

Tomasz Michalski

Szkoła Główna Handlowa w Warszawie

Krystyna Twardowska

Politechnika Warszawska

Barbara Tylutki

Zespół Szkół nr 1 w Rzeszowie

ADAPTACYJNA FILTRACJA W UBEZPIECZENIACH

1. Wstęp

Problemy filtracji w odniesieniu do niegaussowskich układów były rozważane m.in. w pracach Beneša i Karatzasa, De Santisa, Germaniego i Raimondiego, Hausmanna, Makowskiego, Sowersa i Makowskiego (por. [3; 7; 9; 14; 18]). Problemy filtracji dla niegaussowskich układów, występujących w zagadnieniach ubezpieczeniowych, były rozważane w pracach De Jonga i Zehnwirtha, Malcolm, Jamea i Elliota, Zehnwirtha, w książce Goovaerts i innych (por. [6; 13; 21; 8]). Adaptacyjną filtrację w tradycyjnym podejściu można znaleźć w książkach Jazwskiego [10], Andersona i Moore'a, Beneviste'a, Metiviera i Prioureta (por. [10; 2; 4]), a dla rozkładów niegaussowskich np. w pracy Laitonisa i Giannakopulosa [12].

W niniejszym artykule opieramy się na wynikach dotyczących filtracji w odniesieniu do niegaussowskich rozkładów zawartych w pracy Sowersa i Makowskiego (por. [18]). Uśrednioną adaptacyjną filtrację można znaleźć w pracach Polyaka (por. [17]) dla nieskorelowanych szumów oraz Yina i Zhu (por. [20]) dla szumów skorelowanych. Analiza zbieżności filtru jest omówiona w najnowszej pracy autorów (por. [15]), gdzie uogólniono rezultaty z pracy Sowersa i Makowskiego (por. [18]) oraz wykorzystano definicje Andersona i Moore'a (por. [1]) dla zmiennych w czasie niegaussowskich układów.

2. Adaptacyjna uśredniona filtracja

Rozważmy stochastyczny proces $\{(X_t, Y_t)\} \in R^n \times R^m$, zdefiniowany dla $t \in I$, gdzie I jest podzbiorem zbioru Z liczb całkowitych, tzn. t oznacza dyskretne momenty czasowe. Jeden krok czasowy może oznaczać różne przedziały czasowe, np. rok, kwartał, miesiąc itd.

Postawmy następujący problem: znając realizację procesu $\{Y_t : t \in I\}$ do momentu $s \in I$, chcemy jak najdokładniej odtworzyć wartość procesu $\{X_t : t \in I\}$ w momencie $s + u$. Jeśli $u > 0$, to powyższy problem nazywa się prognozowaniem, a gdy $u = 0$ – filtracją.

Dla dowolnego momentu $s \in I$ szukamy więc estymatora \widehat{X}_{s+u} zmiennej X_{s+u} w postaci dowolnej funkcji od obserwowanej realizacji. W zagadnieniach filtracji (dla $u = 0$) zmienna \widehat{X}_s powinna być mierzalna względem σ -ciała \mathbf{Y}_s , gdzie

$$\mathbf{Y}_s = \sigma\{Y_t : t \leq s, t \in I\}.$$

Wśród estymatorów należy wskazać najlepszy, tzn. taki, dla którego wartość

$$E\left\{(X_{s+u} - \widehat{X}_{s+u})(X_{s+u} - \widehat{X}_{s+u})^T\right\},$$

jest minimalna (gdzie T oznacza transponowanie).

W najprostszym przypadku dyskretnej liniowej filtracji Kalmana dany jest układ równań (por. np. Anderson i Moore [2], Jazwinski [10]).

$$X_{t+1} = \Phi_t X_t + v_t, \quad (1)$$

$$Y_t = \Theta_t X_t + w_t, \quad (2)$$

gdzie $\{(X_t, Y_t)\} \in R^n \times R^m$ dla $t \in [t_0, T]$, $\Phi_t \in R^{n \times n}$, $\Theta_t \in R^{m \times n}$. Ponadto v_t, w_t to odpowiednio wektory szumów wejściowych i pomiarowych o wartościach oczekiwanych równych zeru i znanych macierzach kowariancji, Φ_t oznacza znaną macierz układu, Θ_t oznacza znaną macierz (np. zawierającą zera i jedynek; jedynek mamy tam, gdzie nastąpił pomiar, zero – gdzie nie nastąpił). Ciąg $\{w_{t_0}, w_{t_0+1}, \dots, w_T, v_{t_0}, v_{t_0+1}, \dots, v_T\}$ jest ciągiem niezależnych zmiennych gaussowskich o znanych macierzach kowariancji

$$E\{v_t v_t^T\} = R_1(t), \quad E\{v_t\} = 0, \quad (3)$$

$$E\{w_t w_t^T\} = R_2(t), \quad E\{w_t\} = 0, \quad t \in [t_0, T]. \quad (4)$$

Zmienna X_{t_0} ma również rozkład gaussowski o znanej wartości oczekiwanej $E\{X_{t_0}\} = m$ i macierzy kowariancji

$$E\{(X_{t_0} - m)(X_{t_0} - m)^T\} = R_0 = R_{t_0}, \quad (5)$$

przy czym X_{t_0} nie zależy od $\{v_t, w_s : s, t \in [t_0, T]\}$. Macierz $R_2(t)$ jest dodatnio określona.

