PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1108 — 2006
Statystyka aktuarialna – stan i perspektywy rozwoju w Polsce

Helena Jasiulewicz

Politechnika Wrocławska

SKŁADKI ZAUFANIA UWZGLĘDNIAJĄCE WAGĘ KONTRAKTÓW

1. Wstęp

Istotą składki zaufania jest to, że przy jej wyznaczaniu uwzględnia się zarówno dane o ryzyku osiągalne z obserwacji ubezpieczonego w przeszłości, jak i zachowanie się całego portfela w przeszłości. W zależności od danych historycznych polisy, większym lub mniejszym zaufaniem obdarza się indywidualną historię polisy lub historię portfela. Jeżeli wiele danych historycznych jest osiągalnych z obserwacji polisy, to stają się one bardziej godne zaufania i większy jest udział w składce średniej z próby dotyczącej obserwacji polisy z przeszłości, a mniejszy udział w składce średniej z całego portfela. Udział średniej indywidualnej z próby w składce zaufania nazywa się współczynnikiem zaufania do danych historycznych z polisy.

Klasyfikacja ubezpieczonych z uwzględnieniem wiedzy o ryzyku na podstawie indywidualnej obserwacji w przeszłości wymaga modeli niejednorodnych. W modelu niejednorodnym z każdą polisą związany jest parametr losowy Θ, zwany parametrem niejednorodności portfela. Parametr Θ zawiera informacje o ubezpieczonym, które nie były znane ubezpieczycielowi przed zawarciem ubezpieczenia, np. zdolności czy doświadczenie kierującego pojazdem. Te ukryte cechy ubezpieczonego są „odsłaniane" przez dane historyczne osiągalne z obserwacji ubezpieczonego. Wielkość roszczenia lub liczba roszczeń mają rozkłady zależne od parametru Θ. Załóżmy, że historia roszczeń dla i-tej polisy w portfelu jest dana przez szereg czasowy nieujemnych obserwacji $x_{i1}, x_{i2}, \ldots, x_{in_i}$ będących realizacją nieujemnych zmiennych losowych $X_{i1}, X_{i2}, \ldots, X_{in_i}$, gdzie $i = 1, 2, \ldots, N$.

Definicja

Portfelem niejednorodnym nazywamy portfel określony warunkami:

1. Polisa i-ta w portfelu jest opisana przez parę $\left(\Theta_i,\left(X_{ij}\right)_{j\geq 1}\right)$, gdzie Θ_i jest parametrem losowym dla i-tej polisy, a ciąg $\left(X_{ij}\right)_{j\geq 1}$ jest ciągiem wielkości roszczeń lub liczby roszczeń z i-tej polisy w kolejnych okresach ubezpieczenia.

2. Ciąg par $\left(\Theta_1,\left(X_{1j}\right)_{j\geq 1}\right),\left(\Theta_2,\left(X_{2j}\right)_{j\geq 1}\right)\ldots$ jest ciągiem niezależnych zmiennych losowych.

3. Ciąg $\Theta_1,\Theta_2,\ldots$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie.

4. Dla ustalonego $\Theta_i=\theta_i$, ciąg $X_{i1},X_{i2},\ldots$ jest ciągiem zmiennych losowych warunkowo niezależnych o rozkładach warunkowych zależnych od θ_i.

Z warunków modelu wynika, że polisy generują straty niezależnie od siebie, a tym samym historia każdej polisy nie zależy od historii innych polis. Zależność jest tylko możliwa pomiędzy roszczeniami generowanymi przez polisę w kolejnych okresach ubezpieczenia.

Jeżeli składka dla i-tej polisy z opisanego portfela niejednorodnego ma uwzględniać dane historyczne dotyczące tej polisy, to nie jest sensowne, aby składkę indywidualną określać jako $\mathrm{E}X_{ij}$. Przy założeniu, że istnieje $\mathrm{E}X_{ij}$, naturalnym surogatem jest warunkowa wartość oczekiwana $\mathrm{E}\left(X_{ij}\mid\theta_i\right)$ oznaczana przez $m(\theta_i)$. Składka $m(\theta_i)$ dla i-tej polisy jest nieznana, bo nieznany jest parametr θ_i. Z mocnego prawa wielkich liczb wynika, że średnia arytmetyczna z historii polisy zmierza do $m(\theta_i)$, gdy $n_i\to\infty$. Zatem średnia arytmetyczna z obserwacji polisy może być rozważana jako jedna z możliwych aproksymacji dla $m(\theta_i)$. Innym możliwym estymatorem wykorzystującym dane historyczne o polisie jest estymator bayesowski z kwadratową funkcją straty. Z teorii estymacji bayesowskiej wiadomo, że ma on postać $\mathrm{E}\left[m(\Theta_i)\mid X_{i1},\ldots,X_{in_i}\right]$ i jego przeciętny błąd kwadratowy, wynosi $\mathrm{E}\,\mathrm{Var}\left[m(\Theta_i)\mid X_{i1},\ldots,X_{in_i}\right]$. Estymator bayesowski niekoniecznie będzie liniową funkcją danych. Dlatego estymatorów dla $m(\theta_i)$ mających najmniejszy przeciętny błąd kwadratowy będzie się poszukiwać w węższej klasie funkcji od obserwacji, a mianowicie w klasie funkcji liniowych. Taki estymator będzie nazywany liniowym estymatorem bayesowskim dla $m(\theta_i)$. Okazuje się, że liniowy estymator bayesowski jest wypukłą kombinacją średniej z próby będącej obserwacją ubezpieczonego w przeszłości i średniej z obserwacji całego portfela. Z tego powodu liniowy estymator bayesowski nazywany jest również estymatorem zaufania lub składką zaufania.

