

Jacek Szandula

Akademia Ekonomiczna we Wrocławiu

PROGNOZOWANIE SZEREGÓW CZASOWYCH Z NIEREGULARNYMI WAHANIAMI OKRESOWYMI METODĄ NAJMNIEJSZEGO SYMPLEKSU

1. Wstęp

W prognozowaniu szeregów czasowych charakteryzujących się nieregularnością wahań okresowych często powstaje problem określenia:

- a) długości cyklu i co jest z tym związane – punktów zwrotnych cyklu,
- b) wielkości amplitudy cyklu.

Celem artykułu jest zaproponowanie i dokonanie oceny przydatności metody najmniejszego sympleksu do prognozowania szeregów czasowych z nieregularnymi wahaniami okresowymi. Oczywiście istnieje wiele metod radzących sobie lepiej lub gorzej z wcześniej wspomnianymi problemami, niemniej jednak autor chce zweryfikować hipotezę o przydatności prezentowanej metody.

Ze względu na to, że metoda ta nie jest popularna, w pierwszej części artykułu przedstawiono główne jej założenia. Następnie na przykładzie wygenerowanych szeregów czasowych przeprowadzono konfrontację prezentowanej metody z modelami ARIMA.

2. Idea metody

Metoda zaproponowana przez Sugiharę i Maya [1990] została zaprojektowana w celu prognozowania szeregów czasowych charakteryzujących się dużymi wahaniami przypadkowymi. Jednak jak wykazał Drosdowsky [1994], metoda ta jest także użyteczna w prognozowaniu szeregów czasowych, w których w pewnych odstępach czasowych,

zwykle nieregularnych, pojawiają się określone prawidłowości. Oznacza to, że założenie o znajomości mechanizmu generującego szereg czasowy, które jest często stosowane w przypadku prognozowania szeregów czasowych, ustępuje tu założeniu o podobieństwie rozwoju. Z tego względu metodę tę można zaliczyć do analogii historycznych (zob. np. [Cieślak 2005]).

Istotą metody jest przedstawienie kolejnych k -elementowych fragmentów szeregu czasowego jako punktów w przestrzeni k -wymiarowej. Ostatni k -elementowy fragment szeregu czasowego, który autor nazwie punktem odniesienia, w przestrzeni k -wymiarowej ma współrzędne:

$$y_n = [y_{n-(k-1)g} \dots y_{n-2g} y_{n-g} y_n]^T, \quad (1)$$

gdzie: y_n – punkt odniesienia,

g – wielkość odstepu pomiędzy kolejnymi obserwacjami – zwykle $g = 1$.

W analogiczny sposób, jako k -wymiarowe punkty, przedstawiane są pozostałe fragmenty szeregu czasowego:

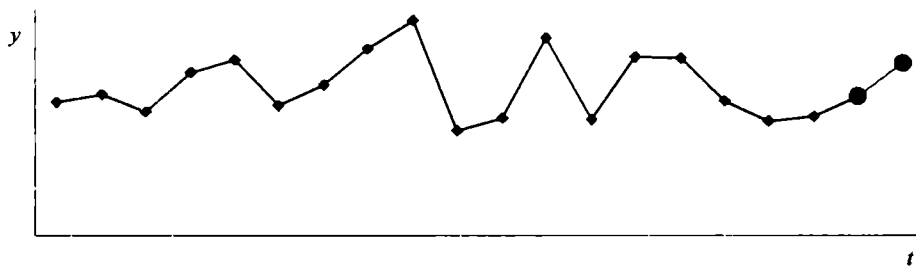
$$x_i = [x_{n-i-(k-1)g} \dots x_{n-i-2g} x_{n-i-g} x_{n-i}]^T, \quad (2)$$

gdzie: x_i – punkt reprezentujący i -ty (licząc w kierunku od obserwacji najmłodszych do najstarszych) fragment szeregu czasowego,

$i = 1, 2, \dots, i \leq n - k$.

