Karina Ostoj[1]

# Analiza predyktorów *credibility* rezerw IBNR

## 1. Wstęp

Celem opracowania jest zbadanie własności bayesowskich predyktorów współczynników łańcuchowych oraz analiza wrażliwości jakości prognozy tzw. rezerw na szkody zaistniałe i niezgłoszone (ang. *Incurred but not Reported* – IBNR) dokonanej z wykorzystaniem rzeczonych predyktorów w zależności od założeń przyjętych w modelu. W artykule poddano weryfikacji hipotezę, że bayesowskie predyktory liniowe współczynników rozwoju szkód, które ze względu na ich postać potraktować można jako uogólnienie klasycznych współczynników *chain ladder*, pozwalają na uzyskanie wyższej precyzji oszacowania wartości rezerw IBNR. Za kryterium oceny przyjęto względny błąd średniokwadratowy, którego wartości otrzymano w drodze symulacji. Wyniki analizy własności predyktorów uzależniono przede wszystkim od założeń dotyczących postaci rozkładu *a priori* zmiennej nieobserwowalnej.

Hipotezę o wyższej jakości wymienionych predyktorów zweryfikowano na podstawie analizy symulacyjnej zaimplementowanej w programie *R*.

W niniejszym artykule podjęto próbę rozwinięcia problemu przedstawionego przez A. Gislera i M.V. Wütricha[2], których praca posłużyła również jako źródło przykładowych danych do analiz. Wykorzystanie bayesowskich wariantów metody *chain ladder* jest dyskutowane w artykułach naukowych z ostatnich lat[3], co potwierdza istotność zagadnienia.

W artykule przedstawiono podstawowe założenia klasycznej metody *chain ladder* wraz z wykorzystywanymi w niej oznaczeniami, założenia bayesowskiego

---

[1] Deloitte, Dział Usług Aktuarialnych; Szkoła Główna Handlowa w Warszawie, doktorant w Kolegium Analiz Ekonomicznych.

[2] A. Gisler, M.V. Wüthrich, *Credibility for the Chain Ladder Reserving Method*, „ASTIN Bulletin" 2008, vol. 38, iss. 2, s. 565–600.

[3] Podobna analiza przeprowadzona została np. w: http://www.dbc.wroc.pl/Content/32942/Kujawski_Oszacowanie_Rezerwy_IBNR_Bayesowskim_Modelem_2016.pdf (data odczytu: 20.03.2018).

wariantu tej metody oraz konstrukcję zgodnego z nimi modelu rozwoju szkód, krótki opis sposobu przeprowadzania symulacji oraz wyniki analizy przeprowadzonej w trzech wariantach.

## 2. Organizacja danych. Podstawowe założenia metody *chain ladder*

Prognozy rezerw IBNR z wykorzystaniem metody *chain ladder* dokonuje się w oparciu o szeregi czasowe o ustalonym kroku czasowym, za który w niniejszym artykule – bez utraty ogólności – przyjęto jeden rok sprawozdawczy. Ustalanie wartości rezerw sprowadza się do analizy czasu, jaki upłynął pomiędzy zaistnieniem szkody a momentem poinformowania zakładu ubezpieczeń o jej wystąpieniu. Indeksem $i \in \{0,\ldots, I\}$ oznaczono numer porządkowy roku zaistnienia szkody (rok bazowy), gdzie $I$ – numer ostatniego pełnego okresu, o którym zakład posiada pełną informację w zakresie szkód zgłoszonych. Indeksem $j \in \{0,\ldots, J\}$ oznaczono tzw. rok rozwoju szkód, czyli liczbę pełnych lat, jakie upłynęły pomiędzy rokiem bazowym oraz rokiem zgłoszenia szkody, gdzie $J$ to maksymalna przewidziana w modelu wartość indeksu $j$. Indeksem $k \in \{1,\ldots, K\}$ oznaczono $k$-ty wyróżniony segment biznesu, który można logicznie wyodrębnić względem pozostałych segmentów, należy jednak do portfela o zbliżonym procesie rozwoju szkód. Przez segmenty biznesu rozumieć można np. kanały dystrybucji, oddziały, agencje albo jednostki geograficzne dla tej samej grupy ubezpieczeń[4].

Niech $C_{i,j}^{k}$ $i \in \{0,\ldots, I\}$, $j \in \{0,\ldots, J\}$, $k \in \{0,\ldots, K\}$, oznacza skumulowaną wartość szkód zaistniałych w roku $i$ w segmencie $k$, zgłoszonych z co najwyżej $j$-letnim opóźnieniem. W zależności od przyjętego modelu elementy $C_{i,j}^{k}$ mogą uwzględniać również regresy i odzyski, ponadto mogą reprezentować również inne kategorie wielkości, jak np. liczbę szkód. Przez zgłoszenie szkody należy w szczególności rozumieć zwiększenia lub zmniejszenia roszczeń (w tym odzyski i regresy) zaistniałych z opóźnieniem $j$ względem daty wystąpienia szkody. W praktyce ubezpieczeniowej dane szkodowe z reguły zorganizowane są w postaci tzw. trójkąta szkód. Zbiór danych zawartych w trójkącie szkód po zakończeniu $I$-tego okresu sprawozdawczego dla $k$-tego segmentu biznesu zdefiniowano jako:

[4] http://prawo.sejm.gov.pl/isap.nsf/download.xsp/WDU20031241151/U/D20031151Lj.pdf (data odczytu: 25.07.2017).

$$C_I^k = \left\{ C_{i,j}^k : i + j \le I \right\},\ k \in \left\{1,\ldots,K\right\}. \tag{1}$$

Dla $I > J$ przyjmuje kształt trapezu, natomiast dla $I = J$ – trójkąta.

Wartość przyszłych zobowiązań z tytułu odszkodowań i świadczeń $R_i^k$, czyli wartości szkód, które zaistniały, lecz nie zostały jeszcze zgłoszone, z dokładnością do roku bazowego obliczana jest zgodnie ze wzorem:

$$R_i^k = C_{i,J}^k - C_{i,I-i}^k,\ i \in \left\{I - J + 1,\ldots,I\right\},\ k \in \left\{1,\ldots,K\right\}, \tag{2}$$

przy czym wartość $C_{i,J}^k$ jest wartością zmiennej losowej, której realizacja pozostaje nieznana w okresie $\left[I\ ,I+1\right)$. Celem analizy jest wyznaczenie predyktora $\hat{R}_i^k$ (wartości rezerwy szkodowej), opisanego wzorem:

$$\hat{R}_i^k = \hat{C}_{i,J}^k - C_{i,I-i}^k,\ i \in \left\{I - J + 1,\ldots,I\right\},\ k \in \left\{1,\ldots,K\right\}, \tag{3}$$

gdzie $\hat{C}_{i,J}^k$ jest predyktorem zmiennej $C_{i,J}^k$. Problem predykcji $R_i^k$ sprowadza się zatem do problemu wyznaczenia wartości predyktora $\hat{C}_{i,J}^k$ dla każdego roku bazowego $i \in \left\{0,\ldots,\ I\right\}$. Istnieje wiele różnych metod wyznaczania wartości $\hat{C}_{i,J}^k$; przegląd tych metod zawarty został np. w opracowaniu P.D. Englanda i R.J. Verralla[5] lub M.V. Wütricha i M. Merza[6]. Wyniki badania z 2004 r., przeprowadzonego na grupie dziesięciu zakładów ubezpieczeń operujących w Polsce wskazały, że prawie wszystkie stosowały metodę Bornhuettera-Fergusona bądź metodę *chain ladder* (w wariancie klasycznym lub jej zmodyfikowaną wersję)[7].

