PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 60 — 2009
Inwestycje finansowe i ubezpieczenia – tendencje światowe a polski rynek

Dominika Tomaszewska

Wyższa Szkoła Bankowa we Wrocławiu

ZASTOSOWANIE KOLEKTYWNEGO MODELU RYZYKA ORAZ MODELU GLM DO KALKULACJI REZERWY SZKODOWEJ

1. Wstęp

Ustalanie rezerw szkodowych jest podstawową czynnością dokonywaną w działach aktuarialnych zakładów ubezpieczeń majątkowych. Rezerwa szkodowa tworzona jest w celu pokrycia szkód, które zaszły w danym w okresie sprawozdawczym, ale odszkodowania lub świadczenia nie zostały w nim wypłacone.

Podstawową wielkością oszacowywaną przez aktuariuszy jest wartość oczekiwana przyszłych odszkodowań i świadczeń. Wartość ta powinna stanowić dolną granicę tworzonej rezerwy szkodowej. Wybór odpowiedniej metody daje możliwość znalezienia również rozkładów prawdopodobieństwa badanej zmiennej.

Kalkulacja rezerwy może się odbywać na podstawie klasycznego modelu kolektywnego ryzyka przy dodatkowych założeniach, że zgromadzone dane są realizacjami zmiennych losowych, które mogą być opisane za pomocą modelu GLM. Obszerność opisu zastosowania modeli jest powodem ograniczenia pracy do przedstawienia kalkulacji wartości oczekiwanej przyszłych odszkodowań.

2. GLM − założenia modelu

Uogólniony model liniowy (GLM, *Generalized Linear Model*) określony jest na następujących założeniach[1]:

1. Zbiór niezależnych zmiennych losowych Y_u, $u = 1,2,...,r$ ma postać:

$$Y_u = h(\mathrm{X_u}\beta) + e_u, \tag{1}$$

gdzie: X_u – jest u- tym wierszem macierzy X o wymiarach $r \times p$,
β – jest wektorem parametrów o długości p,
e_u – jest składnikiem losowym takim, że $E[e_u] = 0$.

[1] G. Taylor, *Loss reserving. An actuarial perspective,* Kluwer Academic Publisher, Boston 2000, s. 171.

2. Przekształcenie $h:\Re \to \Re$ jest przekształceniem wzajemnie jednoznacznym. Ponadto zmienne Y_u mają rozkład z wykładniczej rodziny rozkładów (*expotential dispersion familly*).

3. Dla modelu tego zakłada się, że E[eu] = 0, stąd spełniona jest równość:

$$E[Y_u] = h(\mathrm{X_u}\beta) \quad \text{oraz} \quad h^{-1}(E[Y_u]) = \mathrm{X_u}\beta = \eta_u . \tag{2}$$

4. Funkcja $g = h - 1$ nazywana jest funkcją łączącą (*link function*), a η_u – predyktorem liniowym (*linear* predictor). W zastosowaniu modelu do kalkulacji rezerw szkodowych przyjmuje się, iż funkcja łącząca jest funkcją logarytmiczną.

3. Wykładnicza rodzina rozkładów

Zmienne Y_u mają rozkład z wykładniczej rodziny rozkładów o funkcji gęstości o postaci:

$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right), \tag{3}$$

gdzie: a, b, c – są funkcjami,
θ, ϕ – są parametrami nazywanymi odpowiednio: parametrem kanonicznym oraz parametrem dyspersji (skali, rozproszenia),
w_u – jest wagą obserwacji y_u[2].

W zastosowaniach praktycznych przyjmuje się często, iż

$$a(\phi) = \phi / w_u . \tag{4}$$

W przypadku tym wartość oczekiwana i wariancja zmiennej Y_u mają odpowiednio postać:

$$E[Y_u] = b'(\theta_u) = \mu_u , \tag{5}$$

$$Var[Y_u] = b''(\theta_u)\ \phi / w_u . \tag{6}$$

Funkcja $b''(\theta_u)$ nazywana jest funkcją wariancji (*variance function*) i oznaczana jest przez $V(\mu_u)$.