2. Obserwacje z wagami

Załóżmy, że polisy w portfelu niejednorodnym wykazują wewnętrzną niejednorodność w czasie. Oznacza to, że warunkowe rozkłady zmiennych losowych $X_{i1}, X_{i2}, \ldots$ przy warunku $\Theta_i = \theta_i$ mogą mieć różne rozkłady. Do każdej zmiennej losowej X_{ij} będzie podany jej wolumen wyrażony przez dodatkowy parametr nieujemny w_{ij}. Wielkości w_{ij} są znane i nazywane są wagami. Jeżeli wagi przyjmują wartości całkowite nieujemne, to nazywa się je wagami naturalnymi.

Jest wiele możliwości wyboru wag w_{ij}, zwanych też wolumenem ryzyka w i-tej klasie polis w j-tym roku. Niech Y_{ij} oznacza absolutną wielkość, będącą zmienną losową obserwowaną i oznaczającą całkowitą liczbę roszczeń albo całkowitą kwotę roszczeń w i-tej klasie polis w j-tym roku. Niech X_{ij} oznacza względną wielkość określoną jako

$$X_{ij} = \frac{Y_{ij}}{w_{ij}}.$$

W celu ustalenia uwagi Y_{ij} będą oznaczały całkowitą kwotę roszczenia, a w celu uproszczenia zapisu opuścimy pierwszy indeks i.

1. Niech w_j oznacza liczbę osób ubezpieczonych (ogólnie liczbę jednostek ryzyka) w j-tym roku. Wówczas całkowitą kwotę roszczenia można przedstawić następująco:

$$Y_j = \sum_{l=1}^{w_j} \sum_{k=1}^{N_l} Z_{jk}^{(l)}, \tag{1}$$

gdzie N_l oznacza liczbę roszczeń z l-tej polisy powodujących całkowitą kwotę roszczeń z tej polisy równą $\sum_{k=1}^{N_l} Z_{jk}^{(l)}$. Zmienna losowa $Z_{jk}^{(l)}$ oznacza kwotę k-tego roszczenia z l-tej polisy w j-tym roku. Zmienne losowe $Z_{jk}^{(l)}$ są zmiennymi niezależnymi o jednakowym rozkładzie, takim jak zmienna losowa Z. Załóżmy, że dla $\Theta = \theta$ liczba roszczeń N_l ma rozkład Poissona z parametrem θ dla każdego l oraz zmienne $Z_{jk}^{(l)}$ są niezależne od Θ oraz N_l. Wielkość $X_j = Y_j / w_j$ oznacza przeciętną całkowitą kwotę roszczenia z polisy w l-tym roku (*loss ratio*). Wówczas jej warunkowa wartość oczekiwana wynosi

$$\mathrm{E}(X_j \mid \theta) = \frac{1}{w_j}\mathrm{E}(Y_j \mid \theta) = \frac{1}{w_j} w_j \mathrm{E}(N_l \mid \theta)\mathrm{E}Z_{jk}^{(l)} = m(\theta), \tag{2}$$

gdzie $m(\theta)=\theta \mathrm{E} Z$. Odnotujmy, że warunkowa wariancja zmiennej losowej X_j przy warunku $\Theta=\theta$ wynosi

$$\operatorname{Var}\left(X_j \mid \theta\right)=\left(\frac{1}{w_j}\right)^2 w_j \theta \mathrm{E}\left(Z_{jk}^{(l)}\right)^2=\frac{\theta \mathrm{E} Z^2}{w_j}=\frac{s^2(\theta)}{w_j}, \tag{3}$$

gdzie funkcja $s^2(\theta)=\theta \mathrm{E} Z^2$.

2. Niech w_j oznacza liczbę przejechanych kilometrów przez kierowcę w j-tym roku. Całkowita kwota roszczenia dla tego kierowcy wynosi

$$Y_j=Z_{j1}+\ldots+Z_{jN_j}, \tag{4}$$

gdzie Z_{jk} oznacza wielkość k-tego roszczenia, a N_j oznacza liczbę roszczeń zgłoszonych przez kierowcę. Jeśli przyjmiemy, że zmienne losowe Z_{jk} są niezależnymi zmiennymi losowymi o jednakowych rozkładach takich jak zmienna losowa Z oraz że są niezależne od zmiennej losowej N_j oraz od zmiennej losowej Θ, to otrzymuje się wzór na warunkową wartość oczekiwaną zmiennej losowej $X_j=Y_j/w_j$ postaci

$$\mathrm{E}\left(X_j \mid \theta\right)=\mathrm{E}\left(\frac{Y_j}{w_j}\right)=\frac{1}{w_j} \mathrm{E} Z_{jk} \mathrm{E}\left(N_j \mid \theta\right).$$

Załóżmy, że dla $\Theta=\theta$ liczba roszczeń N_j ma rozkład Poissona z parametrem równym $w_j\theta$. Wówczas $\mathrm{E}\left(X_j \mid \theta\right)=m(\theta)$, gdzie $m(\theta)=\theta \mathrm{E} Z$. Składka netto dla kierowcy przejeżdżającego w_j kilometrów w roku wynosi $w_j m(\theta)$. Analogicznie pokazuje się, że $\operatorname{Var}\left(X_j \mid \Theta\right)=s^2(\theta)/w_j$, gdzie $s^2(\theta)=\theta \mathrm{E} Z^2$.

