gdzie A jest macierzą o wymiarach 2 x 2J (w przypadku k-wymiarowym: k x kJ):

$$A = COV\left(\theta_{iJ+1}, hist\left(X_{iJ}\right)'\right) = \begin{pmatrix} (\sigma^2_{11s})_{s=1,\dots,J} & (\sigma^2_{12s})_{s=1,\dots,J} \\ (\sigma^2_{12s})_{s=1,\dots,J} & (\sigma^2_{22s})_{s=1,\dots,J} \end{pmatrix} \tag{8}$$

z elementami $\sigma^2_{qrs} = \sigma^2_{qr}(\rho_{qr})^{s-1}$. Z kolei B jest macierzą o wymiarach 2J x 2J postaci:

$$B = \begin{pmatrix} b_{11(s,j)} & b_{12(s,j)} \\ b_{12(s,j)} & b_{22(s,j)} \end{pmatrix} \tag{9}$$

z elementami $b_{qr(s,j)} = \sigma^2_{qr}(\rho_{qr})^{|s-j|}$, gdzie $s = 1, \dots, J$ i $j = 1, \dots, J$, które w przypadku dwuwymiarowym tworzą cztery macierze blokowe o wymiarach J x J. Ostatnia z macierzy wprowadzonych w estymatorze (7) ma postać:

$$S_i = E\left(V\left(hist\left(X_{iJ}\right) | hist\left(\theta_{iJ}\right)\right)\right) = diag\left(\frac{1}{\lambda_{i11}}, \dots, \frac{1}{\lambda_{i1J}}, \frac{1}{\lambda_{i21}}, \dots, \frac{1}{\lambda_{i2J}}\right).$$

Postać estymatora (7) jest analogiczna do najlepszych liniowych predyktorów (2) i (4).

Warto zaznaczyć, że zależność czasowa może być modelowana także w sposób ogólniejszy, co zostało zastosowane w empirycznej części pracy. Jednak prezentowana powyżej struktura czasowa odpowiada założeniu o zmieniających się umiejętnościach kierowcy w czasie, a co za tym idzie – słabym wpływie odległej przeszłości na predykcję szkodowości w przyszłych okresach.

4. Analiza danych

W tym punkcie zostanie zaprezentowana analiza danych rzeczywistych pochodzących z zakładu ubezpieczeń działającego na polskim rynku. Dane przedstawiają losową próbkę z portfela klientów indywidualnych (osób fizycznych), którzy na przestrzeni lat 2008–2013 wykupili ubezpieczenia AC i OC komunikacyjne dla swoich samochodów osobowych. Zbiór zawiera informacje o historii szkodowej w obu liniach, długości okresu ochrony (wielkości ekspozycji na ryzyko) oraz oszacowania oczekiwanej rocznej liczby szkód w obu liniach (λ_{ikj} zostały oszacowane w modelu regresji ujemnej dwumianowej w przypadku

AC dla wszystkich szkód, a w przypadku OC dla szkód umiarkowanych, poniżej 70 tys. PLN, na pełnym portfelu klientów). W analizie bierze udział 15 792 klientów podzielonych losowo na dwie grupy w proporcji 80% (grupa A) do 20% (grupa B). Pierwszy podzbiór jest wykorzystywany do oszacowania parametrów każdego z przedstawionych modeli. Natomiast drugi posłuży porównaniu mocy predykcyjnej.

Tabela 1. Podstawowe charakterystyki analizowanego portfela

Grupa	Liczba klientów	Produkt	Liczba szkód $N_{.k.}$	Oczekiwana liczba szkód $\lambda_{.k.}$
A	12 634	autocasco OC komunik.	3 964 2 425	4 200,7 2 407,28
B	3 158	autocasco OC komunik.	952 565	1 085,68 589,6

Źródło: opracowanie własne.

Tabela 1 przedstawia podstawowe charakterystyki dwóch podzbiorów analizowanego portfela. Można z niej odczytać m.in., że AC charakteryzuje się większą częstością niż OC komunikacyjne. Widać również, że oczekiwana liczba szkód odbiega od rzeczywistej liczby zgłoszonych szkód. Jedną z przyczyn tego stanu może być kalkulacja oczekiwanej liczby szkód na poziomie indywidualnym i późniejsza agregacja.