Do wykładniczej rodziny rozkładów o wykładniczej dyspersji należą m.in.:

- rozkład Poissona, gdzie:

$$V(\mu) = \mu^1 , \quad \phi = 1 \ \text{oraz}\ Var[Y] = \phi E[Y], \tag{7}$$

- rozkład gamma(ν, δ), gdzie:

[2] W kalkulacji rezerw zakłada się, że $w_i = 1$ dla wszystkich obserwacji w trójkącie.

$$V(\mu) = \mu^2, \quad \phi = \nu^{-1} \text{ oraz } Var[Y] = \phi \cdot m^2. \tag{8}$$

Parametr ϕ można interpretować w różny osób. Dla rozkładu Poissona $\phi = 1$ i oznacza, że wariancja jest równa średniej; dla rozkładu gamma stała wartość $\phi = \nu^{-1}$ oznacza stały współczynnik zmienności, który wynosi $\phi^{0.5}$.

4. Podstawowe założenia

Zastosowanie modelu GLM odbywa się po uprzedniej klasyfikacji historycznych danych dotyczących wypłaconych odszkodowań. Odszkodowania klasyfikowane są według dwóch wielkości: okresu powstania szkody oraz okresu, w którym nastąpiła wypłata odszkodowania. Klasyfikacja ta prowadzi do prezentowania danych w postaci tzw. trójkąta szkód. Przykład prezentowania danych zawierają tab. 1 i 2.

W metodach kalkulacji rezerw opartych na GLM najczęściej wykorzystuje się następujące założenia:

i. Opisywane zmienne w trójkątach Y_{ij} są generowane przez model multiplikatywny o postaci:

$$Y_{ij} \approx m x_i \; y_j, \tag{9}$$

a więc mogą być wyjaśniane przez parametry charakteryzujące dany okres wypadkowy i oraz okres opóźnienia j;

ii. Zmienne losowe Y_{ij} mają rozkłady z wykładniczej rodziny rozkładów lub wariancję rozkładu można przedstawić za pomocą wartości oczekiwanej przez konstrukcję funkcji wariancji V.

iii. Funkcja łącząca $g = h^{-1}$ jest funkcją logarytmiczną. Wybór ten sprowadza multiplikatywny model:

$$E(Y_{ij}) = \exp(\eta_{ij}) = \exp(\mathrm{X}_{ij}\beta), \tag{10}$$

do postaci liniowej:

$$h^{-1}\left(E(Y_{ij})\right) = \mathrm{X}_{ij}\beta = \eta_{ij}, \tag{11}$$

gdzie: $\mathbf{X}_{ij}$ jest wierszem macierzy, który odpowiada wartości w komórce (i,j) trójkąta.

W szczególnym przypadku predyktor liniowy może mieć postać:

$$\eta_{ij} = \mu + \alpha_i + \beta_j = \mathrm{X}_{ij}\beta. \tag{12}$$

Na podstawie powyższych założeń wartość oczekiwana badanej zmiennej Y_{ij} otrzymywana jest z relacji (12). Wartości przyszłych odszkodowań prawego dolnego trójkąta kalkulowane są zgodnie z formułą.

$$m_{ij} = \hat{E}(Y_{ij}) = \exp\left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j\right). \tag{13}$$

Estymatory parametrów wektora $\beta = \left(\beta_1, ..., \beta_p\right)^t$ otrzymywane są metodą największej wiarygodności. Szczegółowo zagadnienie estymacji parametrów modelu GLM zostało przedstawiono w pracach [Denuit i in. 2001; Mildenhall 1999]. Ze względu na obszerność rozważań temat ten został w pracy pominięty.