3. Niech w_j oznacza część j-tego roku, w którym ubezpieczony był objęty ochroną ubezpieczeniową (*years at risk*). Waga ta nie jest wagą naturalną, bo $0<w_j \leq 1$. Całkowita kwota roszczenia z polisy w j-tym roku wynosi $Y_j=Z_{j1}+\ldots+Z_{jN_j}$, gdzie Z_{jk} oraz N_j oznaczają to samo, co w poprzednim punkcie i spełniają te same założenia. Wówczas zmienna losowa $X_j=Y_j/w_j$ ma warunkową wartość oczekiwaną $\mathrm{E}\left(X_j \mid \theta\right)=m(\theta)$, a warunkową wariancję $\operatorname{Var}\left(X_j \mid \theta\right)=s^2(\theta)/w_j$, gdzie funkcje $m(\theta)$ i $s^2(\theta)$ są takie same jak w punkcie 2. Składka netto za część roku równą w_j wynosi $w_j m(\theta)$.

4. Niech w_j oznacza wiek polisy w j-tym roku (*policy years*). Wówczas całkowitą kwotę roszczenia Y_j określamy wzorem:

$$Y_j = \sum_{i=1}^{w_j} \left(Z_{i1}^{(j)} + \ldots + Z_{iN_i}^{(j)} \right), \tag{5}$$

gdzie Z_{ik} oznacza k-te roszczenie z polisy w wieku i, N_i oznacza zaś liczbę roszczeń z polisy w wieku i mającą rozkład warunkowy Poissona z parametrem θ. Przyjmujemy założenia o zmiennych Z_{ik} takie jak w punkcie 2. Wówczas $\mathrm{E}\left(Y_j \mid \theta\right) = w_j \theta \mathrm{E} Z$, zaś dla zmiennej losowej $X_j = Y_j / w_j$, warunkowa wartość oczekiwana $\mathrm{E}\left(X_j \mid \theta\right) = \theta \mathrm{E} Z$. Natomiast $\mathrm{Var}\left(X_j \mid \theta\right) = \theta \mathrm{E} Z^2 / w_j = s^2(\theta)/w_j$, gdzie $s^2(\theta) = \theta \mathrm{E} Z^2$.

5. Niech w_j oznacza całkowitą sumę ubezpieczenia w j-tym roku. Wówczas sumaryczna kwota roszczenia wynosi $Y_j = \sum_{k=1}^{N_j} Z_{jk}$ przy założeniach takich jak w punkcie 3. Przeciętna wielkość szkody przypadająca na jedną jednostkę sumy ubezpieczenia $X_j = Y_j / w_j$ ma warunkową wartość oczekiwaną i warunkową wariancję taką samą jak w punkcie 3.

W praktyce często wykorzystuje się wagę w_j jako całkowity dochód ze składek w j-tym roku, jednakże ta waga nie jest wygodna, bo zmniejszenie tempa wpływania składek nie zmniejsza ryzyka (zob. [2]).

3. Estymator zaufania uwzględniający wolumen ryzyka

Załóżmy, że przy ustalonym $\Theta_i = \theta_i$ warunkowe wartości oczekiwane i warunkowe wariancje zmiennych losowych X_{ij} są funkcjami θ_i:

$$\mathrm{E}\left(X_{ij} \mid \theta_i\right) = m(\theta_i) \tag{6}$$

oraz

$$\mathrm{Var}\left(X_{ij} \mid \theta_i\right) = \frac{s^2(\theta_i)}{w_{ij}}, \tag{7}$$

dla $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, a stałe w_{ij} są znane. Model niejednorodny zdefiniowany warunkami 1-4 oraz wzorami (6) i (7) nazywa się modelem Bühlmanna-Strauba. Odnotujmy, że ciąg $m(\Theta_1), m(\Theta_2), \ldots$ jest ciągiem zmiennych losowych niezależnych o jednakowych rozkładach oraz ciąg $s^2(\Theta_1), s^2(\Theta_2),\ldots$ jest również ciągiem zmiennych losowych niezależnych o jednakowych rozkładach. Wynika to z faktu, że parametry losowe $\Theta_1, \Theta_2,\ldots$ są zmiennymi losowymi niezależnymi o jednakowym rozkładzie.

Bez straty ogólności można przyjąć, że historie każdej polisy są tej samej długości, tzn. $n_i = n$ dla każdego i. Dotychczas używana nazwa „i-ta polisa" zostanie rozszerzona na nazwę „i-ty kontrakt" lub „i-ta grupa polis". Polisy w danej grupie są jednorodne.

Będziemy poszukiwać najlepszego w sensie najmniejszego przeciętnego błędu kwadratowego, liniowego estymatora średniej $m(\Theta_i) = \mathrm{E}\left(X_{i\,n+1} \mid \Theta_i\right)$ dla i-tego kontraktu w roku $n+1$. W tym celu wprowadźmy następujące oznaczenia:

– indywidualna średnia ważona i-tej grupy polis:

$$\overline{X}_{iw} = \sum_{j=1}^{n} \frac{w_{ij}}{w_i} X_{ij}, \quad \text{gdzie} \quad w_i = \sum_{j=1}^{n} w_{ij}, \tag{8}$$

– ogólna średnia ważona:

$$\overline{X}_{ww} = \sum_{i=1}^{N} \frac{w_i}{w} \overline{X}_{iw}, \quad \text{gdzie} \quad w = \sum_{i=1}^{N} w_i, \tag{9}$$

– parametry struktury ryzyka w portfelu:

$$\mu = \mathrm{E} X_{ij} = \mathrm{E} m(\Theta_i), \quad \varphi = \mathrm{E} s^2(\Theta_i), \quad \psi = \mathrm{Var}\left(m(\Theta_i)\right). \tag{10}$$

Liniowy estymator bayesowski dla $m(\theta_i) = \mathrm{E}\left(X_{i\,n+1} \mid \Theta_i = \theta_i\right)$ jest podany w twierdzeniu 1.