Nieznane parametry modeli wiarygodności są szacowane ważoną metodą najmniejszych kwadratów. W skrócie – parametry są szacowane przez numeryczną minimalizację sumy kwadratów reszt. Wykorzystywana jest metoda Newtona–Raphsona z kryterium zbieżności dla pięciu ostatnich iteracji na poziomie 0,0001. Następujące wyrażenie jest minimalizowane:

$$\sum_{i,j} \omega_{ikj} (N_{ikj} - \lambda_{ikj} F_{ikj})^2 \text{ dla } k = 1,2,$$

gdzie F_{ikj} oznacza estymator dla θ_{ikj}, który zmienia się w zależności od wybranego modelu wiarygodności (i zależy od poszukiwanych parametrów), ω_{ikj} jest długością okresu ochrony dla i-tego klienta w k-tej linii biznesowej podczas j-ego okresu obserwacji. Gdy ustawimy $F_{ikj} \equiv 1$, wtedy nie jest wykorzystywany żaden model wiarygodności i dostajemy model bazowy, który pozwala ocenić poprawę predykcji z wykorzystaniem prezentowanych modeli. Rozważane są modele: jednowymiarowy bez zależności czasowej i z nią oraz dwuwymiarowy bez zależności czasowej oraz z nią. Gdy rozważany jest model jednowymiarowy bez zależności czasowej, wtedy pod F_{ikj} podstawiamy estymator BS_i, który dla k-tej linii biznesowej zależy od parametru σ^2_{kk}. Jeśli do tego dołączymy zależność predykcji od wieku szkody, wtedy F_{ikj} zależy dodatkowo od parametru ρ_{kk}. W przypadku dwuwymiarowym F_{ikj} odpowiada MBS_{ik} z parametrami σ^2_{11}, σ^2_{22} i σ^2_{12}. Dołączenie zależności czasowej w przypadku dwuwymiarowym przekłada się na wprowadzenie do modelu dodatkowych trzech parametrów, mianowicie ρ_{11}, ρ_{22} i ρ_{12}. Oszacowania parametrów wszystkich modeli są przedstawione w tabeli 2. W przypadku dwuwymiarowym oszacowania z modeli jednowymiarowych posłużyły jako wartości startowe dla iteracyjnej metody numerycznej minimalizacji sumy kwadratów reszt.

Oszacowania w tabeli 2 wskazują m.in. na dodatnią korelację między ryzykiem związanym z AC i ryzykiem związanym z OC komunikacyjnym dla tego samego klienta ($\sigma^2_{12} = 0{,}544$), co potwierdza przypuszczenie o zasadności rozważenia podejścia dwuwymiarowego. Widać również pozytywną zależność czasową dla każdego produktu, a nawet pomiędzy nimi ($\rho_{12} = 0{,}628$).

Tabela 2. Oszacowania parametrów analizowanych modeli

Model	σ^2_{11}	σ^2_{22}	σ^2_{12}	ρ_{11}	ρ_{22}	ρ_{12}
Jednowymiarowy, OC, bez czasu	1,687	.	.	.	.	.
z czasem	1,974	.	.	0,569	.	.
Jednowymiarowy, AC, bez czasu	.	1,326	.	.	.	.
z czasem	.	1,406	.	.	0,648	.
Dwuwymiarowy, bez czasu	1,638	1,293	0,544	.	.	.
z czasem	1,752	1,435	0,883	0,483	0,771	0,628

Źródło: opracowanie własne.

Tabela 3. Korelogram dla efektów losowych w AC i OC komunikacyjnym

Opóźnienie h	1	2	3	4	5
$\rho_{11}(h)$	0,114	0,193	0,053	0,324	0,157
$\rho_{22}(h)$	0,330	0,295	0,136	0,581	0,477

Źródło: opracowanie własne.

W tabeli 3 zaprezentowano empiryczne współczynniki autokorelacji dla indywidualnych profili ryzyka. W obu liniach nie potwierdzają one tezy o malejącym geometrycznie w czasie wpływie szkód na predykcję. W obu liniach także szczególnie silny jest wpływ historii szkodowej z okresu *j* – 4 na szkodowość w okresie *j* (dla OC $\rho_{11}(4)=0,324$, dla AC $\rho_{22}(4)=0,581$). Obserwacje te przełożyły się na zastosowanie ogólniejszej postaci struktury czasowej. Jednak poskutkowała ona jedynie zwiększeniem liczby parametrów modeli, a nie przełożyła się na poprawę jakości predykcji.

Na zbiorze testowym (grupa B, 20% danych) przeprowadzono porównanie jakości predykcji każdego z analizowanych modeli. Tabela 4 przedstawia sumy kwadratów reszt wygenerowanych przez poszczególne modele w okresie predykcji.

Tabela 4. Suma kwadratów reszt w próbce testowej

Produkt	Model bazowy	Model jednowymiarowy		Model dwuwymiarowy	
		bez czasu	z czasem	bez czasu	z czasem
OC	137,944	116,220	116,345	115,914	115,450
AC	271,281	254,961	252,500	256,263	252,729

Źródło: opracowanie własne.