Założenia klasycznej metody *chain ladder* na gruncie statystycznym po raz pierwszy[8] sformułowane zostały przez T. Macka. Przedstawiają się one następująco[9]:

1) Dla każdego $k \in \left\{1,\ldots,\ K\right\}$ zmienne losowe $C_{i,j}^k$ opisujące wartość szkód w różnych latach bazowych $i \in \left\{0,\ldots,\ I\right\}$ są niezależne.

---

[5] P.D. England, R.J. Verrall, *Stochastic claims reserving in general insurance*, „British Actuarial Journal", August 2002, vol. 8, iss. 3, s. 443–518.

[6] M.V. Wüthrich, M. Merz, *Stochastic claims reserving methods in insurance*, Wiley, Chichester 2008.

[7] http://www.kalkulator-ubezpieczeniowy.pl/analiza_rezerw_2005.pdf (data odczytu: 10.03.2018).

[8] A. Gisler, M.V. Wüthrich, *Credibility for the Chain Ladder Reserving Method*, „ASTIN Bulletin" 2008, vol. 38, iss. 2, s. 567–568.

[9] T. Mack, *Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates*, „ASTIN Bulletin" 1993, vol. 23, iss. 2, s. 213–225.

2) Istnieją takie stałe $f_j^k$, $\tilde{\sigma}_j^k \in R_+$, że dla każdego $i \in \{0,\ldots, I\}$, $j \in \{0,\ldots,J-1\}$, $k \in \{1,\ldots, K\}$ zachodzi:

$$E\left[C_{i,j+1}^k | C_{i,j}^k\right] = f_j^k C_{i,j}^k, \tag{4}$$

$$Var\left[C_{i,j+1}^k | C_{i,j}^k\right] = (\tilde{\sigma}_j^k)^2 \quad C_{i,j}^k. \tag{5}$$

Na potrzeby artykułu do założeń klasycznej metody *chain ladder*, sformułowanych przez Macka, włącza się indeks segmentu biznesu *k*. W efekcie te same założenia dotyczą każdego z *k* segmentów i nie ustala się żadnej zależności pomiędzy tymi segmentami. Założenia modelu odnoszą się wyłącznie do pierwszych dwóch momentów zmiennej losowej $C_{i,j}^k$, dlatego model określa się mianem *distribution-free* (ang. „pozbawiony założeń odnośnie do postaci rozkładu"). Zaletą modelu jest stworzenie podstaw do wyznaczenia błędu standardowego predykcji. Predykcji wartości $C_{i,J}^k$ dokonuje się z wykorzystaniem klasycznych współczynników rozwoju szkód (inaczej: klasycznych współczynników łańcuchowych), zdefiniowanych zgodnie ze wzorem:

$$\hat{f}_j^k = \frac{{}^kS_{j+1}^{I-j-1}}{{}^kS_j^{I-j-1}}, \quad j \in \{0,\ldots,J-1\}, \quad k \in \{1,\ldots,K\}, \tag{6}$$

gdzie ${}^kS_j^p = \sum_{i=0}^{p} C_{i,j}^k$, $p \le I-j$. Predyktorem zmiennej $C_{i,j}^k$ ze zbioru $\{C_{i,j}^k : i+j > I \wedge i \le I\}$, $k \in \{1,\ldots,K\}$ (nazywanego dolnym trójkątem szkód) jest funkcja:

$${}^{CL}\hat{C}_{i,q}^k = C_{i,I-i}^k \prod_{j=I-i}^{q-1} \hat{f}_j^k, \tag{7}$$

gdzie $i \in \{I-J+1,\ldots,I\}$, $k \in \{1,\ldots,K\}$, a predyktorem wartości rezerw $R_i^k$ funkcja:

$${}^{CL}\hat{R}_i^k = {}^{CL}\hat{C}_{i,J}^k - C_{i,I-i}^k, \tag{8}$$

gdzie $i \in \{I-J+1,\ldots,I\}$, $k \in \{1,\ldots,K\}$.

Klasyczna metoda *chain ladder* jest metodą powszechnie stosowaną, jednak jej predyktory mogą cechować się znacznym błędem prognozy *ex post* w pewnych określonych warunkach, w szczególności dla portfeli małych lub portfeli z krótką historią szkodową. Predykcji rezerwy IBNR dla *k*-tego segmentu biznesu (wobec ograniczonej bazy danych szkodowych) można dokonać w oparciu

o informacje pochodzące z innych $k' \in \{1,\ldots,k-1,\ k+1,\ \ldots,K\}$ segmentów biznesu[10]. Jednym ze sposobów wykorzystania tego typu informacji jest zastosowanie tzw. estymatorów *credibility*, należących do klasy liniowych estymatorów bayesowskich. Teoria *credibility* („teoria wiarygodności", „teoria zaufania") pozwala na włączenie w proces estymacji parametrów rozkładu zmiennej losowej tzw. informacji zewnętrznej (inaczej informacji kolektywnej). Estymatory *credibility* klasycznie stosuje się do wyceny kontraktów w niejednorodnych portfelach ubezpieczeniowych. Zaletą tych estymatorów jest brak konieczności przyjęcia jednoznacznej tezy o homogeniczności bądź heterogeniczności portfela polis. Estymacja polega na uwzględnieniu zarówno informacji o samym ubezpieczonym (informacja indywidualna), jak i o całym portfelu (informacja kolektywna). W artykule podobne podejście zastosowane będzie przy szacowaniu wartości współczynników łańcuchowych, zakładając, że faktyczna wartość współczynników łańcuchowych może stanowić wypadkową informacji o pojedynczym segmencie biznesu i o całym portfelu.

Alois Gisler i Mario V. Wütrich[11] wskazują dwa podstawowe źródła informacji kolektywnej – dane pochodzące z innych segmentów biznesu oraz dane rynkowe, przy czym w swojej publikacji za źródło informacji kolektywnej przyjęli inne segmenty biznesu[12]. Zastosowanie danych rynkowych sprowadzałoby się do stworzenia trójkąta szkód dla całego rynku ubezpieczeniowego i potraktowania go jako jedynego źródła informacji zewnętrznej.

## 3. Bayesowski wariant metody *chain ladder*. Estymatory *credibility*

W porównaniu do modelu Macka w bayesowskim wariancie metody *chain ladder* przyjmujemy bardziej restrykcyjne założenia o rozkładzie współczynnika rozwoju szkód. Wiedzę *a priori* o rozwoju szkód reprezentuje $KJ$-wymiarowych zmiennych losowych:

$$\boldsymbol{F}^k = \left(F_0^k,\ldots,F_{J-1}^k\right)',\ k \in \{1,\ldots,K\}, \tag{9}$$

[10] Ibidem, s. 565.
[11] Ibidem.
[12] Ibidem, s. 565–556.

gdzie zmienne $F_0^k,\ldots,F_{J-1}^k$ są niezależne, a zmienne $\boldsymbol{F}^k,\ k\in\{1,\ldots,K\}$ – niezależne, o tym samym rozkładzie.

