

Zbigniew Tarapata*

Symulacyjna metoda doboru optymalnych parametrów w prognostycznych modelach wygładzania wykładniczego

Wydział Cybernetyki Wojskowej Akademii Technicznej w Warszawie

Streszczenie

W artykule zaprezentowano symulacyjną metodę doboru optymalnego zestawu parametrów w modelach wygładzania wykładniczego Browna. Rozpatrzono trzy podstawowe modele Browna: prosty, klasyczny oraz zmodyfikowaną wersję klasycznego (zapropionowaną przez Z. Pawłowskiego w [9]). Za podstawę w metodzie symulacyjnej posłużyły trzy algorytmy stochastycznego poszukiwania minimum funkcji (minimalizacji błędu standardowego prognozy): najprostszy algorytm z losowaniem punktów próbnych, algorytm z wyznaczaniem kierunków poprawy oraz algorytm adaptacyjny. Dla pierwszego z algorytmów podano oszacowanie liczby powtórzeń eksperymentu symulacyjnego, przy której uzyskamy wartości parametrów modelu z żadaną dokładnością, przy zadanym statystycznym poziomie ufności. Przedstawiono porównanie szybkości zbieżności algorytmów do rozwiązania optymalnego oraz dokładność oszacowań otrzymywanych z badanych algorytmów w zależności od rodzaju modelu wygładzania, na podstawie danych historycznych dotyczących wartości Warszawskiego Indeksu Giełdowego (WIG).

Wstęp

Jednym z podstawowych problemów dotyczących budowy prognostycznych modeli wygładzania wykładniczego jest dobór odpowiednich parametrów w tych modelach, tak aby minimalizować błąd prognozy (np. standardowy błąd prognozy, średni kwadratowy błąd prognozy *ex post*). W literaturze przedmiotu (np. [1], [3], [7], [11]) mówi się, że należy je dobrać doświadczalnie, np. poprzez przeprowadzenie serii eksperymentów komputerowych polegających na stosowaniu różnych kombinacji wartości parametrów $\alpha \in [0,1]$, $\beta \in [0,1]$ (dla modelu Holta) lub $\alpha \in [0,1]$ (dla modelu Browna) przy ustalonym kroku Δ zmiany wartości tych parametrów. Dla przykładu, w pakiecie programowym STATISTICA [10], w którym występuje moduł prognozowania na podstawie szeregów czasowych, znajduje się opcja tzw. przeszukiwania sieciowego parametrów, polegająca na tym, że program powiększa wartość każdego parametru od wartości minimalnej do wartości maksymalnej co ustalony przyrost Δ jego wartości i dla każdej kombinacji wartości parametrów obliczana jest suma kwadratów różnic (6) między wartościami zaobserwowanymi

* Autor pracuje również jako wykładowca w Wyższej Szkole Ekonomicznej w Warszawie.

i prognozowanymi (wygładzanymi). Na tej podstawie wybierana jest najlepsza kombinacja wartości parametrów minimalizująca sumę kwadratów różnic. Należy sobie jednak zdać sprawę, że przestrzeń poszukiwań wartości parametrów przy takim podejściu może być bardzo duża. Gdybyśmy np. w modelu Holta próbowali dobrać najlepszą parę współczynników α , β (ze względu na minimalizację standardowego błędu prognozy) przy przyjętym kroku Δ zmiany wartości tych współczynników równym 0.001, to należałoby sprawdzić 1000^2 par wartości tych współczynników! Z kolei dla $\Delta=0.01$ należałoby wykonać „tylko” 100^2 sprawdzeń. Powracając do pakietu STATISTICA, dodatkowym ograniczeniem jest fakt, że maksymalna liczba kombinacji wartości parametrów, którą można tam sprawdzić jest ograniczona i wynosi 8000.

Zadanie doboru optymalnych parametrów minimalizujących wartość funkcji błędu prognozy można również sprowadzić do zadania optymalizacji nieliniowej lub (dla szczególnych postaci funkcji błędu prognozy) do zadania optymalizacji liniowej, co zostanie pokazane w dalszej części artykułu (rozwiązanie takiego zadania może być znalezione np. poprzez użycie dodatku Solver w arkuszu MsExcel). Obie metody wiążą się jednak z pewnymi niedogodnościami związanymi albo z ich efektywnością (metody optymalizacji nieliniowej i liniowej), albo z ich zbieżnością (lub raczej brakiem zbieżności) do minimum rozpatrywanej funkcji (metody optymalizacji nieliniowej). Z tego też powodu, jak również w przypadkach, kiedy przestrzeń poszukiwań parametrów jest duża przydatne mogą okazać się metody stochastycznego poszukiwania wartości parametrów modelu minimalizujące funkcję błędu prognozy.

W niniejszym artykule przedstawiona zostanie idea metod stochastycznego poszukiwania najlepszych parametrów modeli wykładniczego oraz dokonana zostanie analiza ich zastosowania dla trzech postaci modelu Browna. W [11] przebadano te metody w modelach Holta i Wintersa oraz przeprowadzono porównanie możliwości ich wykorzystania do prognozowania wartości Warszawskiego Indeksu Giełdowego.

Modele wykładniczego Browna

Prosty model wykładniczego znany jest w literaturze dotyczącej prognozowania na podstawie szeregów czasowych od 1959 roku. Autor modelu - R.G. Brown - podczas drugiej wojny światowej pracował dla Marynarki Wojennej USA, gdzie był przydzielony do opracowania systemu śledzącego cel, wykorzystywanego do lokalizacji okrętów podwodnych dla potrzeb sterowania ogniem. Później zastosował tę technikę do prognozowania popytu na części zapasowe. Opisał te pomysły w swojej książce [1] na temat sterowania zapasami.

Prosty model Browna może być stosowany w przypadku występowania w szeregu czasowym prawie stałego poziomu zmiennej prognozowanej oraz wahań przypadkowych.

Model ten może być opisany za pomocą następującego wzoru rekurencyjnego [3]:

$$y_t^*(\alpha) = F_{t-1} = \alpha \cdot y_{t-1} + (1 - \alpha) \cdot y_{t-1}^*(\alpha) \quad (1)$$

lub równoważnie

$$y_t^*(\alpha) = y_{t-1}^*(\alpha) + \alpha \cdot q_{t-1} \quad (2)$$

gdzie:

$$\begin{aligned} y_t^*(\alpha) & - \text{wartość prognozy zmiennej } y \text{ na chwilę } t; \\ y_{t-1} & - \text{wartość zmiennej } y \text{ w szeregu czasowym w chwili } t-1; \\ q_{t-1} & - \text{błąd } ex \text{ post prognozy wyznaczonej na chwilę } t-1, \\ & q_{t-1} = y_{t-1} - y_{t-1}^*(\alpha) \quad (3) \\ \alpha & - \text{parametr modelu, tzw. stała wygładzania, } \alpha \in [0, 1]. \end{aligned}$$

Z (1) wynika, że wartość prognozy na chwilę t zależy, w sposób rekurencyjny, od wartości z szeregu czasowego i prognoz na chwile $t-1, t-2, \dots, 1$. Jako wartość prognozy y_1^* niezbędną do konstrukcji modelu przyjmuje się najczęściej wartość początkową zmiennej prognozowanej w szeregu czasowym, tzn. y_1 lub średnią arytmetyczną pierwszych kilku (np. 5) wartości zmiennej y z szeregu czasowego.

Wartość współczynnika α wpływa na stopień wygładzenia szeregu czasowego (stąd nazwa : model wygładzania wykładniczego): jeżeli $\alpha \approx 1$, to budowana prognoza będzie uwzględniała w wysokim stopniu błędy *ex post* prognoz poprzednich; w przeciwnym przypadku ($\alpha \approx 0$) budowana prognoza będzie uwzględniała w niewielkim stopniu te błędy. W literaturze (np. [3], [5]) można spotkać sugestie, że α powinno należeć do przedziału (0.2; 0.3). Jednak w studiach prowadzonych np. przez Makridakisa w [8] najlepsze prognozy przynosiły często wartości $\alpha > 0.3$.