Twierdzenie 1

W modelu Bühlmanna-Strauba estymator zaufania $\tilde{m}_i$ średniej $m(\Theta_i)$ dla $i = 1,\ldots,N$ ma postać

$$\tilde{m}_i = Z_i \overline{X}_{iw} + (1 - Z_i)\mu, \tag{11}$$

gdzie współczynnik zaufania Z_i dany jest wzorem

$$Z_i = \frac{w_i \psi}{w_i \psi + \varphi}. \tag{12}$$

Dowód

Estymator zaufania $\tilde{m}_i$ jest jedynym rozwiązaniem równań normalnych

$$E(m(\Theta_i)-\tilde{m}_i)=0,$$

$$\mathrm{Cov}(m(\Theta_i)-\tilde{m}_i, X')=0,$$

gdzie $X_i'=(X_{i1},\ldots,X_{in})$ (zob. [3]). Szczegółowe wyprowadzenie estymatora zaufania o postaci (11) znajduje się m.in. w książce [3].

Twierdzenie 2

W modelu Bühlmanna-Strauba przeciętny błąd kwadratowy estymatora zaufania $\tilde{m}_i$ średniej $m(\Theta_i)$ jest równy

$$\mathrm{MSE}_i = E(m(\Theta_i)-\tilde{m}_i)^2=(1-Z_i)\psi, \tag{13}$$

dla $i=1,\ldots,N$.

Dowód w książce [3].

Uwaga

Gdy średnia kolektywna μ jest nieznana i zastąpi się ją przez estymator o postaci

$$\overline{X}_{zw}=\sum_{i=1}^{N}\frac{Z_i}{Z}\overline{X}_{iw}, \tag{14}$$

gdzie $Z=Z_1+\ldots+Z_N$, to wzór (11) przybiera postać

$$\tilde{m}_i^*=Z_i\overline{X}_{iw}+(1-Z_i)\overline{X}_{zw}. \tag{15}$$

Estymator $\tilde{m}_i^*$ nazywany jest jednorodnym estymatorem zaufania dla $m(\Theta_i)$. Jego przeciętny błąd kwadratowy wyraża się wzorem

$$\mathrm{MSE}_i^* = E(m(\Theta_i)-\tilde{m}_i^*)^2=(1-Z_i)\psi\left(1+\frac{1-Z_i}{Z}\right). \tag{16}$$

4. Estymacja parametrów struktury

Niejednorodny estymator zaufania $\tilde{m}_i$ i jednorodny estymator zaufania $\tilde{m}_i^*$ zależą od parametrów struktury

$$\mu=\mathrm{E}m(\Theta_i),\quad \varphi=\mathrm{E}s^2(\Theta_i),\quad \psi=\mathrm{Var}(m(\Theta_i)),$$

które na ogół są nieznane. Wówczas należy znaleźć estymatory tych wielkości na podstawie danych o całym portfelu. Wprowadźmy następujące oznaczenia:

– ważona suma kwadratów odchyleń wewnątrz grupy:

$$\mathrm{WSS}_W = \sum_{i=1}^{N} \sum_{j=1}^{n} w_{ij} \left(X_{ij} - \bar{X}_{iw} \right)^2,$$

– średnia ważona suma kwadratów wewnątrz grupy:

$$\mathrm{MWS}_W = \frac{\mathrm{WSS}_W}{(n-1)N},$$

– ważona suma kwadratów odchyleń pomiędzy grupami:

$$\mathrm{WSS}_B = \sum_{i=1}^{N} w_i \left(\bar{X}_{iw} - \bar{X}_{ww} \right)^2,$$

– średnia ważona suma kwadratów pomiędzy grupami:

$$\mathrm{MWS}_B = \frac{\mathrm{WSS}_B}{N-1}.$$

Twierdzenie 3

W modelu Bühlmanna-Strauba estymatory

- $\hat{\mu} = \bar{X}_{zw}$,
- $\hat{\varphi} = \mathrm{MWS}_W$,
- $\hat{\psi} = \frac{w(N-1)(\mathrm{MWS}_B - \mathrm{MWS}_W)}{w^2 - \sum_{i=1}^{N} w_i^2}$

są nieobciążonymi estymatorami parametrów struktury odpowiednio dla μ, φ, ψ.

Dowód w książce [3].

Za estymator współczynnika Z_i przyjmuje się

$$\hat{Z}_i = \frac{w_i \hat{\psi}}{w_i \hat{\psi} + \hat{\varphi}}, \tag{17}$$

gdzie $\hat{\psi}$ i $\hat{\varphi}$ są takie same jak w twierdzeniu 3. Estymator $\hat{\psi}$ może przyjmować wartości ujemne, wówczas za oszacowanie parametru ψ przyjmuje się wartość estymatora $\max(0, \hat{\psi})$. Oczywiście wtedy nie jest on już estymatorem nieobciążonym.