Niech $\boldsymbol{F}$ oznacza $J\times K$-wymiarową zmienną losową $(\boldsymbol{F}^1,\ldots,\boldsymbol{F}^K)'$. Pod pojęciem bayesowskich współczynników łańcuchowych (inaczej: bayesowskich współczynników rozwoju szkód) rozumie się wektor realizacji zmiennej losowej $\boldsymbol{F}^k$ dla $k$-tej linii biznesowej:

$$\boldsymbol{f}^k=\left(f_0^k,\ldots,\ f_{J-1}^k\right)',\ k\in\{1,\ldots,K\}. \tag{10}$$

Po wprowadzeniu pomocniczej zmiennej (znormalizowaniu wartości trójkąta szkodowego):

$$Y_{i,j}^k=\frac{C_{i,j+1}^k}{C_{i,j}^k},\ i\in\{0,\ldots,I\},\ j\in\{0,\ldots,J-1\},\ k\in\{1,\ldots,K\}, \tag{11}$$

założenia modelu można sformułować w sposób następujący[13]:

1) Warunkowo, przy danym $\boldsymbol{F}$, zmienne losowe $C_{i,j}^k$ dla różnych lat bazowych $i\in\{0,\ldots,I\}$ są niezależne.
2) Niech $\mathcal{B}_j=\left\{C_{i,l}^k:\ i+l\le I\wedge l\le j\wedge\ k\in\{1,\ldots,K\}\right\}$, gdzie $j\in\{0,\ldots,J\}$, będzie odpowiednim podzbiorem zbioru danych pozwalających na stworzenie trójkąta szkodowego $\left\{C_I^k:k\in\{1,\ldots,K\}\right\}$, a $\sigma_j^2\left(F_j^k\right),\ j\in\{0,\ldots,J\}$ – pewną funkcją, zwaną funkcją wariancji. Warunkowo przy danym $\boldsymbol{F}$ oraz zbiorze $\mathcal{B}_j$ rozkład zmiennej $Y_{i,j}^k$ ($i\in\{0,\ldots,I\},\ j\in\{0,\ldots,J-1\},\ k\in\{1,\ldots,K\}$) zależy wyłącznie od $C_{i,j}^k$ i

$$E\left[Y_{i,j}^k|\boldsymbol{F},\mathcal{B}_j\right]=F_j^k, \tag{12}$$

$$Var\left[Y_{i,j}^k|\boldsymbol{F},\mathcal{B}_j\right]=\frac{\sigma_j^2\left(F_j^k\right)}{C_{i,j}^k} \tag{13}$$

oraz skończone są momenty $Var(F_j^k)$ i $E\left[\sigma_j^2\left(F_j^k\right)\right]$.
3) Zmienne $F_j^k,\ j\in\{0,\ldots,J-1\},\ k\in\{1,\ldots,K\}$ są niezależne.

Konsekwencją przyjęcia powyższych założeń jest warunkowa niezależność zmiennych $Y_{i,j}^k$ dla różnych lat wystąpienia szkody $i\in\{0,\ldots,I\}$ oraz ich warunkowe nieskorelowanie dla różnych lat rozwoju szkód $j\in\{0,\ldots,J-1\}$ względem $\boldsymbol{F}$. Warunkowa wartość oczekiwana i warunkowa wariancja zmiennej $Y_{i,j}^k$ zależą

[13] Ibidem, s. 570.

wyłącznie od zmiennej losowej $F_j^k$ i nie zależą od zmiennych $F_m^k$, powiązanych z innymi latami rozwoju szkód $m \neq j,\ m,j \in \{0,\ldots,J-1\}$ (por. (14) i (15)).

Najlepszym estymatorem (ang. *best predictor* – BP) współczynnika $F_j^k$ przy kwadratowej funkcji straty jest estymator:

$$F_{j,k}^{BP} = \arg\min_{h(C_I^k)} E\left[\left(h\left(C_I^k\right) - F_j^k\right)^2 | C_I^k\right] = E\left[F_j^k | C_I^k\right], \tag{14}$$

gdzie $j \in \{0,\ldots,J-1\},\ k \in \{1,\ldots,K\}$, a $h$ jest dowolną funkcją mierzalną zdefiniowaną na obserwacjach z górnego trójkąta szkodowego. Na podstawie wniosku o warunkowym nieskorelowaniu zmiennych $Y_{i,j}^k$ dla różnych lat rozwoju szkód dowodzi się, że predyktor skumulowanej łącznej wartości odszkodowań i świadczeń zaistniałych w roku bazowym $i \in \{0,\ldots,I\}$ również ma własności estymatora bayesowskiego (minimalizuje zadaną funkcję straty opisanej wzorem (14)) i przyjmuje postać[14]:

$${}^{BP}\hat{C}_{i,J}^k = C_{i,I-i}^k \prod_{j=I-i}^{J-1} F_{j,k}^{BP}, \tag{15}$$

gdzie $k \in \{1,\ldots,K\}$, co pozwala na wyznaczenie bayesowskiego predyktora rezerw. Wyznaczenie jego wartości wymaga znajomości łącznego teoretycznego rozkładu zmiennych losowych $\boldsymbol{F}$ i $Y_{i,j}^k$, co stanowi istotne ograniczenie w praktyce ubezpieczeniowej. Rozwiązaniem tej kwestii jest ograniczenie się do klasy predyktorów liniowych. Funkcja $F_{j,k}^{BLP}$ jest najlepszym predyktorem liniowym (ang. *best linear predictor* – BLP) zmiennej $F_j^k$ przy kwadratowej funkcji straty, jeżeli jest funkcją liniową obserwacji $Y_{i,j}^k,\ i \in \{0,\ldots,I\},\ j \in \{0,\ldots,I-i\},\ k \in \{1,\ldots,K\}$ oraz spełnia:

$$F_{j,k}^{BLP} = arg\ \min_{F_j^L} E[\left(F_j^k - F_j^L\right)^2 | C_I^k], \tag{16}$$

gdzie

$$F_j^L = \left\{\hat{F}_{j,k} : \hat{F}_{j,k} = a_{0,j}^k + \sum_{i=0}^{I-j} a_{i,j}^k Y_{i,j}^k \ \wedge\ \forall_{i \in \{0,\ldots,I-j\}}\ a_{i,j}^k \in R\right\},\ j \in \{0,\ldots,J-1\},\ k \in \{1,\ldots,K\}.$$

Bayesowski wariant modelu rozwoju szkód, spełniający założenia przytoczone na początku rozdziału, spełnia jednocześnie założenia modelu Bühlmanna-Strauba, zgodnie z którym najlepszy liniowy predyktor przyjmuje postać:

[14] Ibidem, s. 573.

$$F_{j,k}^{BLP} = \alpha_{j,k}\hat{f}_j^k + \left(1-\alpha_{j,k}\right)f_j^{coll}, \tag{17}$$

$j \in \{0,\ldots,J-1\}$, $k \in \{1,\ldots,K\}$, gdzie:

$$\hat{f}_j^k = \frac{{}^kS_{j+1}^{I-j-1}}{{}^kS_j^{I-j-1}}, \tag{18}$$

$$f_j^{coll} = E[F_j^k], \tag{19}$$

$\alpha_{j,k}$ jest tzw. współczynnikiem zaufania (inaczej współczynnikiem wiarygodności) równym:

$$\alpha_{j,k} = \frac{{}^kS_j^{I-j-1}}{{}^kS_j^{I-j-1} + \frac{s_j^2}{\tau_j^2}}, \tag{20}$$

gdzie $s_j^2 = E\left[\sigma_j^2\left(F_j^k\right)\right]$ oraz $\tau_j^2 = Var\left[F_j^k\right]$[15].