Wprowadźmy oznaczenia:

$$\hat{X}_{t|t} = E\{X_t | Y_t\}, \quad \hat{X}_{t|t-1} = E\{X_t | Y_{t-1}\}, \quad (6)$$

$$S_{t|t} = E\{(X_t - \hat{X}_{t|t})(X_t - \hat{X}_{t|t})^T | Y_t\}, \quad (7)$$

$$S_{t|t-1} = E\{(X_t - \hat{X}_{t|t-1})(X_t - \hat{X}_{t|t-1})^T | Y_{t-1}\}. \quad (8)$$

Wiadomo (por. [2; 10]), że dla gaussowskiego problemu filtracji Kalmana estymator procesu X_t jest wyliczany z następującego układu równań:

$$\hat{X}_{t+1|t} = \Phi_t \hat{X}_{t|t-1} + K_t (Y_t - \Theta_t \hat{X}_{t|t-1}), \quad (9)$$

$$\hat{X}_{0| -1} = E\{X_{t_0}\} = m,$$

gdzie

$$K_t = \Phi_t S_{t|t-1} \Theta_t^T [\Theta_t S_{t|t-1} \Theta_t^T + R_2(t)]^{-1}, \quad T > t \geq t_0 \quad (10)$$

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} - S_{t|t-1} \Theta_t^T [\Theta_t S_{t|t-1} \Theta_t^T + R_2(t)]^{-1} (Y_t - \Theta_t \hat{X}_{t|t-1}), \quad (11)$$

$$S_{t|t} = S_{t|t-1} - S_{t|t-1} \Theta_t^T [\Theta_t S_{t|t-1} \Theta_t^T + R_2(t)]^{-1} \Theta_t S_{t|t-1} \quad (12)$$

z warunkiem początkowym $S_{0| -1} = R_{t_0}$. Macierz K_t jest tzw. macierzą zysku Kalmana.

Ze wzorów (9)-(12) widać, że gdy zmienimy nieco nasze zadanie (ze względów rachunkowych) i poszukujemy informacji o X_t na podstawie pomiarów doko-

nanych do chwili $t - 1$, to faktycznie otrzymujemy rozwiązanie zadania jednokrokowego prognozowania.

Do zdefiniowania problemu adaptacyjnej filtracji określamy tzw. przewidywane reszty (precyzyjniej, przewidywaną wartość błędu reszty):

$$r(t+1|t) = Y_{t+1} - E\{Y_{t+1}|Y_s\} = Y_{t+1} - \Theta_{t+1}\hat{X}_{t+1}. \quad (13)$$

Korekta estymatora stanu będzie proporcjonalna do przewidywanych reszt, a stała proporcjonalności jest zyskiem Kalmana. Adaptacyjna filtracja jest adaptacyjna w następującym sensie: jak długo estymatory reszt nie przekraczają swojej wariancji, wejściowy szum równa się zeru i filtr działa satysfakcjonująco. Gdy estymatory reszt są większe od ich wariancji, wówczas filtr jest rozbieżny. Zamiast dyskutować o różnicach pomiędzy prognozowanymi wartościami stanu układu a ich pomiarami dokonanymi później, proponujemy uwzględnić ich różnice w rzeczywistym czasie wykonywania obliczeń i pozwolić im wpływać na określenie odpowiedniego szumu.

W tej pracy wykorzystujemy procedurę adaptacyjną zaprezentowaną przez Yina i Zhu (por. [20]) dotyczącą skorelowanych szumów oraz procedurę Polyaka (por. [17]) dla nieskorelowanych szumów. Jak widać z przykładów zaprezentowanych przez Yina i Zhu (1992), procedura uśredniająca daje asymptotycznie lepszy wskaźnik zbieżności niż klasyczna metoda adaptacyjna. Mianowicie: do formuły definiującej estymator \hat{X}_{t+1} dodajemy tzw. uśredniający estymator:

$$\begin{aligned} \bar{\hat{X}}_{n+1} &= \frac{1}{n+1} \left[\sum_{j=1}^n \hat{X}_j + \hat{X}_{n+1} \right] = \frac{n}{n+1} \bar{\hat{X}}_n + \frac{1}{n+1} \hat{X}_{n+1} = \\ &= \bar{\hat{X}}_n - \frac{1}{n+1} \bar{\hat{X}}_n + \frac{1}{n+1} \hat{X}_{n+1}, \end{aligned} \quad (14)$$

gdzie \bar{X} oznacza średnią.

3. Model kolektywnego ryzyka

W modelu ryzyka kolektywnego badamy rozkład liczby szkód i wartości (wysokości) szkód X_i , $i = 1, 2, \dots, N(t)$, oraz całkowitą wypłatę $S(t)$:

$$S(t) = X_1 + X_2 + \dots + X_{N(t)}, \quad (15)$$

gdzie $N(t)$ określa liczbę szkód w okresie $[t_0, T]$; zakładamy, że jest to zmienna losowa o rozkładzie Poissona $\Pi(\lambda)$, a X_i , $i = 1, 2, \dots, N(t)$ są niezależnymi zmiennymi losowymi o takich samych rozkładach (będziemy je oznaczać przez X), np.