Jeżeli w niejednorodnym estymatorze zaufania $\tilde{m}_i = Z_i \bar{X}_{iw} + (1 - Z_i)\mu$ współczynnik zaufania Z_i oszacuje się przez $\hat{Z}_i$ dany wzorem (17), a średnią kolektywną μ przez $\bar{X}_{zw}$ daną wzorem (14), to estymator

$$\tilde{m}_i^* = \hat{Z}_i \bar{X}_{iw} + \left(1 - \hat{Z}_i\right) \bar{X}_{zw} \tag{18}$$

nazywa się empirycznym estymatorem zaufania dla i-tego kontraktu w modelu Bühlmanna-Strauba.

Model Bühlmanna-Strauba można również stosować przy brakujących obserwacjach, tzn. wówczas, gdy niektóre x_{ij} są nieosiągalne albo tak wyjątkowe, że nie są brane pod uwagę. Niech T_i oznacza zbiór indeksów j, dla których obserwacje x_{ij} istnieją, $i = 1,\ldots,N$. Wówczas sumowanie po j we wcześniejszych wzorach odbywa się nie od 1 do n, ale po $j \in T_i$. Zatem

$$w_i = \sum_{j \in T_i} w_{ij}, \quad \bar{X}_{iw} = \sum_{j \in T_i} \frac{w_{ij}}{w_i} X_{ij},$$

$$\mathrm{WSS}_W = \sum_{i=1}^{N} \sum_{j \in T_i} w_{ij} \left(X_{ij} - \bar{X}_{iw}\right)^2, \quad \mathrm{MWS}_W = \frac{\mathrm{WSS}_W}{t},$$

gdzie

$$t = \sum_{i=1}^{N} (t_i - 1),$$

a t_i oznacza liczbę elementów w zbiorze T_i. Sformułowane wcześniej twierdzenia pozostają bez zmian.

5. Przykład

Prezentowany przykład będzie dotyczył prognozowania składki netto w ubezpieczeniach komunikacyjnych za rok $n+1$ na podstawie przeciętnej wielkości roszczeń z polisy w latach wcześniejszych. Zmienna losowa Λ, opisująca strukturę portfela, będzie miała przesunięty rozkład gamma. Będzie to odpowiadało sytuacji, gdy kierowcy o bardzo małej szkodowości nie ubezpieczają się. Przykład będzie dotyczył wyznaczania składki netto, gdy przeciętne wielkości roszczeń są duże i mają rozkład Pareto.

Niech obserwowane zmienne losowe X_{ij}, gdzie $i = 1,\ldots,N$, $j = 1,\ldots,n$, będą średnimi arytmetycznymi z w_{ij} zmiennych losowych $Y_{ij}^{(1)}, Y_{ij}^{(2)},\ldots,Y_{ij}^{(w_{ij})}$, niezależnych, o jednakowych rozkładach. Na przykład $Y_{ij}^{(k)}$ można interpretować jako całkowite roczne roszczenie z k-tej polisy należącej do i-tej grupy polis w j-tym roku trwania ubezpieczenia. Wówczas

$$X_{ij} = \frac{1}{w_{ij}} \sum_{k=1}^{w_{ij}} Y_{ij}^{(k)} \tag{19}$$

przedstawia przeciętne roczne roszczenie z polisy w j-tym roku należącej do i-tej grupy polis. Natomiast w ciągu roku polisa generuje losową liczbę $N(\lambda_i)$ roszczeń o wielkościach $Z_1, Z_2, \ldots, Z_{N(\lambda_i)}$, gdzie λ_i jest parametrem ryzyka będącym realizacją zmiennej losowej Λ_i, Zmienne losowe $\Lambda_1, \Lambda_2, \ldots$ są niezależnymi kopiami zmiennej losowej Λ o dystrybuancie $U(\lambda)$. Rozkład $U(\lambda)$ przedstawia strukturę ryzyka rozważanego portfela polis. Zatem całkowita roczna wielkość roszczenia z k-tej polisy należącej do i-tej grupy polis wynosi

$$Y_{ij}^{(k)} = Z_1^{(k)} + Z_2^{(k)} + \ldots + Z_{N(\lambda_i)}^{(k)}, \tag{20}$$

gdzie $Z_l^{(k)}$ są niezależnymi kopiami zmiennej losowej Z o rozkładzie F dla wszystkich l i k. W przykładzie będzie się zakładać, że zmienna losowa N, przedstawiająca liczbę roszczeń z polisy w ciągu roku, ma rozkład Poissona z parametrem λ_i, gdy $\Lambda_i = \lambda_i$. Przy przyjętych wyżej założeniach parametry struktury ryzyka w portfelu μ, φ i ψ przedstawiają się następująco:

$$\mu = (\mathrm{E}Z)(\mathrm{E}\Lambda), \tag{21}$$

$$\varphi = (\mathrm{E}Z^2)(\mathrm{E}\Lambda), \tag{22}$$

$$\psi = (\mathrm{E}Z)^2(\mathrm{Var}\,\Lambda). \tag{23}$$

Wzór (21) wynika z następujących przekształceń:

$$\mu = \mathrm{E}X_{ij} = \mathrm{E}\big(\mathrm{E}(X_{ij} \mid \Lambda_i)\big) = \mathrm{E}\left(\frac{1}{w_{ij}} \sum_{k=1}^{w_{ij}} \mathrm{E}\left(Y_{ij}^{(k)} \mid \Lambda_i\right)\right) =$$

$$= \mathrm{E}\left(\mathrm{E}\left(Y_{ij}^{(k)} \mid \Lambda_i\right)\right) = \mathrm{E}\left(\mathrm{E}\left(\sum_{l=1}^{N(\Lambda_i)} Z_l \mid \Lambda_i\right)\right).$$