Otrzymany estymator przyjmuje postać średniej ważonej klasycznego współczynnika *chain ladder* $\hat{f}_j^k$, reprezentującego informację indywidualną, oraz wartości oczekiwanej $f_j^{coll}$ rozkładu *a priori*. Predyktory $C_{i,J}^k$ wyznaczane są zgodnie ze wzorem:

$${}^{BLP}\hat{C}_{i,J}^k = C_{i,I-i}^k \prod_{j=I-i}^{J-1} F_{j,k}^{BLP}, \; i \in \{0,\ldots,I\}, \; k \in \{1,\ldots,K\}. \tag{21}$$

Należy podkreślić, że ograniczenie do klasy estymatorów liniowych zubaża własności predyktora zmiennej $C_{i,J}^k$. Predyktory *BP* zarówno zmiennych $F_j^k$, jak i $C_{i,J}^k$ spełniają własności predyktora bayesowskiego. Współczynniki $F_{j,k}^{BLP}$ spełniają wprawdzie z definicji założenia estymatora BLP – są najlepszymi liniowymi predyktorami współczynników łańcuchowych, jednak predyktor ${}^{BLP}\hat{C}_{i,J}^k$ nie jest już najlepszym liniowym predyktorem zmiennej $C_{i,J}^k$, a jedynie predyktorem opartym na najlepszych liniowych predyktorach współczynników łańcuchowych.

Oszacowanie $F_{j,k}^{BLP}$ wymaga znajomości teoretycznych parametrów rozkładu *a priori*: $s_j^2$, $\tau_j^2$ oraz $f_j^{coll}$. Nieobciążonym estymatorem parametru $f_j^{coll}$ o najmniejszej wariancji w klasie estymatorów liniowych jest:

[15] Ibidem, s. 575.

$$\hat{f}_j^{coll} = \sum_{k=1}^{K} \frac{\alpha_{j,k}}{\alpha_j} f_j^k, \ j \in \{0,\ldots,J-1\}, \tag{22}$$

gdzie $\alpha_j = \sum_{k=1}^{K} \alpha_{j,k}, \ j \in \{0,\ldots,J-1\}$[16]. Przy założeniu, że znane są wartości parametrów strukturalnych $s_j^2$, $\tau_j^2$ otrzymujemy tzw. najlepszy liniowy nieobciążony predyktor BLUP (ang. *best linear unbiased predictor*) $F_j^k$:

$$F_{j,k}^{BLUP} = \alpha_{j,k} \hat{f}_j^k + \left(1-\alpha_{j,k}\right) \hat{f}_j^{coll}, \tag{23}$$

gdzie $j \in \{0,\ldots,J-1\}$, $k \in \{1,\ldots,K\}$. Przy braku pełnej informacji o postaci funkcyjnej rozkładu *a priori* użyteczność estymatorów $F_{j,k}^{BLUP}$ i $F_{j,k}^{BLP}$ jest mocno ograniczona. Nieznane parametry $s_j^2$, $\tau_j^2$, podobnie jak $f_j^{coll}$, również podlegają estymacji. Wykorzystując estymatory $\hat{s}_j^2$, $\hat{\tau}_j^2$ zaproponowane w pracy H. Bühlmanna i A. Gislera[17] oraz zastępując nimi wielkości $s_j^2$ i $\tau_j^2$ we wzorze na $F_{j,k}^{BLUP}$, otrzymuje się empiryczny estymator $F_{j,k}^{EBLUP}$ oraz empiryczny predyktor ${}^{EBLUP}\hat{R}_i^k$ rezerw:

$$F_{j,k}^{EBLUP} = \hat{\alpha}_{j,k} \hat{f}_j^k + \left(1-\hat{\alpha}_{j,k}\right) \hat{\hat{f}}_j^{coll}, \ j \in \{0,\ldots,J-1\}, \ k \in \{1,\ldots,K\}, \tag{24}$$

$${}^{EBLUP}\hat{C}_{i,J}^k = C_{i,I-i}^k \prod_{j=I-i}^{J-1} F_{j,k}^{EBLUP}, \ i \in \{I-J+1,\ldots,I\}, \ k \in \{1,\ldots,K\}, \tag{25}$$

$${}^{EBLUP}\hat{R}_i^k = {}^{EBLUP}\hat{C}_{i,J}^k - C_{i,I-i}^k, \ i \in \{I-J+1,\ldots,I\}, \ k \in \{1,\ldots,K\}, \tag{26}$$

gdzie $\hat{\alpha}_{j,k} = \dfrac{{}^{k}S_j^{I-j-1}}{{}^{k}S_j^{I-j-1} + \dfrac{\hat{s}_j^2}{\hat{\tau}_j^2}}$.

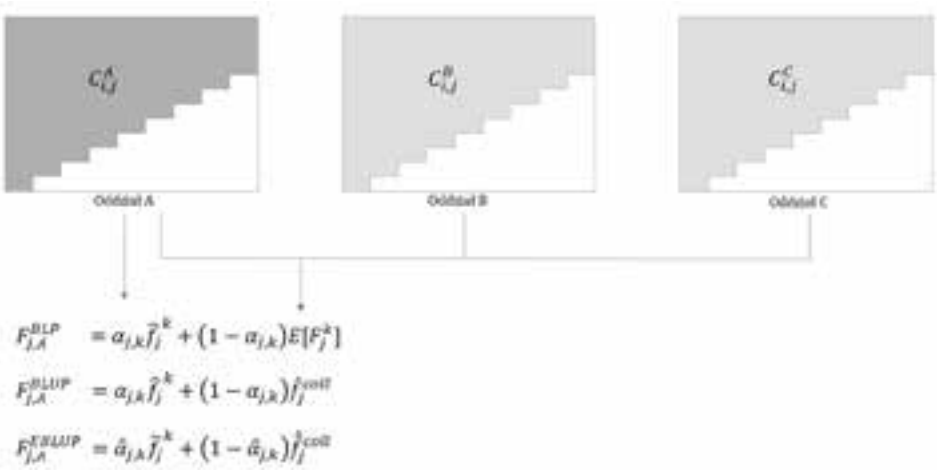
Porównanie wymienionych estymatorów przedstawiono na rysunku 1.

Przed przystąpieniem do analizy symulacyjnej konieczne jest rozwiązanie dwóch problemów natury technicznej. Ponieważ estymator parametru strukturalnego $\hat{\tau}_j^2$ może przyjmować wartości ujemne, zastępujemy go estymatorem

[16] http://www-users.mat.umk.pl/~wniem/Ryzyko/RyzykoUB.pdf (data odczytu: 5.05.2015).

[17] H. Bühlmann, A. Gisler, *A course in credibility theory and its applications*, Springer, Berlin 2005, s. 93–95.

$\hat{\hat{\tau}}_j^{\,2} = max\left(0,\ \hat{\tau}_j^2\right),\quad j \in \{0,\ldots,J-1\}$ [18]. Drugi problem natury implementacyjnej powstaje w momencie, gdy dla każdego $k \in \{1,\ldots,K\}$ zachodzi $\hat{\alpha}_{j,k} = 0$. W efekcie $\hat{\alpha}_j = 0$, przez co niemożliwe jest wyznaczenie $\hat{\hat{f}}_j^{coll}$ zgodnie z empirycznym odpowiednikiem wzoru (22) wykorzystującym predyktory $\hat{\alpha}_{j,k}$. W takich przypadkach przyjęto, że $\hat{\hat{f}}_j^{coll} = \frac{1}{K}\sum_{k=1}^{K}\hat{f}_j^k$. Wymienione ograniczenia i wprowadzone modyfikacje estymatorów BLUP prowadzą do rezygnacji z ważnej cechy estymatorów współczynników łańcuchowych, jaką jest nieobciążoność.