Ponieważ wielkość współczynnika α ma wpływ na jakość modelu prognostycznego i wielkość błędów prognoz (3) na każdą chwilę - a nie można arbitralnie wskazać najlepszej wartości tego współczynnika dla każdego danych - wobec tego podstawowy problem, który dotyczy prostego modelu Browna można zdefiniować następująco:

znaleźć takie α^* , dla którego zachodzi

$$s(\alpha^*) = \min_{\alpha \in [0,1]} s(\alpha) \quad (4)$$

gdzie:

$$\begin{aligned} s(\alpha) & - \text{standardowy błąd prognozy,} \\ s(\alpha) & = \sqrt{\frac{1}{n} \cdot \sum_{t=1}^n (y_t - y_t^*(\alpha))^2} \quad (5) \end{aligned}$$

n - liczba danych w szeregu czasowym.

Zadanie (4) można traktować jako zadanie programowania kwadratowego (pomijając symbol pierwiastka) i rozwiązać jedną ze znanych metod zaprezentowanych np. w [12]. Dla przykładu, we wspomnianym wcześniej pakiecie STATISTICA [10] znajduje się opcja wyznaczania optymalnych wartości parametrów modeli wygładzania wykładniczego polegająca na minimalizacji funkcji sumy kwadratów różnic opisanej następująco:

$$s(\alpha) = \sum_{t=1}^n (y_t - y_t^*(\alpha))^2 \quad (6)$$

przy użyciu metody quasi-Newtona [12].

Zauważmy, że funkcje (5) i (6) osiągają minimum dla tej samej wartości parametru α i w sumie nie ma znaczenia (z punktu widzenia rozwiązania α^*), która z nich będzie minimalizowana (preferowana jest funkcja (6) z racji prostszej postaci).

Czasami stosowana jest funkcja:

$$s(\alpha) = \frac{1}{n} \cdot \sum_{t=1}^n |y_t^*(\alpha) - y_t| \quad (7)$$

która opisuje średni bezwzględny błąd prognozy. Jej postać jest o tyle istotna, że minimalizacja funkcji (7) jest minimalizacją sumy odchyleń bezwzględnych, a ten problem sprowadzalny jest w prosty sposób do łatwiejszego obliczeniowo problemu programowania liniowego [6]. W celu sprowadzenia zadania (4), w którym funkcję celu opisaną przez (5) zastępujemy funkcją opisaną przez (7), do równoważnego zadania programowania liniowego przeprowadźmy rozumowanie przedstawione w [6] (rozdz. 11.3.1). W tym celu oznaczmy:

$$\bar{z}_t = \max \{0, y_t^*(\alpha) - y_t\}, \quad t = \overline{1, n} \quad (8)$$

$$\underline{z}_t = \max \{0, y_t - y_t^*(\alpha)\}, \quad t = \overline{1, n} \quad (9)$$

Wówczas możemy zapisać, że dla każdego $t = \overline{1, n}$ zachodzi:

$$|y_t^*(\alpha) - y_t| = \bar{z}_t + \underline{z}_t \quad (10)$$

$$y_t^*(\alpha) - y_t = \bar{z}_t - \underline{z}_t \quad (11)$$

$$\bar{z}_t \geq 0, \quad \underline{z}_t \geq 0, \quad \bar{z}_t \cdot \underline{z}_t = 0 \quad (12)$$

Po uwzględnieniu powyższych zapisów otrzymujemy następujący problem programowania liniowego:

$$\frac{1}{n} \cdot \sum_{t=1}^n \bar{z}_t + \underline{z}_t \rightarrow \min \quad (13)$$

przy ogr.

$$y_t^*(\alpha) - \bar{z}_t + \underline{z}_t = y_t, \quad t = \overline{1, n} \quad (14)$$

$$1 \geq \alpha \geq 0 \quad (15)$$

$$\bar{z}_t \geq 0, \quad \underline{z}_t \geq 0, \quad t = \overline{1, n} \quad (16)$$

Należy jeszcze tylko zauważyć, że w problemie (13)÷(16) pominęliśmy warunki $\bar{z}_t \cdot \underline{z}_t = 0, t = \overline{1, n}$, ale jak pokazano w [6] nie rozszerza to zbioru rozwiązań optymalnych. Zadanie (13) można rozwiązać wykorzystując standardowe metody rozwiązywania zadań programowania liniowego (np. algorytm simpleksowy [6]). Jednakże problem opisany przez (13)÷(16) może być dużych rozmiarów (liczba zmiennych wynosi $2n+1$, liczba ograniczeń – $3n+2$) i mogą być problemy z efektywnością jego rozwiązywania.

Innym modelem z grupy modeli Browna jest tzw. *klasyczny model Browna*. Został on zaproponowany w [2] również przez F.G. Browna. W modelu tym zakłada się, że szereg czasowy y_t jest sumą składnika systematycznego m_t i składnika losowego ζ_t . Składnik systematyczny jest opisywany wielomianem stopnia p . Szereg czasowy y_t przedstawić wobec tego można w następującej postaci:

$$y_t = m_t + \zeta_t = a_{0,t} + a_{1,t} \cdot t + \frac{1}{2!} \cdot a_{2,t} \cdot t^2 + \dots + \frac{1}{p!} \cdot a_{p,t} \cdot t^p + \zeta_t \quad (17)$$

Istota modelu polega na tym, że wykorzystuje się w nim tzw. operatory wygładzania S_t^k rzędu k , $k = \overline{0, n}$, które zależą rekurencyjnie od wartości szeregu czasowego w chwilach, $t-1, t-2, \dots, 1$ w sposób następujący:

$$\begin{aligned} S_t^0 &= y_t \\ S_t^k(\alpha) &= \alpha \cdot S_t^{k-1} + (1 - \alpha) \cdot S_{t-1}^k, \quad k = \overline{1, n} \end{aligned} \quad (18)$$

Na podstawie układu równań (18) wyznacza się układ równań wiążących oceny współczynników $a_{i,t}$, $i = \overline{1, p}$, $t = \overline{1, n}$ z równania (17) z wartościami operatorów S_t^k , $k = \overline{0, n}$. Dla przypadku, kiedy $p=1$ (tzn. składnik systematyczny ze wzoru (17) opisany będzie wielomianem pierwszego stopnia) pokazuje się, że zależność ta jest następująca [2]:

$$\begin{aligned} a_{0,t}(\alpha) &= 2 \cdot S_t^1(\alpha) - S_t^2(\alpha) \\ a_{1,t}(\alpha) &= \frac{\alpha}{1 - \alpha} \cdot (S_t^1(\alpha) - S_t^2(\alpha)) \end{aligned} \quad (19)$$

Jako wartości S_1^1 oraz S_1^2 przyjmuje się najczęściej początkową wartość z szeregu, tzn. $S_1^1 = S_1^2 = y_1$.

Prognozę y_t^* na chwilę $t > T$ buduje się w sposób następujący:

$$y_t^*(\alpha) = a_{0,T}(\alpha) + a_{1,T}(\alpha) \cdot (t - T), \quad t > T \quad (20)$$

gdzie $1 \leq T \leq n$ oznacza liczbę chwil z szeregu czasowego, które bierzemy pod uwagę do budowy prognozy y_t^* . Najczęściej jest tak, że $T=t-1$. Mówimy wtedy o prognozach budowanych z jednookresowym wyprzedzeniem.

Problem (4), dla klasycznego modelu Browna, przybiera podobną postać jak dla modelu prostego, z tą różnicą, że zamiast funkcji $y_t^*(\alpha)$ ze wzoru (1), występującej w (5), należy przyjąć funkcję (20).