5. Zastosowanie modelu ryzyka kolektywnego

Założenia modelu GLM umożliwiają prognozę rezerwy na podstawie danych zawartych w trójkątach zawierających: liczby $N_{i,j}$ odszkodowań wypłaconych z opóźnieniem j, które powstały w roku wypadkowym i, oraz przeciętnej wartości odszkodowania X_{ij} indywidualnego wypłaconego z opóźnieniem j, które jest wynikiem zdarzenia w roku wypadkowym i. W przypadku tym zakłada się, iż liczby odszkodowań oraz ich wartości są niezależnymi zmiennymi losowymi[3]. Metoda ta znalazła zastosowanie w przypadku, kiedy zmienne S_{ij} mają złożony rozkład Poissona ze szkodą indywidualną X_{ij} o rozkładzie gamma[4]. Zarówno rozkład Poissona, jak i gamma można wyrazić za pomocą funkcji określonej wzorem (3). Rozkłady te należą do wykładniczej rodziny rozkładów.

W celu prognozowania rezerwy wykorzystuje się dwa trójkąty danych[5]. Pierwszy z nich zawiera liczby odszkodowań N_{ij} wypłaconych w poszczególnych okresach (i, j), dla których przyjmuje się, iż mają rozkłady Poissona z parametrem $\mu_{ij}^{'}$. Drugi trójkąt zawiera przeciętne wartości odszkodowań indywidualnych X_{ij} wypłaconych w okresach (i, j). W celu prezentacji metody założono, że X_{ij} mają rozkład gamma (ν, δ_{ij}).

Dla zmiennych tych, zgodnie ze wzorami (9) i (10), prawdziwe są relacje:

$$Var[N_{ij}] = \phi' E\left[N_{ij}\right], \text{ gdzie } \phi' = 1. \tag{14}$$

$$Var[X_{ij}] = \phi \cdot m_{ij}^2, \text{ gdzie } \phi = \nu^{-1}. \tag{15}$$

W celu dokonania prognozy konieczne jest określenie struktury predyktorów liniowych dla zmiennych o rozkładzie Poissona i zmiennych o rozkładzie gamma. W przykładzie tym zdefiniowane zostały one odpowiednio:

$$\ln(E[N_{ij}]) = \eta_{ij} = \mu + \alpha_i + \beta_j, \tag{16}$$

$$\ln(E[X_{ij}]) = \eta'_{ij} = \mu' + \alpha'_i + \beta_j'. \tag{17}$$

[3] M. Denuit, J. Dhaene, M. Goovaerts, R. Kass, *Modern actuarial risk theory,* Kluwer Academic Publishers, Boston 2001, s. 208.

[4] Tamże.

[5] Tamże.

Zastosowanie modelu będzie przedstawione na podstawie danych zawartych w tab. 1-2.

Tabela 1. Trójkąt liczb odszkodowań n_{ij}

n_{ij}	1	2	3	4	5	6	7	8	9	10
1	30,00	22,00	27,00	28,00	20,00	18,00	19,00	9,00	12,00	4,00
2	26,00	33,00	31,00	26,00	30,00	23,00	18,00	16,00	9,00	
3	37,00	27,00	29,00	27,00	30,00	24,00	28,00	12,00		
4	26,00	18,00	16,00	22,00	26,00	16,00	12,00			
5	39,00	28,00	21,00	29,00	31,00	31,00				
6	37,00	26,00	25,00	25,00	33,00					
7	34,00	31,00	32,00	30,00						
8	45,00	39,00	36,00							
9	44,00	32,00								
10	52,00									

Źródło: dane umowne.

Tabela 2. Trójkąt losowych przeciętnych wartości odszkodowań $\overline{x}_{ij}$

$\overline{x}_{ij}$	1	2	3	4	5	6	7	8	9	10
1	49,92	46,05	41,81	40,38	38,57	30,56	34,94	41,27	25,36	33,86
2	56,77	48,45	46,33	44,15	41,58	32,64	37,64	45,30	26,59	
3	59,33	54,78	51,66	48,72	45,34	36,95	40,71	47,10		
4	63,65	57,48	56,31	51,55	44,96	38,02	47,35			
5	68,66	63,10	59,40	56,31	52,15	39,61				
6	74,31	68,80	64,13	61,39	56,94					
7	81,96	71,82	68,23	64,91						
8	84,75	76,46	71,73							
9	91,07	80,94								
10	94,64									

Źródło: dane umowne.