Weibulla, Pareto (który dopuszcza katastroficzne roszczenia), logarytmiczno-normalny lub Burra. W niniejszej pracy będziemy rozpatrywać rozkłady Weibulla.

Zakładamy ponadto, że wszystkie zmienne $N(t)$, $X_1(t)$, $X_2(t)$, ..., $X_{N(t)}(t)$ są wzajemnie niezależne. Zakładamy losowe momenty pojawiania się szkód i losową wysokość szkód. Prawdopodobieństwo pojawiania się dużych szkód nie jest zaniedbywane, dlatego rozpatruje się powyższe rozkłady o tzw. ciężkich ogonach.

Przypomnijmy podstawowe charakterystyki interesującego nas rozkładu Weibulla (α, c) [11] z gęstością prawdopodobieństwa o postaci:

$$f_X(x) = c\alpha x^{c-1} \exp\{-\alpha x^c\} \quad \text{dla } x > 0, 0 < c < 1, 0 < \alpha. \quad (16)$$

Ma on odpowiednio wartość oczekiwaną i wariancję równą:

$$E\{X\} \alpha^{-\frac{1}{c}} \Gamma\left(1 + \frac{1}{c}\right) \quad \text{dla } 0 < c < 1, \quad (17)$$

$$V\{X\} = \left(\frac{1}{\alpha}\right)^{\frac{2}{c}} \left[\Gamma\left(1 + \frac{2}{c}\right) - \left(\Gamma\left(1 + \frac{1}{c}\right)\right)^2 \right] \quad \text{dla } 0 < c < 1. \quad (18)$$

Po skorzystaniu z typowych własności warunkowej wartości oczekiwanej wyliczamy dla $S(t)$ (por. [16]):

$$\begin{aligned} E\{S(t)\} &= E\{E\{S(t)|N(t)\}\} = E\{N(t)E\{X\}\} = \\ &= E\{X\}E\{N(t)\} = \mu E\{N(t)\}, \end{aligned} \quad (19)$$

$$\begin{aligned} V\{S(t)\} &= V\{V\{S(t)|N(t)\}\} + V\{E\{S(t)|N(t)\}\} = \\ &= E\{N(t)V\{X\}\} + V\{E\{X\}N(t)\} = \\ &= E\{N(t)V\{X\} + (E\{X\})^2 V\{N(t)\}\} = \\ &= \sigma^2 V\{N(t)\} + \mu^2 V\{N(t)\}, \end{aligned} \quad (20)$$

gdzie $E\{X\} = \mu$ oraz $V\{X\} = \sigma^2$.

4. Filtracja w ubezpieczeniach

Rozważmy ponownie model (1)-(2). Chcemy znaleźć estymator wektora $\{X_t : t \in I\}$ metodą najmniejszych kwadratów (por. [8, s. 260]). Zastosujemy ten model do prognozowania liczby wypłat oraz jako drugie zadanie – do prognozowania całkowitej wypłaty w modelu kolektywnego ryzyka. Pokażemy, jak otrzymać macierze Φ_t i Θ_t dla obydwu modeli (dla $N(t)$ i $S(t)$, oddzielnie).

Krok 1°. Konstruujemy dwa wektory obserwacji Y : dla liczby wypłat $N(t)$ o rozkładzie Poissona oraz osobno dla całkowitych wypłat $S(t)$ o rozkładzie Weibulla. Zmienne $N(t)$ i $S(t)$ będziemy oznaczać wspólnym symbolem Y_t .

Rozważymy szczególny przypadek, czyli następujące równanie, z którego utworzymy równanie obserwacji (2) (por. [8]):

$$Y_t = \lambda(t) + u_j(t), \quad j = 1, \dots, t \quad (21)$$

ponieważ

$$E[Y_t] = \lambda(t). \quad (22)$$

dla rozkładu Poissona, gdzie λ oznacza parametr Poissona.

Dla całkowitych wypłat tworzymy analogicznie równanie obserwacji, ale zamiast ze wzoru (22), skorzystamy ze wzorów (17) i (19) na wartość oczekiwaną $S(t)$ dla rozkładu Weibulla zmiennych losowych $X_t(t)$ i przyjmiemy

$$E[X_t] = \alpha(t) \frac{-1}{c(t)} \Gamma \left(1 + \frac{1}{c(t)} \right). \quad (23)$$

Aby zapisać równanie (2) filtracji na obserwację, tj.

$$Y_t = \Theta_t X_t + u_t,$$

w przypadku liczby wypłat $N(t)$, modelujemy wektor $\lambda(t)$ jako $\Theta_t X_t$, gdzie Θ_t jest znaną zależną od czasu macierzą (tzn. zawierającą zera i jedynki; zob. pkt 2), natomiast X_t jest poszukiwanym wektorem.

Dla całkowitych wypłat analogicznie tworzymy równanie obserwacji, ale zamiast $\lambda(t)$ rozważamy odpowiednią wartość oczekiwaną.

Krok 2°. Tworzymy równania stanu postaci (1) oddzielnie dla $N(t)$ i dla $S(t)$. Opiszemy teraz, jak uzyskać macierze Φ_t .