Widoczna jest duża różnica między błędami predykcji we wszystkich modelach z nieobserwowanym parametrem ryzyka a błędami w modelu bazowym. Dla produktu OC predykcja jest poprawiona o ok. 15%, a dla AC błąd został zmniejszony o ok. 7%. Dla produktu OC modele dwuwymiarowe są lepsze od jednowymiarowych, ale uwzględnienie struktury czasowej nie wpływa istotnie na jakość predykcji. Natomiast w przypadku drugiego rodzaju ryzyka sytuacja jest odwrotna – uwzględnienie wieku szkody najmocniej wpływa na jakość predykcji, a dodatkowa informacja płynąca z modelu dwuwymiarowego nie zmniejsza istotnie błędów prognoz. Warto wspomnieć również o tym, że rozbudowanie modeli na przypadek wielowymiarowy przekłada się na zwiększenie wariancji oszacowań parametrów, co z biznesowego punktu widzenia nie jest pożądane.

Tabela 5 przedstawia czteroletnią historię szkodową sześciu wybranych klientów. Klienci zostali wybrani tak, by ich historia szkodowa była dosyć urozmaicona (jeden bezszkodowy, szkodowi w różnych momentach czasu czy klienci z różną liczbą szkód), co pozwoli na dokładne porównanie działania poszczególnych modeli. W tabeli są także zaprezentowane oczekiwane liczby szkód dla dwóch produktów w każdym roku. Dla przykładu, klient 2 jest bardziej ryzykowny niż klient 1 w obu liniach biznesowych przez wszystkie lata, co przełożyło się na szkodę zgłoszoną z produktu AC w ostatnim roku.

Tabela 5. Historia szkodowa wybranych klientów

Klient	N_{i1j}	N_{i2j}	λ_{i1j}	λ_{i2j}
1	{0, 0, 0, 0}	{0, 0, 0, 0}	{0,033, 0,033, 0,042, 0,041}	{0,051, 0,052, 0,053, 0,06}
2	{0, 0, 0, 0}	{0, 0, 0, 1}	{0,042, 0,043, 0,05, 0,05}	{0,2, 0,196, 0,193, 0,192}
3	{0, 0, 0, 1}	{0, 0, 0, 0}	{0,055, 0,05, 0,054, 0,053}	{0,08, 0,079, 0,079, 0,075}
4	{0, 0, 0, 0}	{0, 0, 2, 1}	{0,027, 0,03, 0,029, 0,033}	{0,32, 0,33, 0,332, 0,335}
5	{0, 0, 1, 1}	{0, 0, 1, 0}	{0,023, 0,023, 0,02, 0,02}	{0,039, 0,04, 0,042, 0,045}
6	{0, 0, 0, 0}	{3, 0, 0, 0}	{0,016, 0,014, 0,014, 0,015}	{0,04, 0,042, 0,049, 0,038}

Źródło: opracowanie własne.

Tabela 6. Empiryczne estymatory zaufania w modelach wiarygodności dla wybranych klientów

Klient	Linia biznesu	Benchmark	Model jednowymiarowy		Model dwuwymiarowy	
			prosty	z czasem	prosty	z czasem
1	OC	0	0,799	0,837	0,734	0,812
	AC	0	0,777	0,796	0,693	0,743
2	OC	0	0,762	0,825	0,826	0,844
	AC	1,280	1,143	1,284	1,119	1,307
3	OC	4,717	1,979	2,145	1,838	2,083
	AC	0	0,707	0,789	0,893	0,853
4	OC	0	0,833	0,857	1,137	1,141
	AC	2,278	1,813	1,893	1,850	1,907
5	OC	23,256	3,820	4,012	4,016	4,203
	AC	6,024	1,907	1,953	2,801	2,831
6	OC	0	0,909	0,931	2,069	1,534
	AC	17,751	4,068	3,560	4,580	3,781

Źródło: opracowanie własne.