Zmodyfikowana wersja klasycznego modelu Browna została zaproponowana przez Z. Pawłowskiego w [9], s. 241÷245. Modyfikacja ta jest prostsza pod względem rachunkowym a jednocześnie pozwala, jak pisze autor, bardziej efektywnie wnioskować w przyszłość. Ma ona postać:

$$y_t^*(\alpha) = y_T^*(\alpha) + \Delta y_T^*(\alpha) \cdot (t - T), \quad t > T, \quad 2 \leq T \leq n \quad (21)$$

gdzie:

$$\Delta y_T^*(\alpha) = y_T^*(\alpha) - y_{T-1}^*(\alpha) \quad (22)$$

$y_T^*(\alpha)$, $y_{T-1}^*(\alpha)$ - prognozy wyznaczone z (1) odpowiednio dla $t=T$ oraz $t=T-1$;

Z (21) widać, że prognoza na chwilę $t > T$ budowana jest jako suma prognoz wyznaczanych z prostego modelu Browna: prognozy na ostatnią chwilę (z branych pod uwagę) w szeregu czasowym oraz różnicy prognoz z ostatniej i przedostatniej

chwili w szeregu czasowym przemnożonej przez liczbę okresów odległych od chwili T . Dla prognoz budowanych z jednookresowym wyprzedzeniem (tzn. gdy $t=T+1$) formuła (21) przybiera postać :

$$y_{t=T+1}^*(\alpha) = y_T^*(\alpha) + \Delta y_T^*(\alpha) \quad (23)$$

Dla zmodyfikowanego modelu Browna, problem (4) przybiera podobną postać jak dla modelu prostego, z tą różnicą, że zamiast funkcji $y_t^*(\alpha)$ ze wzoru (1), występującej w (5), należy przyjąć funkcję (21).

Idea stochastycznych algorytmów doboru najlepszych parametrów modeli wygładzania wykładniczego

Rozważając algorytmy stochastyczne [13] bierze się zawsze pod uwagę pewną abstrakcyjną przestrzeń probabilistyczną (Ω, G, P) i zakłada się, że wszystkie rozważane wielkości losowe będą w niej zdefiniowane. Oznaczmy przez $S \subset E^k$ zbiór rozwiązań dopuszczalnych rozwiązywanego zadania (np. dla zadania (4) $S=[0,1] \subset E^1$, dla modelu Holta będzie $S=[0,1] \times [0,1] \subset E^2$). Zbiór S rozszerza się do przestrzeni miarowej $(S, \mathcal{A}, \lambda)$, wyróżniając w nim pewne σ -ciało podzbiorów \mathcal{A} oraz pewną miarę λ (ponieważ $S \subset E^k$, więc λ można interpretować jako długość odcinka w E^1 oraz pole powierzchni w E^2). Ponadto będziemy rozpatrywać pewne podzbiory $T \subset S$, które także rozszerzamy do przestrzeni mierzalnych (T, \mathcal{B}, μ) . Przez punkt losowy w T rozumiemy każde mierzalne odwzorowanie Ω w T . Każde takie odwzorowanie generować będzie rozkład prawdopodobieństwa w przestrzeni (T, \mathcal{B}) , który oznaczymy przez Q . Będziemy mówili, że punkt losowy X ma rozkład jednostajny na (T, \mathcal{B}, μ) (krótko: na T), jeżeli $0 < \mu(T) < \infty$ oraz $Q(B) = \mu(B) / \mu(T)$. Będziemy to zapisywać : $X \sim U(T)$. Jeżeli przez f oznaczymy funkcję kryterium (np. (5) dla prostego modelu Browna), to zakłada się o tej funkcji, że jest \mathcal{A} -mierzalna, tzn., że dla każdego punktu losowego X w zbiorze S wielkość $f(X)$ jest zwykłą zmienną losową. Ponadto, jeżeli istnieje rozwiązanie $x^* \in S$ problemu $\min_{x \in S} f(x)$, to liczba $f^* = f(x^*)$ ma następującą własność:

$$P\{\omega \in \Omega: f(X(\omega)) < f^*\} = 0 \quad (24)$$

oraz

$$P\{\omega \in \Omega: f(X(\omega)) < f^* + \varepsilon\} > 0 \quad \text{dla każdego } \varepsilon > 0, \quad (25)$$

gdzie X jest punktem losowym w S .

Przedstawimy jeszcze podstawową definicję, która jest stosowana do określenia i badania zbieżności algorytmów realizujących metody poszukiwania losowego ekstremum funkcji.

Definicja 1 [13]

Jeżeli ciąg zmiennych losowych $(f(X_n))_{n \geq 0}$ dąży z prawdopodobieństwem 1 do f^* , to mówimy, że ciąg punktów losowych $(X_n)_{n \geq 0}$ dąży z prawdopodobieństwem 1 do minimum globalnego funkcji f w przestrzeni $(S, \mathcal{A}, \lambda)$.

Poniżej opiszemy trzy algorytmy stochastyczne zastosowane do badań.

Algorytm 1 (z losowaniem punktów próbnych)

Algorytm ten zalicza się do grupy algorytmów zwanych algorytmami z losowaniem punktów próbnych [13].

Ogólny schemat tego algorytmu przedstawia się następująco:

$$x^{n+1} = \begin{cases} \xi^n, & \text{gdy } f(\xi^n) < f(x^n) \text{ ("sukces")} \\ x^n, & \text{gdy } f(\xi^n) \geq f(x^n) \end{cases} \quad (26)$$

gdzie:

- x^{n+1} - kolejne przybliżenie (w $n+1$ -szym losowaniu) rozwiązania $x \in S, x = \alpha$;
- ξ^n - punkt losowy w zbiorze $S, S=[0,1]$ oraz $\xi^n \sim U(S)$;
- $f(\cdot)$ - minimalizowana wartość funkcji celu opisana przez (5), tzn.

$$f(\cdot) = s(\cdot) \quad (27)$$

w której

$$y_i^*(\alpha) = \begin{cases} (1), & \text{dla prostego modelu Browna} \\ (20), & \text{dla klasycznego modelu Browna} \\ (21), & \text{dla zmodyfikowanego modelu Browna} \end{cases}$$

Idea algorytmu polega na tym, że w każdej $n+1$ -szej iteracji losujemy z rozkładem równomiernym ze zbioru S kombinację parametrów odpowiedniego modelu (tzw. punkt próbny ξ^n ; stąd nazwa algorytmu). Jeżeli wartość funkcji celu (określona przez (27)) dla nowo wylosowanych parametrów jest mniejsza niż dla parametrów z poprzedniego (tzn. n -tego) kroku, to nowym przybliżeniem rozwiązania problemu (4) jest zestaw nowo wylosowanych parametrów, czyli punkt próbny ξ^n . W przeciwnym przypadku pozostajemy przy rozwiązaniu x^n znalezionym do n -tego kroku. Ciąg $(x^n)_{n \geq 0}$ możemy więc traktować jako ciąg coraz lepszych (nie gorszych) przybliżeń minimum funkcji (27).

Mimo, że jest to najprostszy z możliwych algorytmów, to jednak, jak pokażemy w następnych rozdziałach, jest bardzo skuteczny.

Algorytm 2 (z losowaniem kierunku poprawy)

Algorytm ten zalicza się do grupy algorytmów zwanych algorytmami z losowaniem kierunku poprawy [13].