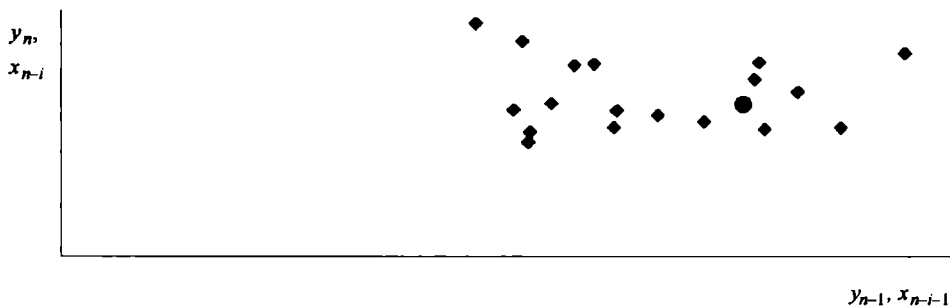
Spośród „sąsiadów” punktu y_n wybierana jest kombinacja $k+1$ punktów tworzących sympleks o najmniejszej średnicy zawierający punkt y_n . Jeżeli prognozowany punkt y_n znajduje się w wierzchołku, na krawędzi lub ścianie sympleksu, wówczas do prognozowania można wybrać sympleks o mniejszej niż k liczbie wymiarów, tzn. zbiór punktów ulega redukcji.

Obliczenie prognozy wymaga wyznaczenia współrzędnych wierzchołków sympleksu po $T - n$ momentach / okresach czasu (przez T rozumiany jest okres prognozowania). Na ich podstawie oraz przy użyciu współrzędnych barycentrycznych punktu y_n w znalezionym sympleksie ustalana jest wartość prognozy.



Rys. 1. Przykładowy szereg czasowy

Źródło: opracowanie własne.

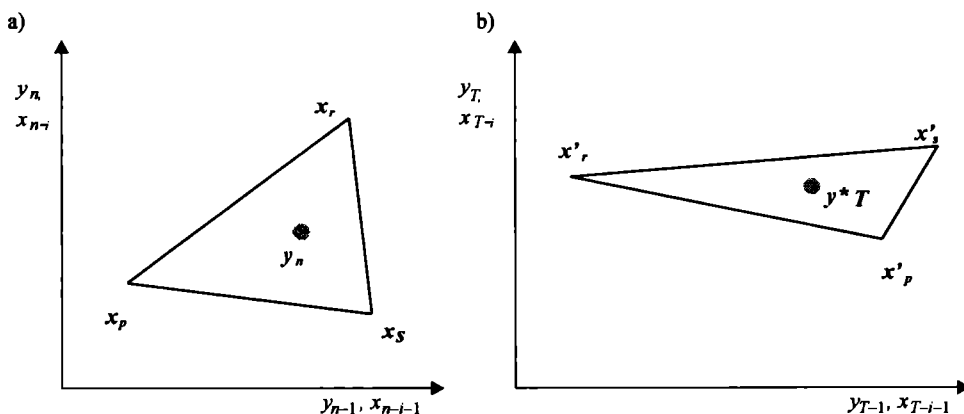


Rys. 2. Szereg czasowy jako zbiór punktów w przestrzeni dwuwymiarowej

Źródło: opracowanie własne.

Poszczególne kroki metody najłatwiej jest prześledzić dla przypadku, gdy $k = 2$ i $g = 1$. Rysunek 1 prezentuje przykładowy szereg czasowy. Dla ustalonego parametru $k = 2$ wybiera się dwie ostatnie obserwacje. Wybrane wartości przedstawia się w przestrzeni dwuwymiarowej jako punkt odniesienia y_n o współrzędnych $[y_{n-1}, y_n]$. Następnie to samo przeprowadza się dla pozostałych par następujących po sobie obserwacji. W ten sposób otrzymuje się zbiór punktów otaczających punkt odniesienia (rys. 2).

Z otrzymanego zbioru wybierane są trzy punkty ($k + 1$), tworzące wierzchołki sympleksu (w tym przypadku jest nim trójkąt) o najmniejszej średnicy spośród sympleksów zawierających punkt odniesienia. Przez średnicę rozumiana jest tu maksymalna odległość między punktami tworzącymi figurę.



Rys. 3. Prognozowanie przy użyciu metody najmniejszego sympleksu w przypadku $k = 2$

Źródło: opracowanie własne.