**Rysunek 1. Postać wybranych bayesowskich estymatorów liniowych**
Źródło: opracowanie własne

Postać liniowych predyktorów wskazuje na to, że stanowią one pewnego rodzaju uogólnienie klasycznych współczynników *chain ladder*, dlatego oczekiwalibyśmy, że liniowe predyktory dadzą lepsze rezultaty niż uzyskane metodą klasyczną. Z drugiej strony ograniczenie do klasy predyktorów liniowych powoduje ograniczenie własności samych predyktorów rezerw, a ponadto w przypadku predyktorów EBLUP z powodów technicznych stosujemy predyktory obciążone. Celem symulacji będzie sprawdzenie, czy włączenie w proces estymacji informacji kolektywnej spowoduje polepszenie jakości predykcji pomimo zastosowania predyktorów obciążonych.

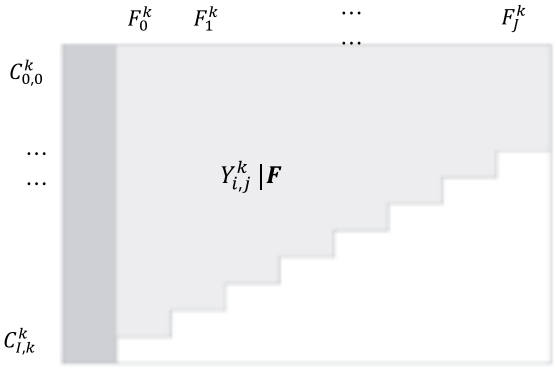
[18] Ibidem, s. 95.

## 4. Konstrukcja modelu rozwoju szkód. Architektura symulacji

Celem opracowania jest porównanie jakości predykcji dokonanej z wykorzystaniem liniowych predyktorów współczynników łańcuchowych oraz zbadanie ich wrażliwości na zmianę wybranych parametrów rozkładu *a priori*, przyjmując założenia o modelu rozwoju szkód przytoczone w punkcie 3. Symulacja wymaga w szczególności przyjęcia założeń o:

- rozkładzie wartości roszczeń zgłoszonych w tym samym okresie, w którym wystąpiła szkoda (pierwsza kolumna górnego trójkąta szkodowego),
- postaci funkcyjnej i parametrach rozkładu *a priori* zmiennej losowej $\boldsymbol{F}$,
- postaci funkcyjnej i parametrach rozkładu warunkowego zmiennej losowej $Y_{i,j}^k$ przy danym $\boldsymbol{F}$ dla $(i,j,k) \in \{0, \ldots, I-1\} \times \{0, \ldots, I-i-1\} \times \{1, \ldots, K\}$.

Oceny jakości predyktorów rezerw dokonano na podstawie trzech rodzajów symulacji, z których dwa polegały na analizie błędu prognozy *ex post*, przyjmując za rozkład *a priori* rozkłady o tych samych macierzach pierwszych i drugich momentach oraz różnej postaci funkcyjnej, natomiast trzecia – na testowaniu własności predyktorów przy rozkładzie *a priori* będącym dyskretną mieszanką dwóch rozkładów.




**Rysunek 2. Opis zmiennych generowanych w toku symulacji**

Źródło: opracowanie własne

Symulacja polega na *N*-krotnym powtórzeniu schematu składającego się z następujących etapów (por. rysunek 2):

1. Generowanie wartości z pierwszej kolumny trójkąta szkodowego $C_{i,0}^k$ dla każdego segmentu biznesu $k \in \{1, \ldots, K\}$.
2. Generowanie realizacji zmiennej losowej $\boldsymbol{F}$ dla $(j,k) \in \{0, \ldots, I-i-1\} \times \{1, \ldots, K\}$.

3. Generowanie górnego trójkąta szkód $C_I^k$ dla $k \in \{1,\ldots,K\}$.
4. Generowanie faktycznej realizacji trójkąta dolnego zgodnie z założeniami wymienionymi w poprzednich punktach.
5. Generowanie predyktorów wartości trójkąta dolnego otrzymanego z wykorzystaniem:
   - klasycznych współczynników *chain ladder*,
   - predyktorów $F_{j,k}^{BLP}$,
   - predyktorów $F_{j,k}^{BLUP}$,
   - predyktorów $F_{j,k}^{EBLUP}$.

Oceny jakości prognozy predyktora $\hat{Q}$ dokonuje się w badaniu na podstawie względnego błędu średniokwadratowego:

$$rRMSE\left(\hat{Q}\right) = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left({}^{n}\hat{Q} - {}^{n}Q\right)^2}}{\left|\frac{1}{N}\sum_{n=1}^{N}{}^{n}Q\right|}, \tag{27}$$

gdzie $n \in \{1, \ldots, N\}$ oznacza $n$-tą iterację, przy czym błąd $rRMSE\left(\hat{Q}\right)$ podlega analizie zarówno dla całego portfela ($Q = \sum_{k=1}^{K}\sum_{i=1}^{I} R_i^k$), jak i w podziale na segmenty biznesu $Q = \sum_{i=0}^{I} R_i^k$ oraz na lata bazowe $Q = \sum_{k=1}^{K} R_i^k$.

Na potrzeby symulacji za wartości parametrów odpowiednich rozkładów przyjęto parametry wyznaczone na podstawie danych zawartych w artykule A. Gislera i M.V. Wütricha[19]. Autorzy zamieścili trójkąty szkód dla towarzystwa Winterthur Insurance Company, zawierające dane z lat bazowych 1986–2006 dla sześciu segmentów biznesu. Model rozwoju szkód przewiduje możliwość zgłoszenia prawa do roszczenia do 10 lat od wystąpienia zdarzenia ubezpieczeniowego oraz uwzględnia wartość regresów i odzysków. Ze względu na poufny charakter danych wartości w trójkątach szkód pomnożone zostały przez nieznaną stałą.

Estymacji podlegały w szczególności parametry rozkładu wartości z pierwszej kolumny górnego trójkąta szkodowego, wektor wartości oczekiwanych i wektor wartości wariancji rozkładu $\boldsymbol{F}$.

Przy założeniach modelu zmienne losowe $C_{i,0}^k$, zawarte w pierwszej kolumnie trójkąta szkodowego, są w zasadzie jedynymi zmiennymi, które mogą zawierać

[19] A. Gisler, M.V. Wüthrich, op.cit., s. 591–596.

informację[20] charakterystyczną o zadanej linii biznesowej oraz o zadanym roku bazowym. Do modelowania łącznej wartości szkód rozważono rozkłady Pareta, Weibulla oraz rozkład log-normalny. Ze względu na małą liczebność próby decyzję o postaci funkcyjnej podjęto na podstawie empirycznych wykresów kwantylowych. Do modelowania wartości z pierwszej kolumny trójkąta górnego przyjęto rozkład log-normalny o parametrach wyznaczonych osobno dla każdego segmentu biznesu.