Ogólny schemat tego algorytmu przedstawia się następująco:

$$x^{n+1} = \begin{cases} x^n + a_n \cdot \xi^n, & \text{gdy } f(x^n + a_n \cdot \xi^n) < f(x^n) \\ x^n, & \text{w przeciwnym przypadku} \end{cases} \quad (28)$$

gdzie:

- x^{n+1} - kolejne przybliżenie (w $n+1$ -szym losowaniu) rozwiązania $x \in S, x = \alpha$;
- ξ^n - punkt losowy (kierunek poprawy) w zbiorze Z ,

$$Z = \{-e_1, e_1\} \quad (29)$$

gdzie wielkość e_1 oznacza wektor osi układu współrzędnych, $\xi^n \sim U(Z)$;

a_n - długość „kroku”, zmienna losowa o rozkładzie równomiernym losowanym ze zbioru T ,

$$T = \begin{cases} [0, 1 - x^n], & \text{gdy } \xi^n = e_1 \\ [0, x^n], & \text{gdy } \xi^n = -e_1 \end{cases} \quad (30)$$

oraz $a_n \sim U(T)$;

$f(\cdot)$ - funkcja celu określona przez (27);

Idea tego algorytmu polega na tym, że w każdej $n+1$ -szej iteracji losujemy z rozkładem równomiernym ze zbioru Z (określonego przez (29)) kierunek ξ^n poszukiwania minimum funkcji, tzw. kierunek poprawy. Kierunek ten można interpretować w ten sposób, że określa on „numer” parametru funkcji celu oraz to, czy wartość tego parametru będzie zmniejszana, czy zwiększana. Np. dla modelu Browna, jeżeli wylosujemy $-e_1$, to oznacza to, że będziemy zmniejszać (bo „-”) wartość pierwszego parametru (bo „numer”=1, czyli α). Następnie losujemy długość „kroku”, czyli wartość zmiany odpowiedniego parametru funkcji celu (w zależności od tego, który kierunek (czyli parametr) wylosowaliśmy) zgodnie z (30). Zauważmy, że wartość długości „kroku” jest losowana w ten sposób (por. (30)), aby punkt $x^n + a_n \cdot \xi^n$ mieścił się w zbiorze S . Jeżeli wartość funkcji celu (określona przez (27)) dla argumentu $x^n + a_n \cdot \xi^n$ jest mniejsza niż dla argumentu x^n z poprzednich kroków, to nowym przybliżeniem rozwiązania problemu (4) jest $x^n + a_n \cdot \xi^n$. W przeciwnym przypadku pozostajemy przy rozwiązaniu x^n znalezionym do n -tego kroku. Podobnie, jak to było w *algorytmie 1* ciąg $(x^n)_{n \geq 0}$ możemy traktować jako ciąg coraz lepszych (nie gorszych) przybliżeń minimum funkcji (27).

Algorytm 3 (z adaptacją rozkładu prawdopodobieństwa kierunku poprawy)

Idea tego algorytmu jest trochę bardziej skomplikowana niż dwóch wcześniej przedstawionych. Istnieje tutaj pewna analogia do *algorytmu 2*. Różnica polega na tym, że o ile w *algorytmie 2* losowaliśmy każdy z kierunków poprawy z jednakowym prawdopodobieństwem w każdym losowaniu, to w *algorytmie 3* rozkład prawdopodobieństwa kierunku poprawy jest zmieniany w trakcie obliczeń, w taki sposób, aby wykorzystać zebrane już informacje o minimalizowanej funkcji. Ponadto inna jest filozofia wyznaczania długości „kroku” w każdym losowaniu. Przedstawiony poniżej algorytm jest zmodyfikowaną wersją algorytmu 5A zaprezentowanego w [13].

Weźmy pod uwagę *algorytm 2* i niech

$$Z = \{e'_1, e'_2, \dots, e'_{2 \cdot K}\} \quad (31)$$

będzie zbiorem rozważanych tam kierunków, przy czym $e'_j = e_j$ dla $j = \overline{1, K}$ oraz $e'_j = -e_{j-K}$ dla $j = \overline{K+1, 2 \cdot K}$, gdzie K oznacza liczbę parametrów modelu (tzn. $K=1$ dla modelu Browna, $K=2$ dla modelu Holta, $K=3$ dla modelu Wintersa, (por.

[11])). Niech $(p_j^{(n)})_{1 \leq j \leq 2 \cdot K}$ będzie rozkładem prawdopodobieństwa, według którego w n -tym kroku iteracyjnym losujemy jeden z kierunków e'_j , $j = \overline{1, 2 \cdot K}$. Niech

$$p_j^{(0)} = \frac{1}{2 \cdot K}, \quad j = \overline{1, 2 \cdot K} \quad (32)$$

i rozważmy następujący sposób modyfikacji rozkładu $(p_j^{(n)})_{1 \leq j \leq 2 \cdot K}$: jeżeli na n -tym etapie obliczeń wylosowaliśmy kierunek $e'_{j(n)}$, to przyjmujemy:

$$p_{j(n)}^{(n+1)} = p_{j(n)}^{(n)} - \theta \cdot (1 - p_{j(n)}^{(n)}) \cdot \operatorname{sgn} [f(x^n + a_n \cdot e'_{j(n)}) - f(x^n)] \quad (33)$$

oraz dla $i = \overline{1, 2 \cdot K}$, $i \neq j(n)$:

$$p_i^{(n+1)} = p_i^{(n)} + \frac{\theta}{2 \cdot K - 1} \cdot (1 - p_{j(n)}^{(n)}) \cdot \operatorname{sgn} [f(x^n + a_n \cdot e'_{j(n)}) - f(x^n)] \quad (34)$$

gdzie $\theta \in [0, 1]$ jest parametrem algorytmu (dla $\theta = 0$ mamy nieadaptacyjny algorytm 2). Ze wzorów (33) i (34) wynika, że w przypadku sukcesu w n -tym kroku (tzn. gdy wartość funkcji celu w n -tym kroku będzie mniejsza niż w kroku poprzednim) zwiększamy prawdopodobieństwo wylosowania tego samego kierunku w następnym kroku i zmniejszamy to prawdopodobieństwo w przypadku porażki.

Analogiczna idea przyświeca wyznaczaniu długości „kroku” a_n w kolejnych iteracjach: a_n losujemy z rozkładem równomiernym ze zbioru:

$$T(n) = \begin{cases} [0, \delta \cdot (1 - x^n)], & \text{gdy } j(n) \in \{1, \dots, K\} \\ [0, \delta \cdot x^n], & \text{gdy } j(n) \in \{K + 1, \dots, 2 \cdot K\} \end{cases} \quad (35)$$

przy czym, jeżeli na n -tym i $n-1$ -szym etapach obliczeń wylosowaliśmy ten sam kierunek $\xi^{n-1} = e'_{j(n-1)} = \xi^n = e'_{j(n)}$, to $\delta \in (0, 1)$; w przeciwnym przypadku $\delta = 1$.

Zauważmy, że jeżeli w poprzednim etapie wylosowaliśmy ten sam kierunek, co w etapie obecnym, to zmniejszamy długość kroku, chcąc w ten sposób jak najlepiej wykorzystać fakt, że poruszamy się w dobrym kierunku i jak najwięcej punktów w tym kierunku poprawy sprawdzić; w przeciwnym przypadku losujemy długość kroku tak, jak w algorytmie 2 (por. (30)).

Możemy obecnie podać ogólny schemat algorytmu 3 :

$$x^{n+1} = \begin{cases} x^n + a_n \cdot \xi^n, & \text{gdy } f(x^n + a_n \cdot \xi^n) < f(x^n) \\ x^n, & \text{w przeciwnym przypadku} \end{cases} \quad (36)$$

gdzie:

- x^{n+1} - kolejne przybliżenie (w $n+1$ -szym losowaniu) rozwiązania $x \in S$, $x = \alpha$;
- ξ^n - punkt losowy (kierunek poprawy) w zbiorze Z określonym przez (31), losowany z rozkładem prawdopodobieństwa określonym przez (33)÷(34);
- a_n - wartość długości „kroku” będąca zmienną losową o rozkładzie równomiernym na zbiorze $T(n)$ określonym przez (35), czyli $a_n \sim U(T(n))$;
- $f(\cdot)$ - funkcja określona przez (27);

Analiza dokładności i zbieżności prezentowanych algorytmów w modelach wygładzania wykładniczego

Podanie warunków na dokładność otrzymanego rozwiązania, jak również warunków zbieżności dla wszystkich rozpatrywanych w poprzednim punkcie algorytmów nie jest trywialne. Stosunkowo najprostsza sytuacja zachodzi dla *algorytmu 1*, dlatego też dla niego podamy analityczne warunki na dokładność otrzymanego rozwiązania oraz zbieżność algorytmu. Dla pozostałych algorytmów przedstawimy empiryczne oszacowania zbieżności i dokładności.