Rysunek 3a przedstawia szukany sympleks $x_p x_r x_s$. Odległości punktu y_n od poszczególnych wierzchołków trójkąta określają współrzędne barycentryczne punktu y_n w sym-

pleksie $x_p x_r x_s$. Aby wyznaczyć prognozę y^*_T , należy znaleźć wierzchołki nowego sympleksu $x'_p x'_r x'_s$, przy czym:

$$x'_i = x_{i-(T-n)}. \quad (3)$$

Po znalezieniu wierzchołków nowego sympleksu wyznacza się punkt prognozy y^*_T , który ma współrzędne barycentryczne w sympleksie $x'_p x'_r x'_s$ identyczne ze współrzędnymi barycentrycznymi punktu y_n w sympleksie $x_p x_r x_s$ (rys. 3b). Ostatnia współrzędna punktu y^*_T wyznacza wartość prognozy na okres T .

Sugihara i May [1990] przedstawiają także procedurę, dzięki której można wybrać optymalny wymiar przestrzeni sympleksu. W tym celu dzieli się szereg czasowy na dwie części. Na podstawie pierwszego fragmentu wyznaczane są prognozy wygasłe. Pozostała część służy weryfikacji trafności wyznaczonych prognoz. Ostateczny wybór parametru k oparty jest na współczynniku korelacji prognoz wygasłych z wartościami rzeczywistymi.

3. Eksperyment

Do sprawdzenia przydatności metody najmniejszego sympleksu użyto szeregów wygenerowanych za pomocą następującego modelu:

$$y_{t,i} = a_i \sin \left(\frac{2\pi(t - \sum_{j=0}^{i-1} b_j)}{b_i} \right), \quad (4)$$

gdzie: t – numer okresu ($t = 1, 2, \dots, n$),

i – numer cyklu,

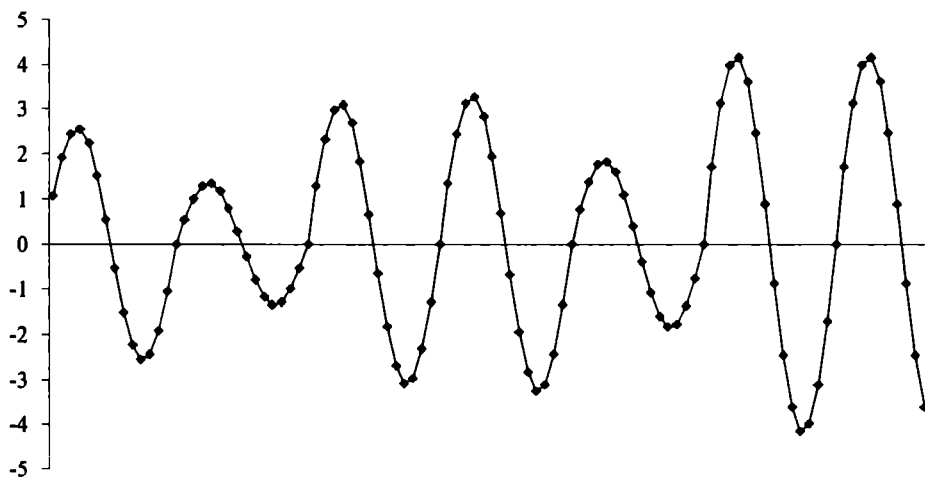
a_i – połowa amplitudy i -tego cyklu,

b_i – długość i -tego cyklu, $b_0 = 0$,

$y_{t,i}$ – wygenerowana wartość w i -tym cyklu i t -tym okresie.