Zgodnie z założeniami bayesowskiego wariantu metody *chain ladder* rozkład *a priori* opisuje wartość oczekiwaną ilorazów $\frac{C_{i,j+1}^{k}}{C_{i,j}^{k}}$ dla każdego roku rozwoju szkód $j \in \{0,\ldots, J-1\}$. Rozsądne wydaje się zatem zidentyfikowanie parametrów tego rozkładu w oparciu o klasyczne współczynniki rozwoju szkód $\hat{f}_j^k$, których wartości przedstawiono w tabeli 1. W pierwszym wariancie symulacji przyjęto, że rozkład ten opisany jest przesuniętym (trójparametrycznym) rozkładem gamma:

$$F_j^k \sim Gamma(\alpha_j,\ \beta_j,\ u_j), \quad j \in \{0,\ldots,J-1\},\ k \in \{1,\ldots,K\}, \tag{28}$$

gdzie trójparametryczny rozkład gamma opisany jest funkcją gęstości:

$$f\left(F_j^k\right) = \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}\left(F_j^k - u_j\right)^{\alpha_j - 1} e^{-\beta\left(F_j^k - u_j\right)}, \tag{29}$$

natomiast w drugim wariancie za rozkład *a priori* przyjęto uogólniony rozkład beta:

$$F_j^k \sim Beta(\gamma_j,\ \delta_j,\ p_j,\ q_j), \quad j \in \{0,\ldots,J-1\},\ k \in \{1,\ldots,K\}, \tag{30}$$

gdzie nośnikiem zmiennej jest przedział $[p_j,\ q_j]$, a funkcja gęstości wspomnianego rozkładu beta przedstawia się następująco:

$$f\left(F_j^k\right) = \frac{\Gamma(\gamma_j + \delta_j)}{\Gamma(\gamma_j)\Gamma(\gamma\delta_j)} \frac{1}{\left(q_j - p_j\right)^{\gamma_j + \delta_j - 1}} \left(F_j^k - p_j\right)^{\gamma_j - 1} \left(q_j - F_j^k\right)^{\delta_j - 1}. \tag{31}$$

[20] Wspomnianej tu informacji nie należy łączyć z informacją indywidualną bądź kolektywną, która dotyczy wyłącznie rozkładu *a priori*.

Parametry obu rozkładów – również ze względu na małą liczebność próby, oszacowano metodą momentów na podstawie obserwacji klasycznych współczynników *chain ladder* zawartych w tabeli 1.

**Tabela nr 1. Wartości współczynników rozwoju wyznaczone klasyczną metodą *chain ladder* dla segmentów 1–6**

| Wartości klasycznych współczynników *chain ladder* | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *k* | Rok rozwoju szkód (*j*) | | | | | | | | | |
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 2,27 | 1,23 | 0,98 | 1,02 | 1,01 | 0,98 | 0,96 | 1,00 | 0,99 | 1,00 |
| 2 | 2,13 | 1,09 | 1,03 | 1,00 | 0,99 | 1,00 | 1,01 | 0,99 | 1,00 | 0,99 |
| 3 | 2,18 | 1,13 | 1,03 | 1,04 | 1,00 | 1,00 | 0,99 | 1,00 | 1,00 | 1,00 |
| 4 | 2,10 | 1,07 | 1,05 | 1,01 | 1,00 | 1,01 | 0,99 | 0,99 | 1,00 | 1,00 |
| 5 | 1,93 | 1,11 | 1,01 | 0,99 | 1,00 | 0,99 | 0,99 | 0,99 | 1,00 | 1,00 |
| 6 | 2,99 | 1,19 | 1,14 | 1,00 | 1,00 | 0,97 | 0,99 | 1,00 | 1,00 | 1,00 |

Źródło: opracowanie własne

W związku z tym oba rozkłady cechowały się tą samą wartością oczekiwaną oraz tą samą wariancją. Istotną różnicą jest natomiast przedział określoności obu rozkładów. W trzecim wariancie za rozkład *a priori* przyjęto dyskretną mieszankę rozkładu beta o parametrach zgodnych z przyjętymi w wariancie drugim oraz rozkładu jednostajnego[21] określonego na tym samym nośniku, przy czym za współczynniki kombinacji liniowej przyjęto wartości $p = 0, \frac{1}{10}, \ldots, \frac{9}{10}, 1$.

Zmienna *Z* ma rozkład opisany dwuelementową mieszanką rozkładów, gdy[22]:

$$f_Z(z) = pf_1(z_1) + (1-p)f_2(z_2), \tag{32}$$

gdzie $p \in [0,1]$, $f_1$, $f_2$ – funkcje gęstości zmiennych losowych określonych na jednakowych nośnikach o dystrybuantach odpowiednio $F_1$, $F_2$. Schemat losowania zmiennej z rozkładu będącego dyskretną mieszanką dwóch rozkładów o znanej postaci oraz zadanym parametrze *p* polega na wylosowaniu zmiennej pomocniczej

[21] Rozkład jednostajny jest szczególnym przypadkiem rozkładu beta, jednak w tekście rozkład beta z parametrami wyznaczonymi na podstawie obserwacji będzie nazywany rozkładem beta, a rozkład jednostajny określony na tym samym nośniku – rozkładem jednostajnym.

[22] S.A. Klugman, H.H. Panjer, E.W. Gordon, *Loss Models. From Data to Decisions*, John Wiley & Sons, Inc., Hoboken 2004, s. 59.

$p'$ z rozkładu jednostajnego na przedziale (0,1), a następnie wylosowaniu zmiennej z rozkładu $F_1$, jeżeli $p' < p$ oraz zmiennej z rozkładu $F_2$ w przeciwnym przypadku. W konsekwencji, teoretyczne wartości wartości oczekiwanej i wariancji zmiennej $f_Z(z)$ opisują wzory:

$$EZ = E\left[E\left(Z|P\right)\right] = pEZ_1 + \left(1-p\right)EZ_2, \tag{33}$$

$$\begin{aligned} VarZ &= E\left[Var\left(Z|P\right)\right] + Var\left[E\left(Z|P\right)\right] \\ &= pVarZ_1 + \left(1-p\right)VarZ_2 + p\left(1-p\right)\left(EZ_1 - EZ_2\right)^2. \end{aligned} \tag{34}$$

Zamierzeniem wykorzystania mieszanki dwóch rozkładów było porównanie wyników dla rozkładu *a priori* z parametrami oszacowanymi na podstawie próby, czyli z założenia – rozkładu wnoszącego możliwe dużo informacji o zmiennej nieobserwowalnej z wynikami dla rozkładu niewnoszącego praktycznie żadnej informacji.

Zgodnie z drugim założeniem bayesowskiego wariantu metody *chain ladder* obserwowalna zmienna $Y_{i,j}^k$ opisana jest rozkładem warunkowym względem realizacji zmiennej $\boldsymbol{F}$. Za postać funkcyjną rozkładu zmiennej obserwowalnej przyjęto trójparametryczny rozkład log-normalny:

$$Y_{i,j}^k \mid \boldsymbol{F} \sim LN\left(F_j^k; \frac{\sigma_j^2\left(F_j^k\right)}{C_{i,j}^k}; x_j\right), \tag{35}$$

gdzie $i \in \{0, \ldots, I-1\}$, $j \in \{0, \ldots, J-1\}$, $k \in \{1, \ldots, K\}$, $x_j$ – punkt początkowy, za który przyjęto minima zmiennej obserwowalnej. Wektor $I \times J \times K$ zmiennych losowych dalej oznaczony zostanie jako $\boldsymbol{Y}$.