Algorytm kalkulacji rezerwy może być zapisany w następujący sposób:

Krok 1. Predykcja brakujących elementów dolnego prawego trójkąta − liczby szkód na podstawie GLM.

Krok 2. Predykcja brakujących elementów dolnego prawego trójkąta – przeciętnych wartości szkód indywidualnych na podstawie GLM.

Krok 3. Oszacowanie wartości średniej $E[S_{ij}]$ na podstawie otrzymanych wartości $E[X_{ij}]$ oraz $E[N_{ij}]$.

Krok 4. Oszacowanie R_i oraz R. Ustalenie rezerwy.

W celu kalkulacji rezerwy korzysta się z podstawowej własności modelu ryzyka kolektywnego określonej wzorem:

$$E[S_{ij}] = E[X_{ij}]E[N_{ij}]. \tag{18}$$

Wartość rezerwy kalkulowana jest na podstawie równań:

$$R_i = S_i = \sum_{j \in \Delta_i}^{n} S_{ij}, \tag{19}$$

$$R = S = \sum_{i=1}^{n} R_i, \tag{20}$$

gdzie: Δ – oznacza prawy dolny trójkąt, dla którego prognozowane są wartości szkód indywidualnych S_{ij}.

Wartość oczekiwaną oraz wariację rozkładu R można znaleźć z wykorzystaniem formuły:

$$E[R_i] = \sum_{j \in \Delta_i}^{n} E[S_{ij}]. \tag{21}$$

6. Prognoza rezerwy − przykład numeryczny

Jak zaznaczono we wstępie, obszerność opisu zastosowania modeli jest powodem ograniczenia pracy do przedstawienia kalkulacji wartości oczekiwanej przyszłych odszkodowań.

Estymacji parametrów predyktorów liniowych oraz estymacji wartości oczekiwanych zmiennych $N_{ij.}$ oraz X_{ij} dokonuje się na podstawie przyjętego założenia dotyczącego struktury predyktorów liniowego dla modelu Poissona i gamma, które są określone wzorami (16), (17). Estymacji parametrów dokonuje się na bazie procedur opisanych w [McCullagh, Nelder 1999; Mildenhall 1999]. Estymowane wartości parametrów przedstawione są w tab. 3.

Dla modelu Poissona parametr dyspersji $\phi' = 1$, w przypadku modelu gamma estymowana wartość $\hat{\phi} = 0{,}0005511$. Na podstawie otrzymanych ocen parametrów możliwe jest określenie wartości predyktorów liniowych dla modelu Poissona GLM oraz gamma GLM. Wartości predyktorów liniowych są podstawą do obliczenia oczekiwanych wartości zmiennych N_{ij} i X_{ij}, w tym danych prognozowanych dla dolnych prawych trójkątów (zob. tab. 4-5).

Kolejnym krokiem jest budowa dolnego prawego trójkąta przyrostowych wartości S_{ij} odszkodowań do wypłaty na podstawie iloczynu estymatorów wartości oczekiwanych N_{ij} i X_{ij} (dla dolnych prawych trójkątów) zgodnie ze znaną formułą (17) charakteryzującą model kolektywny.