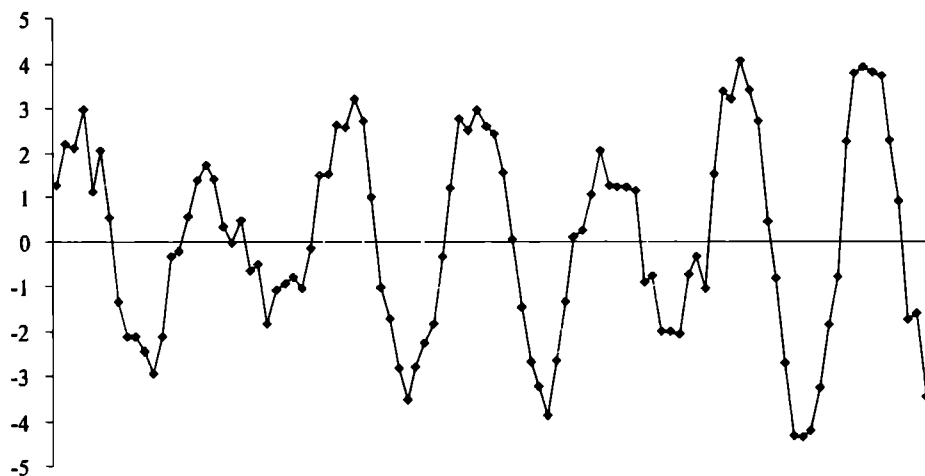
Wygenerowano sześć 900-elementowych szeregów – trzy podstawowe A, B i C oraz trzy powstałe poprzez nałożenie szumu na szeregi podstawowe. Przy generowaniu szeregu A w modelu (4) wartości a_i losowano z rozkładu normalnego o średniej 0 i odchyleniu standardowym 1 ($a_i \sim N(3; 1)$), b_i ustalono na 15 dla wszystkich cykli. Szereg B charakteryzował się stałą amplitudą – $a_i = 1$ – a także zmienną długością cyklu – b_i miało rozkład jednostajny dyskretny na przedziale $[1; 20]$. W szeregu C losowa była zarówno amplituda – $a_i \sim N(3; 1)$ – jak i długość cyklu – b_i ponownie było losowane z rozkładu jednostajnego dyskretnego z przedziału $[1; 20]$. Szeregi z szumem powstały poprzez addytywne nałożenie szumu losowanego z rozkładu normalnego o średniej 0 i odchyleniu

standardowym 0,5 na szeregi podstawowe. Rysunki 4-9 przedstawiają pierwsze 100 obserwacji utworzonych szeregów.



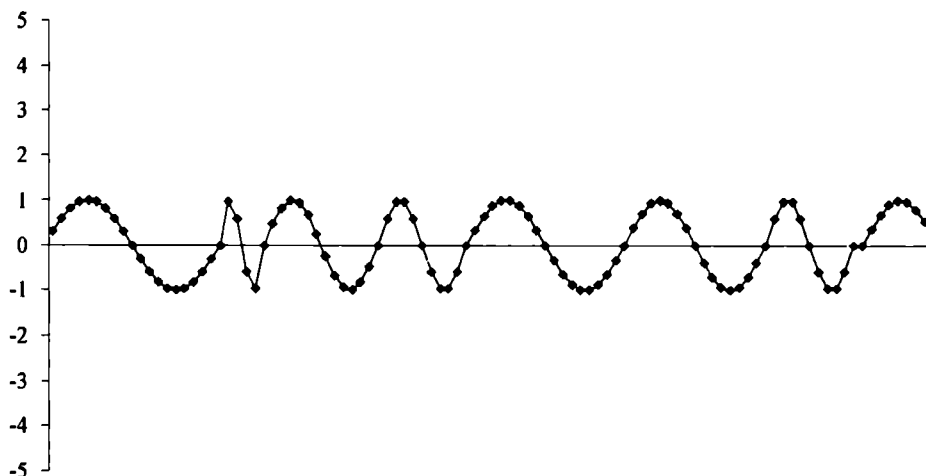
Rys. 4. Szereg A: pierwsze 100 obserwacji

Źródło: opracowanie własne.



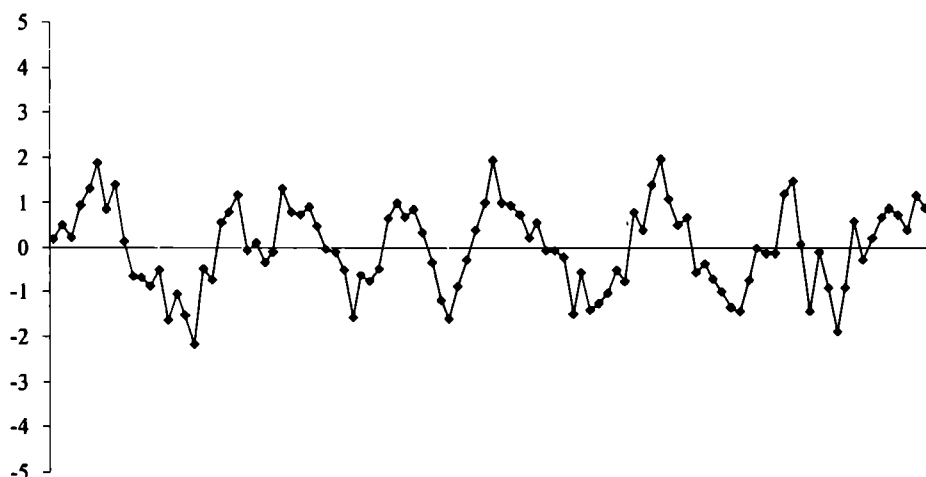
Rys. 5. Szereg A + szum: pierwsze 100 obserwacji

Źródło: opracowanie własne.