Bayesowski wariant metody *chain ladder* zakłada, że wariancja zmiennej obserwowalnej jest funkcją wartości oczekiwanej i opisana jest tzw. funkcją wariancji: $\sigma_j^2\left(F_j^k\right)$. Na podstawie analizy danych towarzystwa Winterthur Insurance Company postanowiono zastosować kwadratową funkcję wartości oczekiwanej o zerowym wyrazie wolnym. Przyjęcie rosnącej funkcji wariancji jako funkcji wartości oczekiwanej wydaje się zasadne, gdyż z reguły najwyższą wariancję mają wczesne lata rozwoju szkód, cechujące się najwyższymi współczynnikami rozwoju. Wartości współczynników łańcuchowych dla pierwszego roku rozwoju kształtują się w granicach 1,5–2,0, co jest zgodne z wynikami badania z 2004 r., przeprowadzonego na rynku polskim dla grup ubezpieczeniowych działu II[23].

[23] http://www.kalkulator-ubezpieczeniowy.pl/analiza_rezerw_2005.pdf (data odczytu: 10.03.2018).

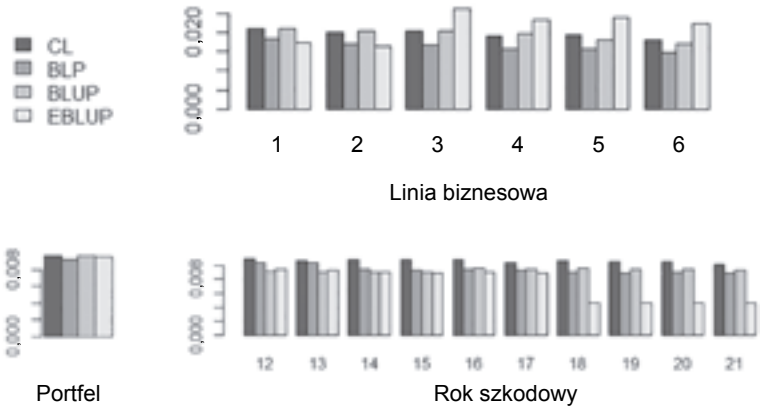
Pierwszy rok rozwoju szkód cechować będą najwyższe wartości współczynników *chain ladder* oraz największa niepewność co do wartości zgłaszanych szkód. W kolejnych latach roszczenia napływają coraz rzadziej, a ich udział w łącznej liczbie szkód z pojedynczego roku bazowego sukcesywnie maleje, co ma wpływ na zmniejszenie wariancji współczynników oraz stabilizację ich wartości wokół jedynki. Fakt uwzględnienia w analizie regresów i odzysków, których natężenie obserwowane jest z pewnym opóźnieniem (w przypadku Winterthur Company natężenie takie zaobserwować można w latach rozwoju szkód: 5, 6 oraz 7, w których wartości współczynników spadły nieco poniżej 1, por. tabela 1), ma nieoczywisty wpływ na wariancję współczynników *chain ladder*. Im bardziej wartość regresów i odzysków będzie zbliżona do wypłacanych przez zakład odszkodowań i świadczeń, tym bardziej stabilizować będą one wartości współczynników rozwoju szkód. Duże udziały regresów i odzysków mogą powodować z jednej strony spadek wartości współczynników poniżej jedynki, a z drugiej wpłynąć na zwiększenie ich wariancji. W przypadku trójkątów szkód Winterthur Insurance Company dla lat charakteryzujących się współczynnikami *chain ladder* zbliżonymi do jedności wariancja wyrażona w jednostkach wypłaconych odszkodowań i świadczeń jest relatywnie niska. Cecha ta charakteryzuje zarówno lata ze znaczącym udziałem regresów, jak i późne lata rozwoju szkód, a dla lat „regresowych" nieznaczny wzrost wariancji odnotowano jedynie w przypadku piątej linii biznesowej. Eskalacja wariancji zauważalna jest dla pierwszego roku rozwoju szkód, dla których oczekiwana wartość współczynnika wynosi około dwóch.

Zaprezentowana kalibracja modelu pozwala na zbadanie własności rozważanych estymatorów *credibility* współczynników łańcuchowych i błędów predykcji rezerw oraz przeprowadzenie analizy wrażliwości. Dla pierwszych dwóch wariantów symulacji przyjęto $N = 50\ 000$ iteracji, natomiast dla analizy wrażliwości przeprowadzono 11 symulacji, każda o liczbie $N = 10\ 000$ iteracji.

## 5. Wyniki symulacji

Zgodnie z założeniami pierwszego i drugiego wariantu symulacji, za rozkład *a priori* przyjmuje się odpowiednio rozkład gamma i rozkład beta o tej samej wartości oczekiwanej oraz wariancji. Ważnym założeniem jest ustalenie przedziału określoności obu rozkładów – w obu przypadkach zastosowano uogólnione postacie funkcyjne rozkładów z parametrami położenia, przy czym

punkt początkowy dla obu wariantów jest taki sam, natomiast nośnik rozkładu beta jest ograniczony, w przeciwieństwie do rozkładu gamma. Rozważenie obu przypadków ma pewne odniesienie do praktyki ubezpieczeniowej, gdyż często z estymacji opartej na wykorzystaniu trójkątów szkód wyklucza się obserwacje odstające od przeciętnie obserwowanych wartości. Rozkład gamma odzwierciedlałby zatem podejście, w którym wartość szkód włączona w proces estymacji współczynników łańcuchowych jest nieograniczona, natomiast rozkład beta uwzględnia wyłącznie szkody o wartościach nieprzekraczających pewnej maksymalnej założonej łącznej wartości szkód.



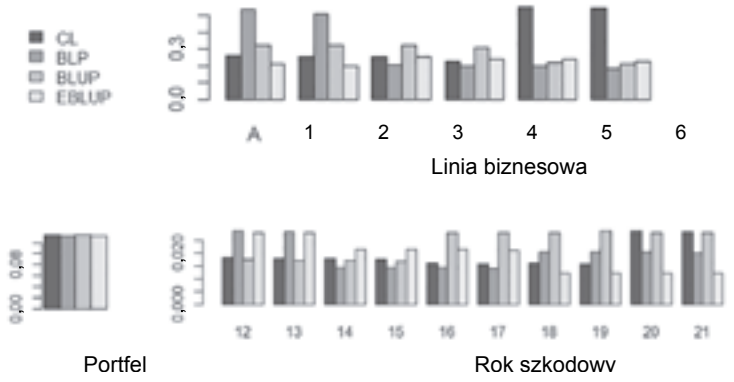

**Rysunek 3. Wartości względnego błędu średniokwadratowego w różnych przekrojach wyznaczone w sposób symulacyjny przy rozkładzie *a priori* gamma**

Źródło: opracowanie własne

Rysunek 3 przedstawia wyniki symulacji dla rozkładu gamma. Na wykresach pokazano przeciętną wartość względnego błędu średniokwadratowego z wykorzystaniem kolejnych wariantów współczynników łańcuchowych uzyskanych za pomocą: metody klasycznej (CL), najlepszych predyktorów liniowych (BLP), najlepszych nieobciążonych predyktorów liniowych (BLUP) oraz najlepszych empirycznych nieobciążonych predyktorów liniowych (EBLUP).