Tabela 3. Estymatory parametrów modeli: Poissona GLM oraz gamma GLM

POISSON GLM			GAMMA GLM		
Parametry	wartość	średni błąd szacunku	parametry	wartość	średni błąd szacunku
mu	3,3738	0,0937	mu'	3,9249	0,0113
alfa2	0,1362	0,1006	alfa2'	0,0808	0,0111
alfa3	0,2000	0,1011	alfa3'	0,1707	0,0116
alfa4	-0,1886	0,1141	alfa4'	0,2372	0,0121
alfa5	0,2047	0,1069	alfa5'	0,3107	0,0128
alfa6	0,1535	0,1136	alfa6'	0,4015	0,0136
alfa7	0,2449	0,1193	alfa7'	0,4643	0,0148
alfa8	0,4615	0,1229	alfa8'	0,5134	0,0164
alfa9	0,3664	0,1433	alfa9'	0,5798	0,0193
alfa10	0,5775	0,1674	alfa10'	0,6252	0,0261
beta2	-0,2169	0,0840	beta2'	-0,1040	0,0111
beta3	-0,2401	0,0895	beta3'	-0,1614	0,0116
beta4	-0,2052	0,0951	beta4'	-0,2148	0,0121
beta5	-0,1233	0,0991	beta5'	-0,2919	0,0128
beta6	-0,3458	0,1146	beta6'	-0,5181	0,0136
beta7	-0,4639	0,1325	beta7'	-0,3601	0,0148
beta8	-0,9770	0,1787	beta8'	-0,2133	0,0164
beta9	-1,0928	0,2313	beta9'	-0,7082	0,0193
beta10	-1,9875	0,5086	beta10'	-0,4027	0,0261

Źródło: opracowanie własne.

Tabela 4. Wartości estymowane $\hat{E}\ [N_{ij}]$

$\hat{E}\ [N_{ij}]$	1	2	3	4	5	6	7	8	9	10
1	29,19	23,50	22,96	23,77	25,80	20,65	18,35	10,99	9,79	4,00
2	33,45	26,93	26,31	27,24	29,57	23,67	21,03	12,59	11,21	4,58
3	35,65	28,70	28,04	29,04	31,51	25,23	22,42	13,42	11,95	4,89
4	24,17	19,46	19,01	19,69	21,37	17,10	15,20	9,10	8,10	3,31
5	35,82	28,83	28,17	29,17	31,66	25,35	22,52	13,48	12,01	4,91
6	34,03	27,40	26,77	27,72	30,08	24,08	21,40	12,81	11,41	4,66
7	37,29	30,02	29,33	30,37	32,96	26,39	23,45	14,04	12,50	5,11
8	46,30	37,28	36,42	37,71	40,93	32,77	29,12	17,43	15,52	6,35
9	42,10	33,90	33,12	34,29	37,22	29,79	26,48	15,85	14,12	5,77
10	52,00	41,86	40,90	42,35	45,97	36,80	32,70	19,58	17,43	7,13

Źródło: opracowanie własne.

Tabela 5. Wartości estymowane $\hat{E}\,[X_{ij}]$

$\hat{E}\,[X_{ij}]$	1	2	3	4	5	6	7	8	9	10
1	50,65	45,64	43,10	40,86	37,82	30,17	35,33	40,92	24,94	33,86
2	54,91	49,49	46,73	44,30	41,01	32,71	38,31	44,36	27,04	36,71
3	60,08	54,14	51,12	48,46	44,87	35,78	41,91	48,53	29,59	40,16
4	64,20	57,86	54,63	51,79	47,95	38,24	44,79	51,87	31,62	42,92
5	69,10	62,27	58,80	55,75	51,61	41,16	48,21	55,83	34,03	46,20
6	75,67	68,20	64,39	61,05	56,51	45,08	52,79	61,14	37,27	50,59
7	80,57	72,61	68,56	65,00	60,17	47,99	56,21	65,09	39,68	53,86
8	84,63	76,27	72,01	68,27	63,20	50,41	59,04	68,37	41,68	56,58
9	90,44	81,51	76,96	72,96	67,54	53,87	63,09	73,07	44,54	60,46
10	94,64	85,29	80,54	76,35	70,68	56,38	66,02	76,46	46,61	63,27

Źródło: opracowanie własne.

W efekcie otrzymujemy kolejne dane przedstawione tab. 6, które są prognozami wartości oczekiwanych odszkodowań S_{ij} do wypłaty.