Ponieważ

$$\mathrm{E}\left(\sum_{l=1}^{N(\Lambda_i)} Z_l \mid \Lambda_i\right) = \Lambda_i \mathrm{E}Z,$$

$\mu = (\mathrm{E}\Lambda_i)(\mathrm{E}Z)$. Wzór (22) uzasadnia się następująco:

$$\varphi = \mathrm{E}\,\mathrm{Var}\left(Y_{ij} \mid \Lambda_i\right) = \mathrm{E}\,\mathrm{Var}\left(\sum_{l=1}^{N(\Lambda_i)} Z_l \mid \Lambda_i\right) = \mathrm{E}\left(\Lambda_i \mathrm{Var}\,Z + \Lambda_i (\mathrm{E}Z)^2\right).$$

Natomiast wzór (23) wynika ze wzoru $\mathrm{E}\left(X_{ij} \mid \Lambda_i\right) = \Lambda_i \mathrm{E}Z$, ponieważ parametr struktury $\varphi = \mathrm{Var}\,\mathrm{E}\left(X_{ij} \mid \Lambda_i\right)$.

Poniżej zostanie opisany sposób generowania danych x_{ij}, które są realizacjami zmiennych X_{ij} przy znanych wagach w_{ij} będących liczbami naturalnymi. Przedział (a,b), będący nośnikiem gęstości $u(\lambda)$ zmiennej losowej Λ, dzieli się na N rozłącznych przedziałów $I_1 = (a,a_1], I_2 = (a_1,a_2], \ldots, I_N = (a_{N-1},b)$, gdzie $a < a_1 < a_2 < \ldots < a_{N-1} < b$. Z każdego przedziału I_i wybiera się liczbę λ_i według wzoru $\lambda_i = \mathrm{E}(\Lambda \mid \Lambda \in I_i)$.

Dla ustalonego i oraz j generuje się w_{ij} liczb losowych według rozkładu Poissona z parametrem λ_i. Wygenerowane liczby oznaczmy przez $k_l^{(i)}$, gdzie $l = 1, 2, \ldots, w_{ij}$. Liczbę $k_l^{(i)}$ można interpretować jako całkowitą liczbę roszczeń w ciągu j-tego roku z l-tej polisy należącej do i-tej grupy polis. Następnie dla ustalonego l generuje się $k_l^{(i)}$ liczb według rozkładu F zmiennej losowej Z. Wygenerowane liczby $z_1^{(l)}, \ldots, z_{k_l(i)}^{(l)}$ przedstawiają wielkości kolejnych roszczeń z l-tej polisy w ciągu roku. Suma

$$y_{ij}^{(l)} = z_1^{(l)} + \ldots + z_{k_l^{(i)}}^{(l)}$$

wyraża całkowitą wielkość roszczenia z l-tej polisy w ciągu roku. Średnia arytmetyczna x_{ij} z liczb $y_{ij}^{(1)}, y_{ij}^{(2)}, \ldots, y_{ij}^{(w_{ij})}$ jest realizacją zmiennej losowej X_{ij}, czyli

$$x_{ij} = \frac{y_{ij}^{(1)} + y_{ij}^{(2)} + \ldots + y_{ij}^{(w_{ij})}}{w_{ij}}. \tag{24}$$

Dane x_{ij} zostaną wygenerowane przy założeniu, że zmienna losowa Λ ma przesunięty rozkład gamma z parametrami α, β, γ, gdzie γ jest parametrem przesunięcia, oraz że zmienna losowa Z ma rozkład Pareto. W rozkładach zmiennych Λ i Z przyjęto następujące parametry:

$$\gamma = 0,05,$$

$$\mathrm{E}\Lambda = \frac{\alpha}{\beta} + \gamma = 0,2,$$

$$\mathrm{Var}\,\Lambda = \frac{\alpha}{\beta^2} = 0,01,$$

$$\mathrm{E}Z = 15,$$

$$\sqrt{\mathrm{Var}\,Z} = 8.$$

Przyjęte założenia mogą być stosowane np. przy wyznaczaniu składki w ubezpieczeniach komunikacyjnych AC z dużymi wypłatami.

Wagi naturalne w_{ij} są generowane według rozkładu jednostajnego ze zbioru $\{1, 2, ..., 100\}$. Dla $N = 12$, $n = 7$ wygenerowane dane x_{ij} wraz z ich wagami w_{ij} są zamieszczone w tab. 1.

Tabela 1. Przeciętna roczna wielkość roszczenia z polisy

i \ j	1	2	3	4	5	6	7
1	2,57 29	0,941 70	0 2	0 32	1,61 52	1,86 48	0,895 36
2	2,55 27	1,58 17	0,818 90	1,28 51	0,901 41	1,72 95	2,78 49
3	3,61 33	2,93 24	2,08 96	2,44 78	1,76 56	1,11 47	3,39 11
4	2,99 68	2,76 78	2,5 24	2,67 57	2,71 81	6,29 2	1,95 76
5	1,7 57	3,98 50	1,42 47	3,73 38	1,17 93	3,94 34	2,92 10
6	3,67 3	3,93 66	3,34 84	1,73 94	2,17 79	0 11	0,587 27
7	1,71 96	1,83 45	4,16 39	1,45 78	1,73 21	2,68 88	0 1
8	0 2	3,52 87	3,85 84	4,15 47	2,69 90	1,81 79	1,57 38
9	4,28 94	3,08 96	1,93 14	4,68 38	4,51 56	1,36 25	2,71 66
10	3,97 26	3,28 88	11,1 1	6,37 20	3,46 21	3,49 55	4,48 16
11	5,03 67	4,65 14	3,86 89	4,21 62	6,8 13	11,6 24	3,62 36
12	5,69 89	5,24 39	9,81 29	8,39 75	6,38 43	6,16 77	5,84 92

Źródło: opracowanie własne.