W przypadku liczby wypłat $N(t)$ bierzemy następujące wartości: dla każdego kwartału wypłacania odszkodowań uaktualniamy wielomian interpolacyjny Newtona odpowiedniego rzędu (zależnie od liczby rozpatrywanych kwartałów) przechodzący przez punkty $(t_i, N(t_i))$ dla $i = 0, \dots, k$ dla k kwartałów. Otrzymujemy

$$N(t) = N(t_0) + \frac{\Delta N(t_0)}{\Delta t} \frac{t-t_0}{1!} + \frac{\Delta^2 N(t_0)}{\Delta t^2} \frac{(t-t_0)(t-t_1)}{2!} + \dots + \frac{\Delta^k N(t_0)}{\Delta t^k} \frac{(t-t_0)(t-t_1)\dots(t-t_{k-1})}{k!}, \quad (24)$$

gdzie

$$\Delta^k N(t) = N(t+kh) - \binom{k}{1} N[t+(k-1)h] + \binom{k}{2} N[t+(k-2)h] + \dots + (-1)^k \binom{k}{k} N(t)$$

oraz $h = \Delta t$.

Dla całkowitych wypłat tworzymy równanie stanu, zakładając, że

$$S(t) = \alpha_0(t) + \sum_{i=1}^5 \alpha_i(t) f_i(t), \quad (25)$$

gdzie $f_1(t), \dots, f_5(t)$ oznaczają odpowiednio wskaźnik monitorowania działalności towarzystwa ubezpieczeniowego, wskaźnik zabezpieczenia wypłat odszkodowań, wskaźnik płacenia zobowiązań (w dniach), współczynnik szkodowości na udziale własnym oraz kapitały własne do funduszu ubezpieczeniowego. Parametry: $\alpha_0(t), \alpha_1(t), \dots, \alpha_5(t)$ zostały wyliczone za pomocą modelu regresji wielokrotnej w pakiecie komputerowym STATGRAPHICS. Odpowiednie wartości dla $f_1(t), \dots, f_5(t)$ są zadane w zbiorze danych dla modelu.

5. Trójkąt wypłat

Trójkąt wypłat (*run-off triangle*) tworzy się w celu wyliczenia przyszłych wypłat. Stosujemy go do oceny i szacowania wyników IBNR (*incurred but not reported* – poniesionych, lecz nie udokumentowanych). Na osi pionowej znajduje się rok początkowy (tworzący fakty), a na poziomej osi pojawiają się lata udokumentowanych wypłat. Prowadzi to do wyliczenia nieskumulowanych lub skumulowanych wartości trójkąta, pod warunkiem że znamy np. procentowy prognozowany rozkład

udokumentowania wypłat. Chcemy więc dokonać prognozy wysokości całkowitych wypłat $S(t)$, które miałyby miejsce w następnych miesiącach, kwartałach czy też latach (w zależności od przyjętego kroku czasowego). Zagęszczenie momentów obserwacji daje dokładniejszy obraz badanego obiektu, natomiast w oficjalnych statystykach, np. w instytucjach sprawujących nadzór ubezpieczeniowy w danym kraju, dane są gromadzone w ujęciu kwartalnym i rocznym. Najbardziej popularny w porównaniach jest okres roczny. Dalej, w celu ustalenia uwagi, przyjmijmy roczny interwał czasu.

Załóżmy, że mamy dane wysokości całkowitych wypłat ubezpieczyciela, które pojawiały się w latach 1, 2, ..., t oraz w tych latach były udokumentowane. Budujemy następujący trójkąt dla wysokości wypłaty \tilde{s}_{js} . Zwykle wartości \tilde{s}_{js} dla $j + s \leq t + 1$ są obserwowane, podczas gdy pozostałe wartości są przewidywane. Można zbudować podobny trójkąt dotyczący liczby roszczeń:

Trójkąt wypłat obserwowanych wartości $S(t)$

Rok wystąpienia	Rok zaobserwowania lub prognozowania			
	1	2	$t - 1$	t
1	\tilde{s}_{11}	\tilde{s}_{12}	$\tilde{s}_{1,t-1}$	$\tilde{s}_{1,t}$
2	\tilde{s}_{21}	\tilde{s}_{22}	$\tilde{s}_{2,t-1}$	
...		
T	\tilde{s}_{T1}			

Poszczególne elementy \tilde{s}_{js} oznaczają wysokość wypłat, które pojawiły się w roku początkowym j . Rok s jest kolejnym rokiem, dla którego roszczenia z roku j są udokumentowane lub prognozowane (jest on rokiem kalendarzowym $j + s - 1$) dla $j, s = 1, 2, \dots$; zostały one udokumentowane w roku $j + s - 1 \leq t$. Przykładowo pierwsza kolumna oznacza roszczenia, które zostały w pełni udokumentowane i wypłacone przed końcem roku pojawienia się straty. W rozwiązaniach modelowych uwzględniamy naturalne opóźnienia o jeden lub kilka okresów, a konstruujemy trójkąt rozchodów ze względu na to, że sytuacje rynkowe ubezpieczeń charakteryzują się różnorodnością w procesie realizacji roszczeń. Trójkąt pozwala na wyłonienie tych różnorodności. Wyliczamy trójkąt wypłat dla prognozowanych wysokości wypłat na lata 1998-2002, gdyż tylko dla tego okresu mieliśmy odpowiednie dane.