Rys. 6. Szereg B: pierwsze 100 obserwacji

Źródło: opracowanie własne.

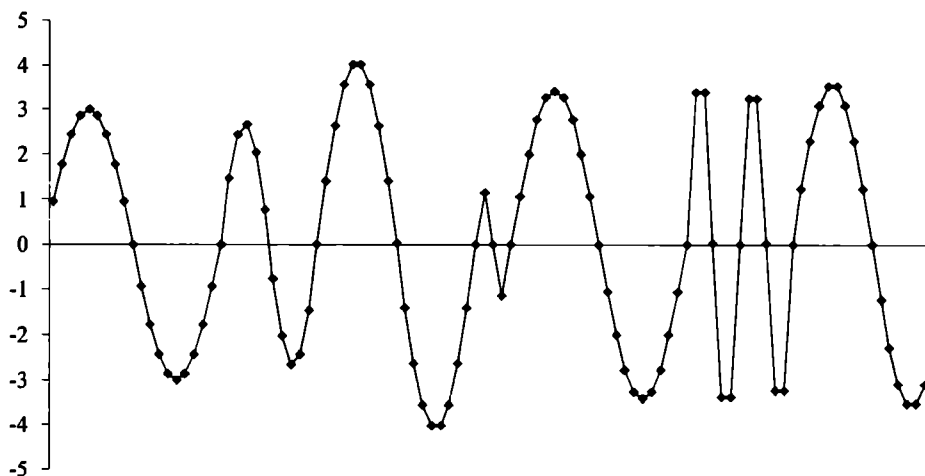


Rys. 7. Szereg B + szum: pierwsze 100 obserwacji

Źródło: opracowanie własne.

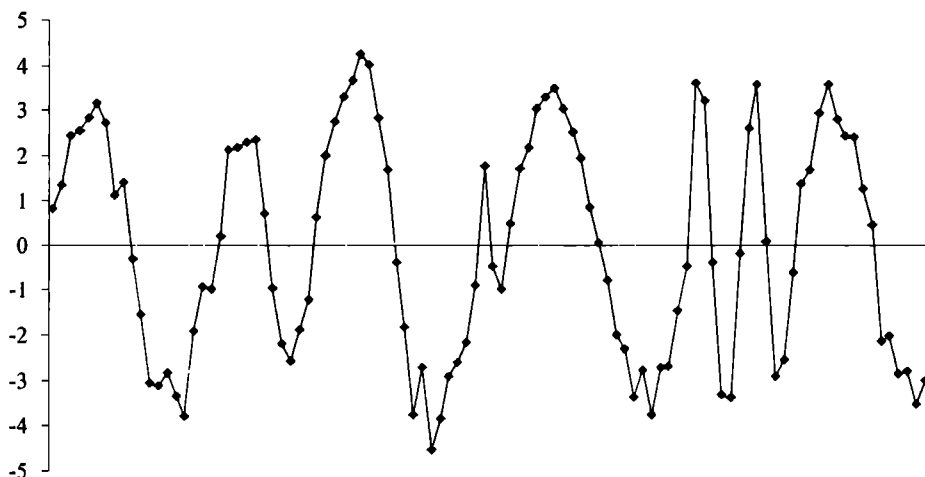
Każdy szereg został podzielony na dwa fragmenty. Pierwszy, składający się z 600 obserwacji, służył za bibliotekę wzorów. Kolejne 300 obserwacji posłużyło sukcesywnemu wyznaczaniu prognoz oraz ocenie trafności tych prognoz. Prognozy wyznaczano dla wyprzedzeń od 1 do 3. W przypadku metody najmniejszego sympleksu biblioteka wzorów została użyta do wyznaczenia zbioru punktów będących potencjalnymi wierzchołkami sympleksów. W przeprowadzonym badaniu rozpatrywano wymiary symplek-

sów od $k = 2$ do 7. W przypadku modeli ARIMA (SARIMA) na podstawie pierwszego fragmentu szeregu oszacowano parametry modelu.



Rys. 8. Szereg C: pierwsze 100 obserwacji

Źródło: opracowanie własne.