Tabela 2. Wartości estymatorów parametrów struktury

Parametry struktury	μ			φ	ψ
	3			57,8	2,25
Oszacowania parametrów	$\bar{X}_{zw}$	$\bar{X}_{ww}$	$\bar{X}_{\dot{z}w}$	$\hat{\varphi}$	$\hat{\psi}$
	3,04	3,1	3,04	66,1	2,22

Źródło: opracowanie własne.

Tabela 3. Wartości estymatorów zaufania

i	m_i	$\tilde{m}_i$	$\sqrt{MSE_i}$	$\tilde{m}_i^*$	$\sqrt{MSE_i^*}$	$\overline{X}_{iw}$	Z_i
1	1,16	1,43	0,443	1,44	0,445	1,12	0,913
2	1,51	1,64	0,382	1,64	0,383	1,66	0,935
3	1,78	2,28	0,395	2,28	0,396	2,47	0,931
4	2,03	2,65	0,375	2,65	0,376	3,13	0,938
5	2,29	2,41	0,404	2,41	0,405	2,69	0,928
6	2,55	2,51	0,385	2,51	0,386	2,2	0,934
7	2,83	2,21	0,383	2,21	0,384	1,94	0,935
8	3,15	2,97	0,357	2,98	0,358	2,51	0,943
9	3,54	3,49	0,373	3,49	0,374	3,22	0,938
10	4,03	3,73	0,478	3,74	0,48	5,16	0,898
11	4,74	4,79	0,418	4,79	0,42	5,69	0,922
12	6,48	6,36	0,351	6,36	0,352	6,79	0,945

Źródło: opracowanie własne.

Dla danych z tab. 1 wyznacza się wartości estymatorów parametrów struktury. Wyniki obliczeń wraz z ich teoretycznymi odpowiednikami są podane w tab. 2.

Dla wygenerowanych danych z tab. 1 zostały obliczone średnie grupowe $\overline{x}_{iw}$ współczynniki zaufania Z_i, jednorodne i niejednorodne estymatory zaufania $\tilde{m}_i$, $\tilde{m}_i^*$ oraz odpowiednio ich błędy $\sqrt{MSE_i}$ i $\sqrt{MSE_i^*}$. Wyniki tych obliczeń podano w tab. 3.

Dla danych z tab. 1 zostały wyznaczone wartości estymatorów $\hat{Z}_i$, a następnie wartości empirycznych estymatorów zaufania $\hat{m}_i$ oraz $\hat{m}_{iw}$, gdzie $\hat{m}_i = \hat{Z}_i\overline{X}_{iw} + \left(1-\hat{Z}_i\right)\overline{X}_{zw}$ oraz $\hat{m}_{iw} = \hat{Z}_i\overline{X}_{iw} + \left(1-\hat{Z}_i\right)\overline{X}_{ww}$. Wyniki tych obliczeń są zawarte w tab. 4.

Tabela 4. Wartości empirycznych estymatorów zaufania

i	m_i	$\hat{m}_i$	$\hat{m}_{iw}$	$\hat{Z}_i$	Z_i
1	1,16	1,46	1,46	0,9	0,913
2	1,51	1,65	1,66	0,926	0,935
3	1,78	2,29	2,29	0,921	0,931
4	2,03	2,65	2,66	0,928	0,938
5	2,29	2,42	2,42	0,917	0,928
6	2,55	2,52	2,52	0,924	0,934
7	2,83	2,22	2,23	0,925	0,935
8	3,15	2,98	2,98	0,935	0,943
9	3,54	3,48	3,49	0,929	0,938
10	4,03	3,73	3,73	0,884	0,898
11	4,74	4,77	4,77	0,911	0,922
12	6,48	6,33	6,34	0,937	0,945

Źródło: opracowanie własne.

W celu oszacowania przeciętnego błędu kwadratowego estymatora $\hat{m}_i$ wielkości $\mathrm{E}\left(X_{ij} \mid \lambda_i\right)$ powtórzono 500 razy generowanie tabel z danymi x_{ij}. Niech $\hat{m}_i^{(j)}$ będzie wartością $\hat{m}_i$ dla j-tej symulacji dla $j = 1, ..., 500$. Przeciętny błąd kwadratowy estymatora $\hat{m}_i$ wyznacza się według wzoru

$$e_i^2 = \frac{1}{500}\sum_{j=1}^{500}\left(m_i - \hat{m}^{(j)}\right)^2.$$

Przeciętny błąd kwadratowy estymatora $\hat{m}_{iw}$ wyznacza się według wzoru

$$e_{iw}^2 = \frac{1}{500}\sum_{j=1}^{500}\left(m_i - \hat{m}^{(j)}\right)^2.$$

Wyniki obliczeń dla e_i oraz e_{iw} wraz z empiryczną wartością oczekiwaną i empiryczną wariancją estymatorów zaufania $\hat{m}_i$ oraz $\hat{m}_{iw}$ są zawarte w tab. 5.