6. Niegaussowska filtracja

Filtr Kalmana dany wzorem (9), rozważany w poprzednim punkcie, jest filtrem o minimalnej wariancji, niezależnie od tego, czy występujące zmienne losowe są gaussowskie (por. [2, § 3.2]). Byle tylko występujące rozkłady były całkowicie opisane

przez swoje pierwsze i drugie momenty. W przypadku odstępstw od rozkładu gaussowskiego wyprowadzone wzory (9)-(12) są dalej słuszne, ale dają one niepełną informację o wyższych momentach rozkładu prawdopodobieństwa stanu układu, nie pozwalając wnioskować o funkcjach gęstości rozkładu prawdopodobieństwa.

Gęstość rozkładu prawdopodobieństwa warunkowego $p_{x|y}(x|y)$, jeśli wartość y jest ustaloną, a x jest zmienną, zawiera pełną informację o X , wynikającą z tego, że $Y = y$. Ponieważ jest to funkcja, a nie pojedyncza liczba rzeczywista, powstaje pytanie, czy nie byłoby lepiej, tracąc część informacji, przyjąć prostszą charakterystykę niż funkcja, np. przyjmując pojedynczą estymatę wartości przyjmowanej przez X przy $Y = y$. Użyteczna okazuje się w takich sytuacjach tzw. estymata o minimalnej wariancji.

Przez \hat{x} oznaczmy estymatę wartości przyjmowanej przez X , wiedząc, że $Y = y$. Zdefiniujemy estymatę (\hat{x}) o minimalnej wariancji jako taką, dla której zachodzi nierówność:

$$E\{\|X - \hat{x}\|^2 | Y = y\} \leq E\{\|X - z\|^2 | Y = y\} \quad (26)$$

dla wszystkich wektorów z określonych na podstawie y .

Oczywiście, w przypadku rozkładów niegaussowskich mogą istnieć estymatory nieliniowe lepsze niż filtr Kalmana (liniowy). Dopuszcza się także ogólniejsze założenie o szumach wejściowym i wyjściowym oraz to, że mogą one być skorelowane. Możemy używać różnych metod do rozważania skorelowanych gaussowskich szumów.

Autorzy w pracy [15] uogólniają twierdzenie Sowersa i Makowskiego (por. [18]), które mówi, że przy pewnych założeniach, dla niegaussowskich rozkładów F , różnica między błędem prognozy otrzymanej metodą filtru Kalmana i filtru o minimalnej wariancji dąży do zera niezależnie od F , gdy $t \rightarrow \infty$. Oznacza to, że filtr Kalmana zastosowany w punkcie 4 również prowadzi do rozwiązania problemu.

Przepiszmy układ (1)-(2) w postaci:

$$X_{k+1} = \Phi(t_{k+1}, t_k) X_k + v_k, \quad (27)$$

$$Y_k = \Theta(t_k) X_k + w_k, \quad (28)$$

dla $k = 0, 1, 2, \dots$

Zdefiniujmy (por. [10, s. 231-232]) tzw. macierz informacyjną dla układu (27)-(28)

$$\rho(k, 1) = \sum_{i=1}^k F^T(i, k) \Theta^T(i) R_{2,i}^{-1} \Theta(i) F(i, k) \quad (29)$$

i tzw. macierz sterującą dla układu (27)-(28)

$$\phi(k, 0) = \sum_{i=0}^{k-1} F(k, i+1) R_{1, i+1}(i) F^T(k, i+1). \quad (30)$$

gdzie $F(i, k)$ oznacza macierz przejścia dla zmodyfikowanego równania (27), bez szumu wejściowego (por. [10, s. 231-232]).

Korzystając z powyższych definicji, możemy uogólnić w odniesieniu do układów zmiennych w czasie następujące definicje stabilizowalności i wykrywalności (zob. [1; 10]), które będą potrzebne w twierdzeniu 1.

Definicja 1 [5, § 3.4]

Para macierzy $(F; G)$, gdzie $F \in R^{n \times n}$ i $G \in R^{n \times m}$ jest stabilizująca, jeśli istnieje macierz $K \in R^{n \times n}$, taka że wartości własne macierzy $F - GK$ leżą wewnątrz zbioru $\{z: |z| < 1\}$, gdzie $z \in C$ (C – zbiór liczb zespolonych). Mówimy, że para macierzy $(H; F)$ jest wykrywalna, jeśli para $(H^T; F^T)$ jest stabilizująca.

Można sformułować następujące twierdzenie.

Twierdzenie 1 [15]

Założmy, że stan początkowy X_0 ma rozkład F , który posiada skończony pierwszy i drugi moment odpowiednio μ oraz σ^2 , jest całkowalny z kwadratem i jest niezależny od skorelowanych szumów v_t oraz w_t , spełniających odpowiednie warunki. Macierze kowariancji oraz σ są dodatnio określone. Para macierzy (Φ_t, Θ_t) , występująca w równaniach (27) i (28), jest wykrywalna, a para $(\bar{\Phi}_t, \bar{\Theta}_t)$ jest stabilizująca, gdzie

$$\bar{\Phi}_t = \Phi_t - \Gamma_{vw} \Gamma_w^{-1} \Theta_t, \quad \bar{\Theta}_t = \Gamma_v - \Gamma_{vw} \Gamma_w^{-1} \Gamma_{vw} \quad (31)$$

dla $\Gamma = \begin{pmatrix} \Gamma_v & \Gamma_{vw} \\ \Gamma_{vw} & \Gamma_w \end{pmatrix}$ (Γ jest macierzą korelacji).