Rozważmy *algorytm 1* opisany przez ciąg formuł (26)-(27) w poprzedniej części artykułu. Zgodnie z założeniami tam poczynionymi ξ^n , $n=1, 2, \dots$ są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym w przestrzeni $(S, \mathcal{A}, \lambda)$. Dla każdego podzbioru $A \subset S$, $A \in \mathcal{A}$, mamy:

$$P\{\xi^n \in A\} = \lambda(A) / \lambda(S) \quad (37)$$

Przypuśćmy, że za pomocą *algorytmu 1* wykonaliśmy N kroków iteracyjnych i że jako przybliżenie optymalnego rozwiązania $x^* \in S$ otrzymaliśmy X^N . Chcemy odpowiedzieć na pytanie: z jaką dokładnością oszacowanie X^N przybliży nieznaną wartość x^* minimum globalnego funkcji f (opisanej przez (27)) na zbiorze S ? Aby odpowiedzieć na to pytanie przeprowadzimy rozumowanie przedstawione w [13].

Rozważmy ciąg wartości funkcji $f_i = f(\xi^i)_{i=1, \dots, N}$, w losowych punktach ξ^i , $i = \overline{1, N}$.

Jest to ciąg niezależnych zmiennych losowych o jednakowym rozkładzie prawdopodobieństwa. Oznaczmy przez F dystrybuantę tego rozkładu:

$$F(y) = P(f_i < y) = \frac{\lambda\{x \in S : f(x) < y\}}{\lambda(S)} \quad (38)$$

Ponieważ wartości funkcji nie są znane (gdyż dotyczą losowych argumentów), to dystrybuanta F również nie jest znana.

Dla rozważanego *algorytmu 1*, punkt X^N jest najlepszym punktem spośród punktów $X^1 = \xi^1, X^2, \dots, X^N$ w tym sensie, że

$$f(X^N) = \min \{f_1, f_2, \dots, f_N\} \quad (39)$$

Wynika stąd, że :

$$P\{f(X^N) \geq y\} = [1 - F(y)]^N \quad (40)$$

Fakt ten umożliwia ocenę dokładności rozwiązania przybliżonego X^N przez oszacowanie „wielkości” zbioru punktów $x \in S$, w których funkcja f przyjmuje wartości mniejsze (czyli lepsze) od $f(X^N)$, przy czym przez wielkość tego zbioru rozumiemy jego względną miarę:

$$\Lambda_N = \frac{\lambda\{x : f(x) < f(X^N)\}}{\lambda(S)} \quad (41)$$

Z (41) wynika, że dla jednowymiarowego przypadku (np. dla $S=[0,1]$) wielkość ta jest równa stosunkowi sumy długości odcinków wewnątrz których wartość funkcji

f jest mniejsza od wartości $f(X^N)$ do długości odcinka $[0,1]$; dla dwuwymiarowego przypadku (np. dla $S=[0,1] \times [0,1]$) wielkość ta jest równa stosunkowi sumy pól figur wpisanych w kwadrat o boku długości 1, takich, że dla każdego punktu x' należącego do tej sumy wartość funkcji f jest mniejsza od wartości $f(X^N)$ do pola kwadratu o boku długości 1.

Ponieważ X^N jest wielkością losową, więc Λ_N jest zmienną losową. Dla dystrybuanty tej zmiennej losowej prawdziwe jest następujące twierdzenie.

Twierdzenie 1 [13]

Jeżeli $\xi^1, \xi^2, \dots, \xi^N$ jest ciągiem niezależnych punktów losowych o rozkładzie jednostajnym na przestrzeni $(S, \mathcal{A}, \lambda)$, $X^1 = \xi^1, X^2, \dots, X^N$ jest ciągiem określonym wzorem (26) oraz zbiór Λ_N jest określony przez (41), to dla każdego $\varepsilon \in (0,1)$

$$P\{\Lambda_N > \varepsilon\} \leq (1 - \varepsilon)^N \quad (42)$$

Dowód tego twierdzenia zaprezentowany został w [13].

Z powyższego twierdzenia wynika, że jeżeli ustalimy liczbę $\eta \in (0,1)$ i dobierzemy najmniejsze $N=N(\varepsilon, \eta)$ takie, że

$$(1 - \varepsilon)^N \leq 1 - \eta \quad (43)$$

to jeżeli za pomocą algorytmu 1 wykonamy $N=N(\varepsilon, \eta)$ kroków iteracyjnych, to otrzymamy :

$$P\left\{\frac{\lambda\{x : f(x) < f(X^N)\}}{\lambda(S)} > \varepsilon\right\} \leq 1 - \eta \quad (44)$$

Z (43) otrzymujemy, że:

$$N \geq \left\lceil \frac{\log(1 - \eta)}{\log(1 - \varepsilon)} \right\rceil \quad (45)$$

gdzie symbol $\lceil x \rceil$ oznacza najmniejszą liczbę całkowitą nie mniejszą od x . Z (45) wynika, że aby np. z prawdopodobieństwem η zlokalizować minimum funkcji f jednej zmiennej z dokładnością do przedziału o długości nie przekraczającej ε , a więc z dokładnością do części zakresu zmienności argumentu równej ε należy wylosować co najmniej N punktów określonych wzorem (45). W ogólności, jeśli mamy funkcję k zmiennych określoną na przestrzeni $[0,1]^k$ i jeśli chcemy zlokalizować minimum tej funkcji z dokładnością do części zakresu zmienności każdego z argumentów równej ε , to musimy zlokalizować je z dokładnością do ε^k rozmiaru całej przestrzeni $[0,1]^k$, czyli liczba punktów, które należy wylosować wynosi:

$$N \geq \left\lceil \frac{\log(1 - \eta)}{\log(1 - \varepsilon^k)} \right\rceil \quad (46)$$

Dla przykładu, jeżeli w modelu Browna chcielibyśmy z prawdopodobieństwem $\eta=0.95$ zlokalizować minimum funkcji f (określonej przez (5)) z dokładnością do przedziału (bo funkcja (5) jest funkcją jednoargumentową) o długości nie przekraczającej $\varepsilon=0.1$, czyli z dokładnością do dziesiątej części zakresu zmienności argumentu, to należałoby wylosować

$$N \geq \left\lceil \frac{\log(1-0.95)}{\log(1-0.1)} \right\rceil = 28 \quad (47)$$

punktów.

Z kolei, jeżeli w modelu Holta chcielibyśmy z prawdopodobieństwem $\eta=0.95$ zlokalizować minimum funkcji f z dokładnością do dziesiątej części zakresu zmienności każdego z argumentów (funkcja f w modelu Holta jest dwuargumentowa, por. [11]), to należałoby już wylosować

$$N \geq \left\lceil \frac{\log(1-0.95)}{\log(1-0.1^2)} \right\rceil = 299 \quad (48)$$

punktów.

Należy zauważyć, że przedstawione wyżej zagadnienie jest równoważne znanemu w statystyce matematycznej zagadnieniu nieparametrycznych przedziałów (granic) tolerancji. W [4], s. 329÷331 przedstawiono formalne ujęcie tego zagadnienia.