Tabela 6. Wartości estymowane $\hat{E}\,[S_{ij}]$ – model kolektywny

$\hat{E}\,[S_{ij}]$	1	2	3	4	5	6	7	8	9	10	$\hat{E}\,[S_i]$
1											0,00
2										168,27	168,27
3									353,64	196,21	549,85
4								471,97	256,26	142,18	870,41
5							1 085,65	752,72	408,69	226,75	2 473,81
6						1 085,54	1 129,67	783,25	425,26	235,95	3 659,67
7					1 983,38	1 266,33	1 317,81	913,68	496,08	275,24	6 252,52
8				2 574,78	2 587,00	1 651,72	1 718,87	1 191,76	647,06	359,01	10 730,20
9			2 548,69	2 502,14	2 514,02	1 605,12	1 670,38	1 158,14	628,80	348,88	12 976,17
10		3 570,56	3 293,99	3 233,84	3 249,18	2 074,50	2 158,84	1 496,80	812,68	450,90	20 341,29
										Ogółem	58 022,19

Źródło: opracowanie własne.

Sumowanie elementów trójkąta dla danego roku odniesienia daje wartość oczekiwaną przyszłych odszkodowań S_i dla roku odniesienia i. Suma wszystkich elementów trójkąta stanowi wartość oczekiwaną odszkodowań S do wypłaty ogółem.

7. Podsumowanie

W literaturze aktuarialnej coraz częściej podkreśla się fakt, iż ustalanie rezerwy szkodowej w wysokości równej wartości oczekiwanej przyszłych odszkodowań i świadczeń jest poziomem niewystarczającym. W związku z tym ustalona rezerwa powinna zawierać margines bezpieczeństwa. Decyzje o poziomie rezerwy można podjąć na podstawie informacji wariancji, odchyleniu standardowym, współczynniku zmienności lub o rozkładzie prawdopodobieństwa zmiennej *S.* Zastosowanie stochastycznego modelu GLM daje możliwość określenia rozkładu przyszłych odszkodowań. Oszacowanie rozkładu pozwala na określenie prawdopodobieństwa, iż wartość odszkodowań i świadczeń przekroczy rozważany poziom rezerwy szkodowej.

Wprowadzenie do kalkulacji rezerwy dwóch trójkątów danych umożliwia zastosowanie kolektywnego modelu ryzyka. Obserwacja danych w przedstawionych trójkątach oraz prognoza brakujących elementów trójkątów wnosi dodatkowe informacje, których nie można uzyskać na podstawie jednego trójkąta zawierającego przyrostowe wartości odszkodowań wypłacanych z opóźnieniem w stosunku do okresu odniesienia. Wykorzystanie jednego trójkąta jest dominującym podejściem opisywanym w literaturze aktuarialnej, jednak posiadanie danych uporządkowanych w dwóch trójkątach pozwala również ocenić i prognozować tzw. zjawisko eskalacji liczby i wartości roszczeń, które może wynikać np. ze wzrostu świadomości ubezpieczeniowej osób ubezpieczających się lub z procesu inflacji roszczeń. Informacje mogą również brać udział w podejmowaniu decyzji w stosunku do wartości rezerw szkodowych.

Literatura

Denuit M., Dhaene J., Goovaerts M., Kass R., *Modern actuarial risk theory,* Kluwer Academic Publishers, Boston 2001.

McCullagh P., Nelder J.A., *Generalized linear models,* Chapman & Hall, Cambridge 1999.

Mildenhall S.A., *Systematic relationship between minimum bias and Generalized Linear Model*, PCAS 1999 vol. XXXVI.

Wang S.S., *Aggregation of correlated risk portfolios: models and algorithms,* PCAS 1998, vol. XXXV.

IMPLEMENTATION OF COLLECTIVE RISK MODEL AND GENERALIZED LINEAR MODEL FOR A LOSS RESERVE CALCULATION

Summary

This article describes an implementation of Generalized Linear Model and Collective Risk Model for a loss reserve calculation. A loss reserve is the main value calculated by actuaries. A process of calculating starts from building two data triangles. The first one contains the number of losses, the second one contains average values of losses. The data is presented in order by an accident year and by a development year. The article presents also an application of Fast Fourier Transform to calculate a loss distribution.