Jakość predykcji na poziomie całego portfela wydaje się jednakowa. Zaledwie dla dwóch spośród czterech segmentów biznesu współczynniki EBLUP cechują się lepszą jakością predykcji niż klasyczne współczynniki łańcuchowe, natomiast w przypadku rozkładu beta względny błąd średniokwadratowy współczynników EBLUP osiąga wartość zbliżoną lub niższą niż klasyczne współczynniki (por. rysunek 3 i 4). W przypadku rozkładu *a priori* gamma wykorzystanie bayesowskich współczynników liniowych skutkuje lepszymi rezultatami w podziale

na lata bazowe, dla rozkładu beta jakość predykcji jest zróżnicowana pomiędzy kolejnymi latami bazowymi. Praktyczne zastosowanie tego wniosku jest jednak ograniczone ze względu na fakt, iż ta sama metoda powinna zostać zastosowana w odniesieniu do całego pojedynczego trójkąta szkodowego, a zatem do wszystkich lat bazowych. W pewnych warunkach zastosowanie bayesowskich współczynników liniowych skutkuje bardziej precyzyjnymi oszacowaniami predyktorów rezerw dla pojedynczych segmentów biznesu, jednak na poziomie całego portfela jakość predykcji wydaje się zbliżona (por. rysunek 4). O ile w kontekście praktyki ubezpieczeniowej zastosowanie tych predyktorów wydaje się ograniczone z punktu widzenia oszacowania dla całego portfela, o tyle bayesowskie współczynniki liniowe mogą dostarczyć informacji o strukturze biznesu na poziomie niższym niż cały portfel (segmenty biznesu).



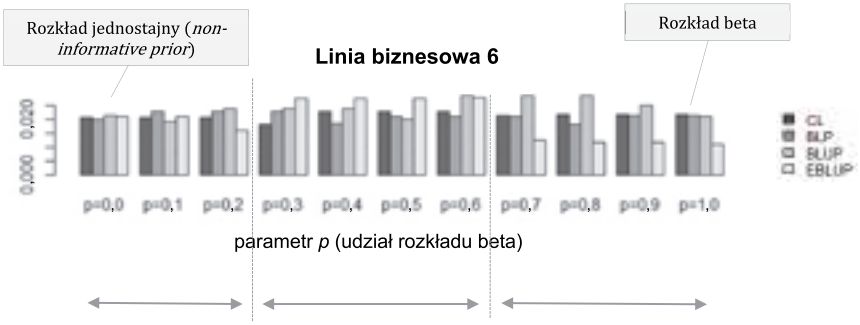

**Rysunek 4. Wartości względnego błędu średniokwadratowego w różnych przekrojach wyznaczone w sposób symulacyjny przy rozkładzie *a priori* beta**

Źródło: opracowanie własne

Analiza wrażliwości polegała na porównaniu własności predyktorów ze względu na rozkład *a priori* stanowiący mieszankę rozkładów beta z różnymi udziałami tych rozkładów. Przebieg zmienności względnego błędu średniokwadratowego był podobny dla wszystkich segmentów biznesu, przykładowy (i zarazem typowy) przebieg dla wybranej linii biznesu przedstawiono na rysunku 5.

Wartość średniego względnego błędu kwadratowego w szczególności dla liniowych współczynników typu EBLUP wykazuje silną, niemonotoniczną zależność od parametru $p$. Najwyższe wartości średniego błędu kwadratowego obserwuje się dla parametrów $p \in \{0,3, 0,4, 0,5, 0,6\}$, czyli takich parametrów, dla których udział rozkładu beta i rozkładu jednostajnego są zbliżone. Dla rozkładu

jednostajnego jakość predykcji jest porównywalna dla wszystkich czterech rozpatrywanych rodzajów predyktorów. W przypadku rozkładów z przeważającym wkładem rozkładu beta ($p \in \{0,7, 0,8, 0,9, 1,0\}$) estymatory EBLUP cechują się najniższym błędem predykcji.




**Rysunek 5. Wartości względnego błędu średniokwadratowego w różnych przekrojach wyznaczone w sposób symulacyjny przy rozkładzie *a priori* będącym dyskretną mieszanką rozkładów beta – na przykładzie segmentu biznesu 6**

Źródło: opracowanie własne

## 6. Podsumowanie i kierunki dalszych badań

Bayesowskie liniowe predyktory stanowią uogólnienie klasycznych współczynników *chain ladder*. Szereg przeprowadzonych symulacji wskazuje, że nie można w sposób jednoznaczny wykazać, który rodzaj predyktorów cechuje się lepszą jakością prognozy, gdyż zależy to silnie od założeń modelu. W pewnych określonych warunkach zastosowanie bayesowskich współczynników liniowych prowadzi do bardziej precyzyjnych predyktorów rezerw dla pojedynczych segmentów biznesu, jednak na poziomie całego portfela jakość predykcji wydaje się zbliżona. W artykule testowano własności predyktorów liniowych względem różnych rozkładów *a priori*. Wyniki szeregu symulacji pomocniczych wskazywały na silne uzależnienie również od innych założeń przyjętych w modelu – jednym z najsilniej wpływających na ostateczne wnioski jest postać funkcji wariancji. Przeprowadzona analiza wrażliwości wskazuje na fakt, że trafność prognozy silnie zależy od kształtu rozkładu *a priori*, gdyż najsłabsze predyktory rezerw otrzymano dla mieszanek rozkładów beta oraz rozkładu jednostajnego

o zbliżonym udziale w ostatecznej postaci rozkładu. Wniosek ten należy jednak rozpatrywać w kategoriach hipotezy, która wymaga dalszej weryfikacji.

## Bibliografia

Bühlmann H., Gisler A., *A course in credibility theory and its applications*, Springer, Berlin 2005.

England P.D., Verrall R.J., *Stochastic claims reserving in general insurance*, „British Actuarial Journal" 2002, vol. 8, iss. 3, s. 443–518.

Gisler A., Wütrich M.V., *Credibility for the Chain Ladder Reserving Method*, „ASTIN Bulletin" 2008, vol. 38, iss. 2, s. 565–600.

Klugman S.A., Panjer H.H., Gordon E.W., *Loss Models. From Data to Decisions*, John Wiley & Sons, Inc., Hoboken 2004.

Mack T., *Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates*, „ASTIN Bulletin" 1993, vol. 23, iss. 2, s. 213–225.

Wütrich M.V., Merz M., *Stochastic claims reserving methods in insurance*, Wiley, Chichester 2008.

### Źródła sieciowe

http://prawo.sejm.gov.pl/isap.nsf/download.xsp/WDU20031241151/U/D20031151Lj.pdf (data odczytu: 25.07.2017).

http://www.dbc.wroc.pl/Content/32942/Kujawski_Oszacowanie_Rezerwy_IBNR_Bayesowskim_Modelem_2016.pdf (data odczytu: 20.03.2018).

http://www.kalkulator-ubezpieczeniowy.pl/analiza_rezerw_2005.pdf (data odczytu: 10.03.2018).

http://www-users.mat.umk.pl/~wniem/Ryzyko/RyzykoUB.pdf (data odczytu: 5.05.2015).

* * *

## Analysis of the IBNR reserve credibility predictors

**Abstract**

Chain ladder is one of the simplest and the most frequently used method for the estimation of the IBNR (Incurred but not Reported) reserve. Due to its popularity and wide recognition, it abided many modifications, including its generalization within the credibility theory presented in Gisler and Wütrich (2008). The idea relies on a distinction between so-called individual information and collective information relating

to various parts of a single insurance portfolio, where the share of both sources of information constitutes a subject to estimation. The simulation analysis presented in this paper enabled the comparison of the prediction quality based on the classical chain ladder and of its Bayesian counterpart.

**Keywords:** Bayesian statistics, credibility theory, IBNR, simulation analysis, mean square error