Tabela 6 przedstawia oszacowania $\hat{F}_{ikj}$ dla wybranych sześciu klientów według wszystkich modeli. Kolumna „Benchmark" prezentuje oszacowanie

$\frac{N_{ik.}}{\lambda_{ik.}}$ niewygładzone przez modele. Bezszkodowy klient 1 powinien według wszystkich modeli dostać rabat w obu liniach biznesowych za dotychczasowy przebieg ubezpieczenia. Okazuje się, że wysokość zniżki jest najwyższa w tym przypadku w modelach nieuwzględniających wieku szkody. Klient 2 zgłosił w ostatnim roku szkodę z AC, co według wszystkich modeli przekłada się na zwyżkę składki w tym produkcie. W modelach bez zależności czasowej jest to zwyżka rzędu 12–14%, a modele uwzględniające młody wiek szkody sugerują zwyżkę na poziomie ok. 30%. Zastosowanie modeli dwuwymiarowych w tym przypadku przekłada się na niższą zniżkę w linii bezszkodowej, czyli OC komunikacyjnym. Analogicznie wygląda sytuacja klienta 3 z tą różnicą, że szkoda została zgłoszona z innego produktu. Ciekawy jest przypadek klienta 4, który jest bezszkodowy w OC, ale ma na tyle dużo szkód w AC, że modele dwuwymiarowe szacują dla niego zwyżkę w OC na poziomie ok. 14% składki. Jeszcze silniej ten efekt modeli wielowymiarowych widać na przykładzie klienta 6, dla którego trzy szkody sprzed 4 lat w AC przekładają się na 53,4% zwyżki w OC (model z czasem) lub 106% zwyżki przy nieuwzględnieniu wieku szkód.

5. Podsumowanie

W pracy tej przedstawiono empiryczną analizę zależności między historią szkodową w dwóch liniach biznesowych ubezpieczeń komunikacyjnych, która uwzględnia strukturę czasową zgłaszanych szkód. Historia szkodowa bardzo często znajduje zastosowanie w ustalaniu przyszłych składek przez system bonus-malus. Prezentowane podejście jest nowatorskie ze względu na szersze spojrzenie na ryzyko danego klienta przez wielowymiarowe modele wiarygodności, które uwzględniają korelacje pomiędzy poszczególnymi produktami.

Przedstawione wyniki pokazują duży potencjał biznesowy do jeszcze dokładniejszej selekcji klientów na podstawie przebiegu ich szkodowości. Uwzględnienie zależności czasowej, która nakłada większą karę za szkody zgłoszone niedawno, dostarcza silnego bodźca do ostrożniejszej jazdy. Z kolei mniejsza zwyżka składki za dawne szkody brzmi atrakcyjnie z marketingowego punktu widzenia i może posłużyć do odróżnienia się na konkurencyjnym rynku.

Warto zaznaczyć również to, że najbardziej rozbudowany z prezentowanych modeli – dwuwymiarowy model uwzględniający wiek szkody – najlepiej sprawdził się w predykcji szkodowości na danych rzeczywistych, osiągając niższy błąd prognozy od prostego modelu Bühlmanna–Strauba.

Bibliografia

Bühlmann H., Gisler A., *A course in credibility theory and its applications*, Springer-Verlag, Berlin–Heidelberg–New York 2005.

Desjardins D., Dionne G., Pinquet J., *Experience rating schemes for fleets of vehicles*, „ASTIN Bulletin" 2001, vol. 31, s. 81–105.

Englund M., Guillén M., Gustafsson J., Nielsen L.H., Nielsen J.P., *Multivariate latent risk: A credibility approach*, „ASTIN Bulletin" 2008, vol. 38(1), s. 137–146.

Frees E.W., Young V.R., Luo Y., *Case studies using panel data models*, „North American Actuarial Journal" 2001, vol. 5(4), s. 24–42.

Lemaire J., *Bonus-Malus Systems in Automobile Insurance*, Kluwer Academic Publishers, New York 1995.

Ohlsson E., Johansson B., *Non-life insurance pricing with generalized linear models*, Springer-Verlag, Berlin–Heidelberg 2010.

Pinquet J., Guillén M., Bolancé C., *Allowance for the age of claims in Bonus-Malus systems*, „ASTIN Bulletin" 2001, vol. 31, s. 337–348.

Schnieper R., *Robust Bayesian experience rating*, „ASTIN Bulletin" 2004, vol. 34(1), s. 125–150.

Sobiecki D., *Dwustopniowe modelowanie składki za ubezpieczenie komunikacyjne OC*, „Prace Naukowe" UE we Wrocławiu, z. 312, Wrocław 2013, s. 116–134.

Thuring F., *Multivariate credibility with application to cross-selling financial services products*, niepublikowana praca doktorska, City Univeristy London, 2012.

* * *

Experience rating with dependence between MTPL and MOD claims

Summary

The subject of the article is experience rating with dependence between two lines of vehicle insurance. The paper presents the application of multidimensional credibility models, which are a generalisation of the Bühlmann-Straub model for the age of claim. Conclusions drawn from the analysis of real data portfolio can be applied in cross-selling of MTPL and MOD insurance.

Keywords: cross-selling, experience rating, multidimensional credibility models