Niech X oznacza zmienną losową ciągłą o gęstości $g(x)$. Rozpatrzmy całkę W określoną następująco:

$$W = \int_{L_1}^{L_2} g(x) dx \quad (49)$$

gdzie L_1 i L_2 są jednoznaczными funkcjami statystyk pozycyjnych w N -elementowych próbach prostych z populacji, której cechę X badamy. Noszą one nazwę granic tolerancji. Ponieważ L_1 i L_2 są zmiennymi losowymi, więc całka W też jest zmienną losową (o rozkładzie beta (por. [4], rozdz. 5.9)) i można ją utożsamiać z frakcją elementów populacji generalnej zawartą pomiędzy losowymi granicami L_1 i L_2 . W [4] pokazano, że dla dowolnej zmiennej losowej X typu ciągłego i dla dowolnego $\eta < 1$ można dobrać takie N , że

$$P(W \geq \varepsilon) = N \cdot (N-1) \cdot \int_{\varepsilon}^1 t^{N-2} \cdot (1-t) dt = 1 - \eta \quad (50)$$

a więc przy N odpowiednio dobranym prawdopodobieństwo, iż frakcja elementów populacji generalnej zawartych między najmniejszym i największym elementem w N -elementowej próbie jest co najmniej równa ε , jest równe $1-\eta$. Zauważmy, że jeżeli za W podstawimy

$$W = \frac{\lambda\{x: f(x) < f(X^N)\}}{\lambda(S)} \quad (51)$$

to (50) można utożsamiać z (44).

Powiedzmy jeszcze o zbieżności algorytmu 1. Mówi o tym następujące twierdzenie, które podamy bez dowodu (dowód znajduje się w [13]).

Twierdzenie 2 [13]

Jeżeli $(X^n)_{n \geq 1}$ jest ciągiem punktów losowych generowanych przez algorytm 1 (por. wzór (26)), to ciąg ten zbiega z prawdopodobieństwem 1 do minimum globalnego funkcji f (opisanej przez (27)).

Istotnym problemem, którym możemy być zainteresowani jest wyznaczenie takiej liczby losowań N , przy której wartość funkcji f z (27) będzie wyznaczona z odpowiednią dokładnością w stosunku do wartości optymalnej $f(\alpha^*)$. Oszacowanie tej liczby zostanie przedstawione w następnym rozdziale (por. (54)).

Analiza wyników symulacji

Przedstawimy obecnie wyniki badań zbieżności prezentowanych algorytmów oraz dokładności rozwiązań, które z nich otrzymujemy. W celu przeprowadzenia badań zbudowano symulator w języku *VisualBasic for Applications* działający w środowisku arkusza kalkulacyjnego *MsExcel*, realizujący prezentowane wcześniej algorytmy losowego poszukiwania minimum funkcji (27). Pojedynczy eksperyment symulacyjny polegał na wylosowaniu $N=100$ punktów losowych. Liczbę tę dobrano nieprzypadkowo i jak się okaże była wystarczająca do tego, aby osiągnąć dużą dokładność wartości funkcji $f(\alpha^N)$ z (27) w stosunku do wartości $f(\alpha^*)$ dla rozwiązania optymalnego α^* . Odpowiada ona w przybliżeniu wartościom $\varepsilon=0.02$ oraz $\eta=0.1$ z (45). Wykonywano $M=30$ eksperymentów symulacyjnych. Oznaczmy przez $\alpha_{i,j,k}^N$ wartość współczynnika α otrzymaną po N losowaniach w i -tym ($i=\overline{1, M}$) eksperymencie symulacyjnym, dla j -tego ($j=\overline{1,3}$) algorytmu oraz k -tej metody wygładzania (dla $k=1$ mamy prosty model Browna, $k=2$ – klasyczny m. Browna, $k=3$ – zmodyfikowany m. Browna). Ponieważ α może przyjmować różne wartości w różnych eksperymentach symulacyjnych, więc jest zmienną losową. W tym kontekście $\alpha_{i,j,k}^N$ jest i -tą realizacją zmiennej losowej α . Estymator punktowy wartości oczekiwanej parametru wygładzania α dla j -tego algorytmu oraz k -tego modelu wygładzania po N losowaniach punktów próbnych wyznaczmy następująco:

$$\tilde{\alpha}_{j,k}^N = \frac{1}{M} \cdot \sum_{i=1}^M \alpha_{i,j,k}^N, \quad j = \overline{1,3}, \quad k = \overline{1,3} \quad (52)$$

Zależność (52) przedstawia oszacowanie wartości parametru α dokonane przez j -ty algorytm zastosowany dla k -tego modelu wygładzania przy N losowaniach punktów próbnych.

Oznaczmy przez $\mathbf{N}=\{1, \dots, N\}$ zbiór numerów wylosowanych punktów próbnych. Z kolei przez $n_{i,j,k}$ oznaczmy następującą wielkość:

$$n_{i,j,k} = \mathbf{min} \left\{ n \in \mathbf{N} : \frac{f(\alpha_{i,j,k}^n) - f(\alpha_k^*)}{f(\alpha_k^*)} \leq 0.01 \right\} \quad (53)$$

gdzie $\alpha_{i,j,k}^n$ oznacza wartość parametru α otrzymaną do n -tego losowania w i -tym eksperymencie symulacyjnym ($i = \overline{1, M}$), dla j -tego algorytmu ($j = \overline{1, 3}$) i k -tego modelu wygładzania (dla $k=1$ mamy prosty model Browna, $k=2$ – klasyczny m. Browna, $k=3$ – zmodyfikowany m. Browna), a α_k^* oznacza wartość optymalną parametru α dla k -tego modelu wygładzania. Interpretacja wielkości $n_{i,j,k}$ jest następująca: jest to najmniejszy numer punktu próbnego (losowego), taki, że różnica wartości funkcji f dla tego rozwiązania (punktu próbnego) i dla rozwiązania optymalnego (α^*) jest nie większa niż jeden procent wartości $f(\alpha^*)$, dla j -tego algorytmu i k -tego modelu wygładzania. Ponieważ punkt $\alpha_{i,j,k}^n$ jest punktem losowym, więc $n_{j,k}$ jest zmienną losową a $n_{i,j,k}$ jest jej i -tą realizacją. Przeprowadzając M eksperymentów symulacyjnych otrzymamy pewien rozkład empiryczny zmiennej $n_{j,k}$, który może być reprezentowany np. za pomocą dystrybuanty empirycznej $F_{j,k}^e$. Wyznamy estymator punktowy $\hat{n}_{j,k}^{0.9}$ kwantyla rzędu 0.9 zmiennej losowej $n_{j,k}$. Jest to taka liczba, która spełnia warunek:

$$F_{j,k}^e(\hat{n}_{j,k}^{0.9}) = P(n_{j,k} < \hat{n}_{j,k}^{0.9}) = 0.9 \quad (54)$$

Interpretacja liczby $\hat{n}_{j,k}^{0.9}$ (dla j -tego algorytmu i k -tego modelu wygładzania) jest następująca: jest to minimalna liczba losowań punktów próbnych potrzebnych do tego, aby z prawdopodobieństwem 0.9 można było stwierdzić, że $f(\alpha_{\hat{n}_{j,k}^{0.9}})$ różni się od $f(\alpha^*)$ nie więcej niż o jeden procent. Pozostało jeszcze tylko dodać, że dla algorytmu 3 przyjęto $\theta=0.5$ (por. (33), (34)).

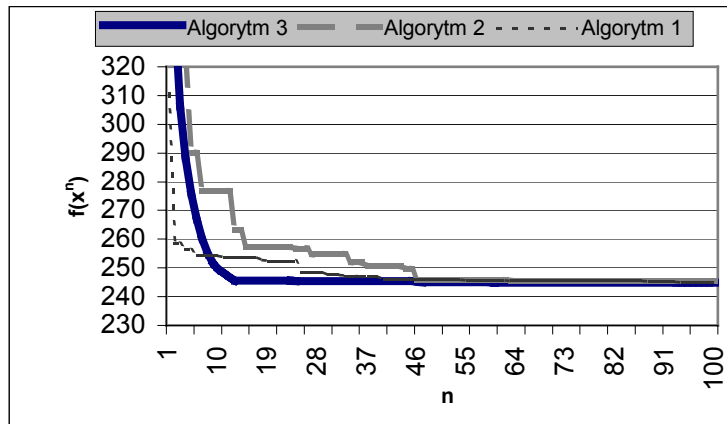
Przejdziemy teraz do interpretacji wyników.