Wówczas dla

$$\varepsilon_t = E \left\{ \left\| \hat{X}_t^0 - \hat{X}_t^K \right\|^2 \right\}, \quad (32)$$

gdzie \hat{X}_t^0 oznacza estymatę o minimalnej wariancji i \hat{X}_t^K oznacza estymatę Kalmana, można scharakteryzować granicę ciągu błędów:

$$\lim \varepsilon_t = 0, \quad \text{przy } t \rightarrow \infty. \quad (33)$$

Można zauważyć, że spełniona jest zależność:

$$\varepsilon_t = \varepsilon_t((\Phi_t, \Theta_t, \Gamma), F).$$

7. Obliczenia numeryczne

W punkcie tym przedstawiono obliczenia numeryczne potwierdzające otrzymane rezultaty teoretyczne. Rozwiązujemy problem adaptacyjnej filtracji całkowitych wypłat dla danych kwartalnych dotyczących wypłat ubezpieczenia *autocasco* w Polsce z okresu 31 marca 1996 r.-30 czerwca 2004 r. (tab. 1). Dane otrzymano z Komisji Nadzoru Ubezpieczeń i Funduszy Emerytalnych (KNUiFE). Dane otrzymane z KNUiFE są sumaryczne dla ok. 60 towarzystw ubezpieczeniowych działających w Polsce, więc – w konsekwencji – prezentowane wyniki bardziej dotyczą sektora. Należy jednak zaznaczyć, że prezentowana metoda, przy zastosowaniu danych dotyczących każdego towarzystwa oddzielnie, może być wykorzystywana do prognozowania na poziomie towarzystwa ubezpieczeniowego.

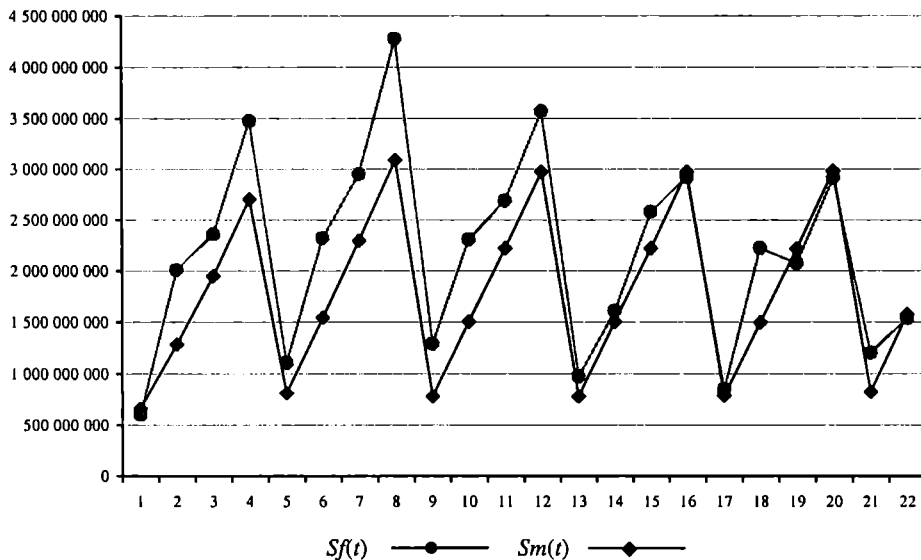
Tabela 1. Dane dotyczące wysokości całkowitych wypłat $S(t)$

Data	31 marca 1999	30 czerwca 1999	30 września 1999	31 grudnia 1999	31 marca 2000
	1	2	3	4	5
$Sm(t)$	663 712 745,5	1 285 481 197,3	1 952 157 595,8	2 704 176 286,2	812 610 696,4
$Sf(t)$	600 468 688,0	2 009 190 609,6	2 360 709 330,7	3 478 274 874,4	1 108 707 692,7
$Sm(t) - Sf(t)$	63 244 057,4	-723 709 412,3	-408 551 734,9	-774 098 588,3	-296 096 996,3
Data	30 czerwca 2000	30 września 2000	31 grudnia 2000	31 marca 2001	30 czerwca 2001
	6	7	8	9	10
$Sm(t)$	1 544 597 079,4	2 297 557 645,2	3 088 174 550,6	782 458 208,3	1 504 912 421,0
$Sf(t)$	2 322 766 757,8	2 953 737 574,9	4 278 891 647,0	1 294 462 813,5	2 308 436 790,0
$Sm(t) - Sf(t)$	-778 169 678,4	-656 179 929,7	-1190717096,4	-512 004 605,2	-803 524 369,0
Data	30 września 2001	31 grudnia 2001	31 marca 2002	30 czerwca 2002	30 września 2002
	11	12	13	14	15
$Sm(t)$	2 221 339 674,9	2 974 221 832,8	782 458 208,3	1 504 912 421,0	2 221 339 674,9
$Sf(t)$	2 689 691 128,2	3 570 910 768,6	974 746 112,9	1 608 565 560,5	2 579 214 133,4
$Sm(t) - Sf(t)$	-468 351 453,3	-596 688 935,8	-192 287 904,5	-103 653 139,5	-357 874 458,5
Data	31 grudnia 2002	31 marca 2003	30 czerwca 2003	30 września 2003	31 grudnia 2003
	16	17	18	19	20
$Sm(t)$	2 974 221 832,8	788 663 772,3	1 498 511 633,8	2 218 786 352,5	2 984 205 999,5
$Sf(t)$	2 921 801 696,3	850 442 333,1	2 221 692 458,7	2 075 773 888,9	2 912 862 247,0
$Sm(t) - Sf(t)$	52420136,5	-61 778 560,8	-723 180 824,9	143 012 463,6	71 343 752,5
Data	31 marca 2004	30 kwietnia 2004			
	21	22			
$Sm(t)$	825 581 889,7	1 576 469 068,0			
$Sf(t)$	1 205 185 454,6	1 538 701 354,7			
$Sm(t) - Sf(t)$	-379 603 564,8	37 767 713,3			