Tabela 5. Błąd empirycznego estymatora zaufania

i	m_i	$\mathrm{E}\hat{m}_i$	$\sqrt{\mathrm{Var}\,\hat{m}_i}$	e_i	$\mathrm{E}\hat{m}_{iw}$	$\sqrt{\mathrm{Var}\,\hat{m}_{iw}}$	e_{iw}
1	1,16	1,26	0,222	0,245	1,26	0,222	0,245
2	1,51	1,58	0,26	0,268	1,58	0,26	0,268
3	1,78	1,85	0,278	0,286	1,85	0,278	0,286
4	2,03	2,08	0,274	0,279	2,08	0,275	0,279
5	2,29	2,3	0,308	0,308	2,3	0,308	0,308
6	2,55	2,58	0,339	0,34	2,58	0,338	0,34
7	2,83	2,85	0,358	0,359	2,85	0,358	0,359
8	3,15	3,15	0,401	0,401	3,15	0,401	0,401
9	3,54	3,5	0,422	0,424	3,5	0,421	0,423
10	4,03	3,97	0,374	0,38	3,97	0,374	0,38
11	4,74	4,68	0,508	0,511	4,68	0,507	0,511
12	6,48	6,33	0,498	0,522	6,32	0,498	0,523

Źródło: opracowanie własne.

Globalny błąd względny estymatorów empirycznych $\hat{m}_i$ obliczony według wzoru

$$\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} e_i^2} - \sqrt{\frac{1}{N}\sum_{i=1}^{N}\mathrm{MSE}_i^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\mathrm{MSE}_i^2}}100\%$$

dla danych e_i z tab. 4 jest bliski zeru. Podobnie globalny błąd względny estymatorów empirycznych $\hat{m}_{iw}$ obliczony według wzoru

$$\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} e_{iw}^2} - \sqrt{\frac{1}{N}\sum_{i=1}^{N} \mathrm{MSE}_i^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} \mathrm{MSE}_i^2}} 100\%$$

dla danych e_{iw} z tab. 4 jest również bliski zeru. Wyniki obliczeń dla e_i oraz e_{iw}, wraz z empiryczną wartością oczekiwaną i empiryczną wariancją estymatorów zaufania $\hat{m}_i$ oraz $\hat{m}_{iw}$ są zawarte w tab. 5. Wartości oczekiwane i wariancje estymatorów $\hat{\varphi}$ i $\hat{\psi}$, wyznaczone jako empiryczne wartości oczekiwane i empiryczne wariancje, są podane w tab. 6.

Tabela 6. Symulacyjna ocena estymatorów $\hat{\varphi}$ i $\hat{\psi}$

φ	$E\hat{\varphi}$	$\sqrt{\mathrm{Var}\,\hat{\varphi}}$	ψ	$E\hat{\psi}$	$\sqrt{\mathrm{Var}\,\hat{\psi}}$
57,8	58,2	17,5	2,25	2,47	0,425

Źródło: opracowanie własne.

Wniosek 1. Przykład ten wskazuje, że model Bühlmanna-Strauba jest dobrym modelem do prognozowania składki netto w niejednorodnych portfelach, niezbyt małych i obserwowanych przez pewien okres.

Wniosek 2. Jeżeli we wzorze na składkę zaufania wstawi się estymator $\overline{X}_{ww}$ zamiast estymatora $\overline{X}_{zw}$ dla parametru μ, to jakość prognozy składki netto nie ulega istotnemu pogorszeniu, natomiast zyskuje się bardzo na prostocie obliczeń, ponieważ wartości estymatora $\overline{X}_{ww}$ oblicza się bardzo łatwo, w przeciwieństwie do obliczeń wartości estymatora $\overline{X}_{zw}$.

Literatura

[1] De Vylder F., *Practical Credibility Theory with Emphasis on Optimal Parameter Estimation*, „ASTIN Bulletin" 1981, nr 12, s. 115-131.
[2] Goovaerts M.J., Hoogstad W.J.. *Credibility Theory*, Nationale-Nederlanden N.V., Rotterdam 1987.
[3] Herzog T.N., *Introduction to Credibility Theory*, ACTEX Publications, Winsted 1996.
[4] Jasiulewicz H., *Teoria zaufania. Modele aktuarialne*, AE, Wrocław 2005.

[5] Kaas R., Goovaerts M.J., Dannenburg D.R., *Practical Actuarial Credibility Models*, Institute of Actuarial Science, Amsterdam 1996.
[6] Kaas R., Goovaerts M.J., Dhane J., Denuit M., *Modern Actuarial Risk Theory*, Kluwer Academic Publishers, Boston 2001.
[7] Mikosch T., *Non-Life Insurance Mathematics*, Springer-Verlag, Berlin 2004.
[8] Sundt B., *Linearly Sufficient fun with Credibility*, „Insurance Math. Econom." 1991, nr 10, s. 69-74.

CREDIBILITY PREMIUMS RESPECTING THE WEIGHTS OF CONTRACTS

Summary

The paper presents a conception of credibility premiums, which are not well known in the Polish literature on insurance. The Bühlmann-Straub model for setting net premiums for heterogeneous insurance policies is analyzed. Natural and unnatural weights of insurance contracts will be discussed. Next, the influence of the weights on credibility premiums and properties of such predictors will be presented. The theoretical considerations will be ilustrated by a numerical example concerning motor vehicle insurance. This example indicates the potential applications of these concepts in insurance practice.