W **Tabeli 1** przedstawiono wyniki estymacji $\tilde{\alpha}_{j,k}^N$, α^* , $f(\tilde{\alpha}_{j,k}^N)$, $f(\alpha^*)$, $\hat{n}_{j,k}^{0.9}$ dla rozpatrywanych modeli Browna oraz algorytmów (dla WIG-u w okresie X.1994÷XI.1999). Zauważmy, że optymalna wartość α^* współczynnika α dla prostego modelu Browna wynosi 1, czyli najlepsze efekty daje prognoza na postawie metody naiwnej (por. (1)), gdyż jest ona wyznaczana na chwilę t jako wartość z szeregu czasowego w chwili $t-1$.

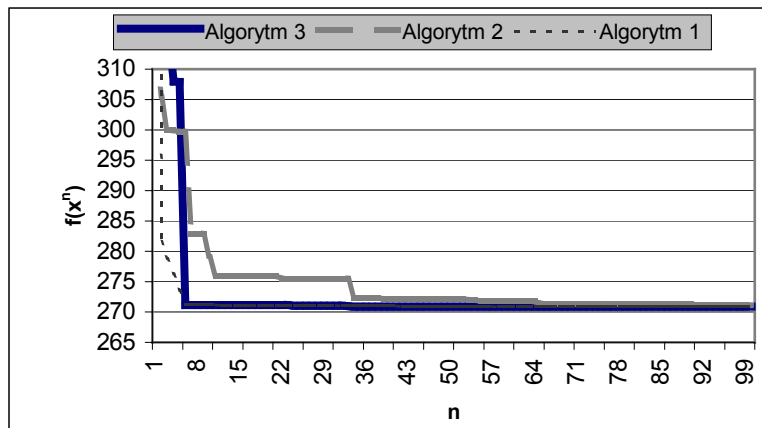
Na **Wykresach 1÷3** zaprezentowano szybkość zmniejszania się wartości funkcji (27) w zależności od numeru kroku n (numeru punktu losowego) dla WIG-u długookresowego (X.1994÷XI.1999). Widzimy, że szybkość ta jest duża, co potwierdzają wartości $\hat{n}_{j,k}^{0.9}$ w **Tabeli 1**.

Na **Wykresie 4** przedstawiono zmiany wartości indeksu WIG we wrześniu 1999r. oraz prognozy (liczone z jednookresowym wyprzedzeniem) wartości WIG-u na podstawie wszystkich trzech modeli. Zauważmy, że optymalna wartość α^* współczynnika α dla prostego modelu Browna wynosi również 1 (por. **Tabelę 2**), tak jak to było dla WIG-u długookresowego (por. **Tabelę 1**). W **Tabeli 2** przedstawiono wartości $\tilde{\alpha}_{j,k}^N$, α^* , $f(\tilde{\alpha}_{j,k}^N)$, $f(\alpha^*)$, $\hat{n}_{j,k}^{0.9}$ dla rozpatrywanych modeli oraz algorytmów (dla WIG-u na wrzesień 1999).

Wykres 1
Porównanie szybkości zbieżności algorytmów dla prostego modelu Browna



Wykres 2
Porównanie szybkości zbieżności algorytmów dla klasycznego modelu Browna



Wykres 3
Porównanie szybkości zbieżności algorytmów dla zmodyfikowanego modelu Browna

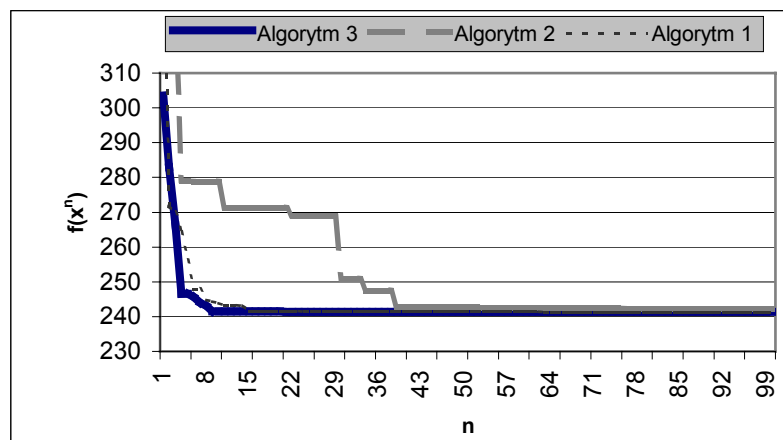


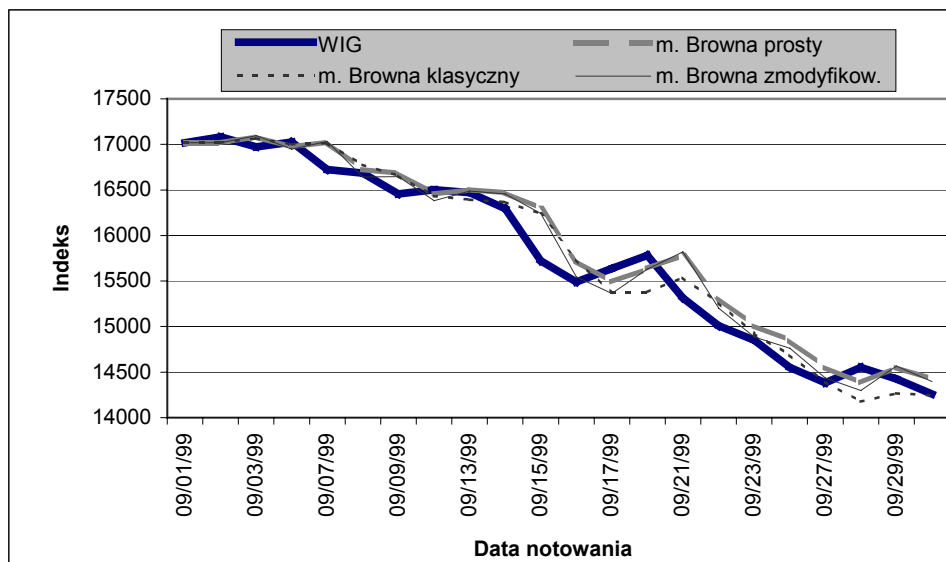
Tabela 1
Oszacowane wartości współczynników modeli Browna oraz odpowiadające im wartości funkcji f błędu standardowego prognozy dla WIG-u (X.1994÷XI.1999)

Rodzaj modelu Browna j	Rodzaj algorytmu k	$\tilde{\alpha}_{j,k}^N$	α^*	$f(\tilde{\alpha}_{j,k}^N)$	$f(\alpha^*)$	$\hat{n}_{j,k}^{0.9}$
Prosty ($j=1$)	1	0.985	1	245.20	244.45	40
	2	0.971		245.96		46
	3	0.997		244.59		12
Klasyczny ($j=2$)	1	0.635	0.635	270.88	270.88	5
	2	0.634		270.88		34
	3	0.633		270.88		6
Zmodyfikowany ($j=3$)	1	0.568	0.569	241.37	241.37	10
	2	0.563		241.40		39
	3	0.575		241.39		8

Tabela 2
Oszacowane wartości współczynników modeli Browna oraz odpowiadające im wartości funkcji f błędu standardowego prognozy dla WIG-u na wrzesień 1999r.

Rodzaj modelu Browna j	Rodzaj algorytmu k	$\tilde{\alpha}_{j,k}^N$	α^*	$f(\tilde{\alpha}_{j,k}^N)$	$f(\alpha^*)$	$\hat{n}_{j,k}^{0.9}$
Prosty ($j=1$)	1	0.985	1	235.20	233.95	24
	2	0.971		236.41		>100
	3	0.997		234.20		34
Klasyczny ($j=2$)	1	0.403	0.398	221.57	221.56	9
	2	0.394		221.57		27
	3	0.399		221.56		7
Zmodyfikowany ($j=3$)	1	0.635	0.635	219.89	219.89	24
	2	0.634		219.89		27
	3	0.639		219.89		7

Wykres 4
Wartości historyczne i najlepsze prognozy na podstawie różnych modeli Browna dla WIG-u na wrzesień 1999r.