Źródło: opracowanie własne na podstawie danych KNUiFE.

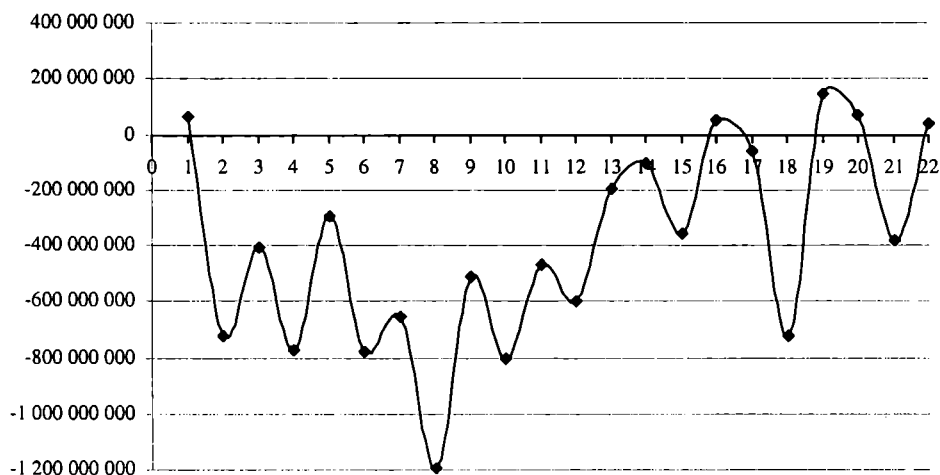
Dwanaście pierwszych obserwacji (31 marca 1996 r.-31 grudnia 1998 r.) zostało wykorzystanych do obliczeń tzw. statystyk początkowych (m , R_t i $R_2(t)$). Z tego względu prognozy rozpoczynają się od daty 31 marca 1999 r.

Na wykresach zaprezentowanych na rys. 2-3 przedstawiono zmierzone wysokości wypłat $Sm(t)$, prognozowane wysokości wypłat $Sf(t)$ oraz różnice $Sm(t) - Sf(t)$. Wartości na osiach pionowych wykresów są wyrażone w tysiącach złotych.



Rys. 1. Wysokość całkowitych wypłat – zmierzona $Sm(t)$ i prognozowana $Sf(t)$

Źródło: opracowanie własne.



Rys. 2. Różnice $Sm(t) - Sf(t)$

Źródło: opracowanie własne.

Ponadto został wyznaczony trójkąt wypłat (*run-off triangle*) $S(t)$ na podstawie znajomości procentowego rozkładu wysokości udokumentowanych wypłat kwartalnych w latach 1999-2003.

Trójkąt (*run-off triangle*) całkowitych wypłat $S(t)$

	0	1	2	3	4
1999	6 755 535 344,8	1 552 860 675,8	24 9234 983,3	176 576 649,2	9 293 507,9
2000	8 704 041 417,4	1 984 589 693,4	66 117 442,8	15 996 155,5	
2001	7 735 944 226,7	1 777 402 970,4	55 235 608,4		
2002	6 408 446 411,7	1 506 110 213,8			
2003	6 492 950 982,2				

8. Podsumowanie

W książce Goovaerts i innych (por. [8]) podano wiele przykładów, z których wynika, że stosując złożone metody ilościowe, możemy prognozować różne wielkości występujące w ubezpieczeniach. Zastosowaliśmy metodę filtru Kalmana do prognozy wysokości wypłat w kolejnych kwartałach dla szkód podlegających ubezpieczeniu *autocasco*. W wyniku zastosowania tej metody otrzymaliśmy zadowalające rezultaty, tzn. zmniejszające się w kolejnych krokach różnice $S_m(t) - S_f(t)$ pomiędzy prognozą a późniejszym pomiarem, co jest widoczne na przedstawionych wykresach. Zaproponowana metoda uśredniającej procedury adaptacyjnej pozwala przewidzieć nie tylko wysokość całkowitej wypłaty i liczbę roszczeń dla wartości skumulowanych i nieskumulowanych, ale także „koszty palące” (czyli całkowite wypłaty podzielone przez wielkość składek), wypłaty przekraczające rezerwy ubezpieczyciela czy też procent sfinansowanych wypłat odszkodowań w stosunku do zgłoszonych roszczeń, strukturę portfela lub wysokość składki.