Uwagi i wnioski

Z przeprowadzonych badań wynika kilka istotnych wniosków. Pierwszy to taki, że prognozowanie za pomocą prostego modelu Browna sprowadzało się w zasadzie do prognozowania metodą naiwną, gdyż $\alpha=1$ świadczy o tym, że przy budowie prognozy w ogóle nie bierzemy pod uwagę prognoz poprzednich (por. (1)). Potwierdziła się znana prawda, że prosty model Browna nie radzi sobie z szeregami czasowymi, w których występuje tendencja rozwojowa (metoda nie była w stanie zaproponować lepszego rozwiązania niż metoda naiwna (bo $\alpha=1$)). Okazuje się, że w szeregach czasowych o dużej zmienności wartości zmiennej prognozowanej metoda naiwna często daje najlepsze (ze względu na minimalizację błędu prognozy) wartości prognoz. Mimo to, model ten nie okazał się najgorszy (por. **Tabełę 1** i **Tabełę 2**). Jest to częściowym potwierdzeniem badań przeprowadzonych przez Makridakisa i innych ([7], [8]), z których wynikało, że prosta metoda Browna dawała najlepsze prognozy na jeden okres naprzód spośród 24 innych metod analizy szeregów czasowych zastosowanych przez nich w badaniach.

Bezwzględnie najlepszym modelem prognozowania okazał się model zaproponowany przez Z. Pawłowskiego w [9] (*zmodyfikowany klasyczny model Browna*), choć różnica w wartości błędu prognozy dla α^* była nieznaczna w stosunku do pozostałych modeli (por. **Tabełę 1** i **Tabełę 2**). To, że model ten okazał się najlepszy jest tylko potwierdzeniem hipotezy postawionej przez Z. Pawłowskiego w [9], że zmodyfikowana wersja klasycznego modelu Browna jest bardziej elastyczna od wersji klasycznej (i oczywiście od wersji prostej).

Jeżeli chodzi o badane algorytmy, to zauważmy, że liczba $\hat{n}_{j,k}^{0.9}$ losowań potrzebnych do tego, aby wartość funkcji błędu $f(\hat{n}_{j,k}^{0.9})$ nie różniła się więcej niż o 1 procent od wartości funkcji f dla optymalnej wartości α^* parametru α , w zasadzie była mniejsza od 15, dla najlepszego algorytmu z punktu widzenia szybkości zbieżności, czyli dla *algorytmu 3*. Należało się tego spodziewać, gdyż z idei tego algorytmu wynika, że jeżeli „złapie” on w którejś iteracji rozwiązanie lepsze od dotychczasowego, to poprzez zwiększanie prawdopodobieństwa wylosowania tego samego kierunku (który spowodował, że uzyskaliśmy lepsze rozwiązanie) oraz zmniejszanie długości kroku zmiany wartości parametru próbuje jak najlepiej wykorzystać fakt (poprzez przeszukanie otoczenia lepszego rozwiązania), że znalazł się w pobliżu tego rozwiązania. *Algorytmowi 3* dorównuje, pod względem szybkości zbieżności, *algorytm 1* natomiast „odstaje” od nich *algorytm 2*. Szczególnie jest to widoczne w modelach wieloparametrowych (por. [11]). Nie jest zaskoczeniem, że *algorytm 1* osiąga tak dobre wyniki w szybkości zbieżności, gdyż Z. Zieliński w [13] udowadnia, iż szybkość zbieżności tego najprostszego algorytmu losowego jest najlepsza spośród prezentowanych przez niego w cytowanej pracy. Zauważmy ponadto, że wyestymowane wartości $\tilde{\alpha}_{j,k}^N$, $j=1,3$, $k=1,3$ przedstawione w **Tabeli 1** i **Tabeli 2** w większości przypadków nie różnią się więcej niż o 0.02 od odpowiedniej wartości α^* . Jest to potwierdzeniem oszacowania ze wzoru (45) na liczbę punktów losowych, gdyż dla $N=100$ otrzymujemy na 90-cio procentowym poziomie ufności (tzn. $\eta=0.1$), że wartości parametru α^N nie powinny różnić się więcej niż o $\varepsilon=0.02$ od wartości α^* . Zauważmy ponadto, że gdybyśmy chcieli, aby wartość parametru α^N nie różniła się o więcej niż 0.02 od wartości α^* , to dokonując doboru wartości parametru α^N metodą przeszukiwania sieciowego (o której wspomnieliśmy we **Wstępie**) należałoby sprawdzić $N=100/2=50$ wartości parametru α (startując od $\alpha=0$ i przyjmując krok Δ zmian wartości α równy $\Delta=0.02$), podczas gdy używając *algorytmu 3* wykonywaliśmy z reguły mniej niż 15 kroków (por. wartości $\hat{n}_{j,k}^{0.9}$ w **Tabeli 1** i **Tabeli 2**). Ta różnica jest szczególnie widoczna w modelach wieloparametrowych (por. [11]).

Zaprezentowane algorytmy doboru optymalnych parametrów modeli wygładzania wykładniczego mają charakter uniwersalny i mogą być stosowane również do znajdowania optymalnych parametrów tych modeli dla innych szeregów czasowych. Szereg czasowy przedstawiający wartości WIG-u był tylko przykładem, dla którego dokonano analizy własności tych algorytmów.

Analizę wykorzystania modeli wygładzania wykładniczego dla szeregów niestacjonarnych do prognozowania wartości WIG-u przedstawiono w [11]. Ponadto pracę [11] można traktować jako uzupełnienie tego artykułu, gdyż wykorzystano tam również metody stochastycznego poszukiwania minimum funkcji błędu prognozy, przedstawione w niniejszej pracy, do wyznaczania optymalnego zestawu parametrów w dwuparametrowym modelu Holta oraz trzyparametrowym modelu Wintersa (addytywnym i multiplikatywnym). Dopiero stosowanie rozpatrywanych w pracy algorytmów do optymalizacji funkcji wieloparametrowej daje pełne porównanie korzyści z ich stosowania.

Literatura

- [1]. Brown R.G.: Statistical Forecasting for Inventory Control. *McGrow Hill*, New York 1959.
- [2]. Brown R.G.: Smoothing, Forecasting and Prediction of Discrete Time Series. *McGrow Hill*, New York 1963.
- [3]. Cieślak M. (red.): Prognozowanie gospodarcze. *PWN*, Warszawa 1997.
- [4]. Fisz M.: Rachunek prawdopodobieństwa i statystyka matematyczna. *PWN*, Warszawa 1958.
- [5]. Gardner, E. S., Jr.: Exponential smoothing: The state of the art. *Journal of Forecasting*, 4 (1985), 1-28.
- [6]. Grabowski W.: Programowanie matematyczne. *PWE*, Warszawa 1980.
- [7]. Makridakis S., Wheelwright S.C.: Forecasting Methods and Applications. *John Wiley & Sons*, New York 1989.
- [8]. Makridakis S., Andersen A., Carbone R., Fildes R., Hibon M., Lewandowski R., Newton J., Parzen R., Winkler R.: The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1 (1982), 11-153.
- [9]. Pawłowski Z.: Prognozy ekonometryczne. *PWN*, Warszawa 1973.
- [10]. STATISTICA PL dla Windows (Tom III): Statystyki II. *StatSoft Polska Sp. z o.o.*, Kraków 1997.
- [11]. Tarapata Z.: Analiza możliwości wykorzystania wybranych modeli wygładzania wykładniczego do prognozowania wartości WIG-u. *Zeszyty Naukowe Wyższej Szkoły Ekonomicznej*, Warszawa 2000. (w druku)
- [12]. Zangwill W. I.: Programowanie nieliniowe. *WNT*, Warszawa 1974.
- [13]. Zieliński R., Neumann P.: Stochastyczne metody poszukiwania minimum funkcji. *WNT*, Warszawa 1986.