Nasze zbiory danych nie były zbyt liczne ze względu na krótki okres funkcjonowania w Polsce gospodarki rynkowej. Na podstawie zaproponowanej metody, wykorzystując dostępne dane, można sformułować pewne wnioski dotyczące realizacji ryzyka w ubezpieczeniach. Prezentowana metoda pozwala na interesującą od strony portfelowej analizę, m.in. umożliwia identyfikację trendów, a tym samym wychwycenie sezonowości analizowanych rodzajów ryzyka w ubezpieczeniach gospodarczych.

Niniejszą pracę autorzy traktują jako początek szerszych badań nad wypracowaniem systemu prognozowania na rynku ubezpieczeniowym. Przedstawione wyniki dotyczą jedynie małego obszaru, jednak stosowane narzędzia mogą być wykorzystywane w procesie prognozowania realizacji różnego rodzaju ryzyka charakterystycznego dla ubezpieczeń gospodarczych.

Literatura

- [1] Anderson B.D, Moore J.B., *Detectability and Stabilizability of Time-Varying Discrete-Time Linear systems*, SIAM J. „Control and Optimization” 1981, 19 (1), 20-32.
- [2] Anderson B.D., Moore J.B., *Optimal Filtering*, Prentice Hall, Englewood Cliffs, New Jersey 1979.
- [3] Beneš V.E., Karatzas I., *Estimation and Control for Linear, Partially Observable Systems with Non-Gaussian Initial Distributions*, „Stochastic Processes and their Applications” 1983, 14, 233-248.
- [4] Beneviste A., Metivier M., Priouret P., *Adaptive Algorithms and Stochastic Approximations*. Springer, Berlin 1990.
- [5] Caines P.E., *Linear Stochastic Systems*, John Wiley and Sons, New York 1988.
- [6] De Jong P., Zehnwirth B., *Credibility Theory and the Kalman Filter*, „Insurance: Mathematics and Economics” 1983, 2, 281-286.
- [7] De Santis A., Germani A., Raimondi M., *Optimal Quadratic Filtering of Linear Discrete-Time Non-Gaussian Systems*, „IEEE Transactions on Automatic Control” 1995, 40 (7), 1274-1278.
- [8] Goovaerts M.J. i in., *Effective Actuarial Methods*, North-Holland, Amsterdam 1990.
- [9] Haussmann U.G., *A Conditionally Almost Linear Filtering Problem with Non-Gaussian Initial Conditions*, „Stochastics” 1998, 23, 244-275.
- [10] Jazwinski A.H., *Stochastic Processes and Filtering Theory*, Academic Press, New York 1970.
- [11] Johnson N.L., Kotz S., Balakrishnan N., *Continuous Univariate Distributions*, vol. I, John Wiley and Sons, New York 1994.
- [12] Laitonis D.G., Giannakopoulos P.K., *Adaptive Distributed Partitioning Filters: Non-Gaussian Initial Conditions*, „Stochastic Analysis and Applications” 1999, 17 (3), 405-419.
- [13] Malcolm W.P., Jansen M.R., Elliot R.J., *Risk Sensitive Filtering with Poisson Process Observations*, „Applied Mathematics and Optimization” 2000, 41, 387-402.
- [14] Makowski A.M., *Filtering Formulae for Partially Observed Linear Systems with Non-Gaussian Initial Conditions*, „Stochastics” 1986, vol. 16, 1-24.
- [15] Michalski T., Twardowska K., Tylutki B., *The Convergence Analysis in Adaptive Filtering in the Presence of Heavy-Tailed Distributions*, to appear in „Applicationes Mathematicae” 2005.
- [16] Panjer H.H., Willmot G.E., *Insurance Risk Models*, Society of Actuaries, Schaumburg 1992.
- [17] Polyak B.T., *New Method of Stochastic Approximation Type*, „Automation and Remote Control” 1990, 51, 937-946.
- [18] Sowers R., Makowski A.M., *Discrete-Time Filtering For Linear Systems With Non-Gaussian Initial Conditions: Asymptotic Behaviour of the Difference between the MMSE Estimates*, „IEEE Transactions on Automatic Control” 1992, vol. 37, nr 1, 114-120.
- [19] Stewart G.W., *Introduction to Matrix Computations*, Academic Press, New York 1973.
- [20] Yin G.G., Zhu Y., *Averaging Procedures in Adaptive Filtering: An Efficient Approach*, „IEEE Transactions on Automatic Control” 1992, 37 (4), 466-475.
- [21] Zehnwirth B., *A Generalization of the Kalman Filter for Models with State-Dependent Observation Variance*, „Journal of the American Statistical Association” 1988, 83(401), 164-167.

ADAPTIVE FILTERING ANALYSIS IN INSURANCE

Summary

We apply the so-called Kalman filter technique to predict the number of claims for future periods and the total claim payments in the collective risk model in insurance problems. The methods, based

on adaptive linear discrete filtration and its averaging procedure, were examined for the probability distributions appearing in insurance.

We have used the method to predict the loss reserving in automobile insurance for the next successive quarters. We get a satisfactory result.

The method proposed here allows us to predict and control not only the claim amounts and numbers of claims using the absolute or cumulative values but also the loss ratios, the burning costs, the level of the reserves for unpaid compensations, the level of the catastrophic reserves or premiums.