Rys. 9. Szereg C + szum: pierwsze 100 obserwacji

Źródło: opracowanie własne.

Tabela 1 podsumowuje modele ARIMA uzyskane dla poszczególnych szeregów. Jedynie dla szeregu A oraz A + szum uzyskano modele sezonowe. W pozostałych przypadkach, gdy długość cyklu nie była stała, uzyskane modele (najlepsze w sensie, dopasowania, istot-

ności parametrów i autokorelacji składnika losowego) nie rozpoznawały sezonowości. Próba narzucenia postaci sezonowej powodowała pogorszenie jakości modeli.

Tabela 1. Postacie modeli ARIMA uzyskane dla poszczególnych szeregów czasowych

Szereg	Postać modelu ARIMA (SARIMA)	Standardowy błąd oceny modelu
A	$(1,0,2) \times (1,1,0)_{15}$	0,211
B	$(1,0,1)$	0,324
C	$(1,0,2)$	0,859
A + szum	$(1,0,0) \times (2,0,0)_{15}$	0,827
B + szum	$(1,0,0)$	0,718
C + szum	$(1,0,2)$	1,174

Źródło: opracowanie własne.

Do analizy trafności wyznaczonych prognoz wybrano:

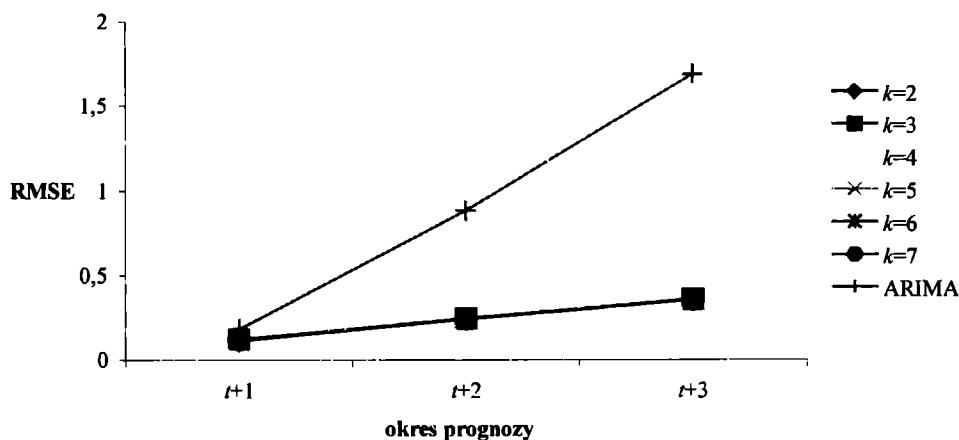
$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2}, \quad (5)$$

gdzie: RMSE – pierwiastek średniego kwadratowego błędu prognoz *ex post*,

N – liczba weryfikowanych prognoz,

y_t – wartość rzeczywista w t -tym okresie,

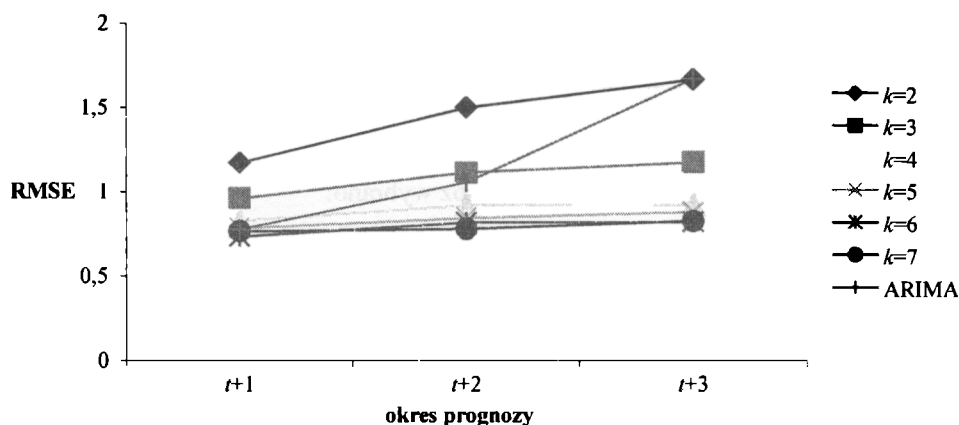
y_t^* – prognoza na t -ty okres.



Rys. 10. Szereg A: błędy prognoz wygasłych

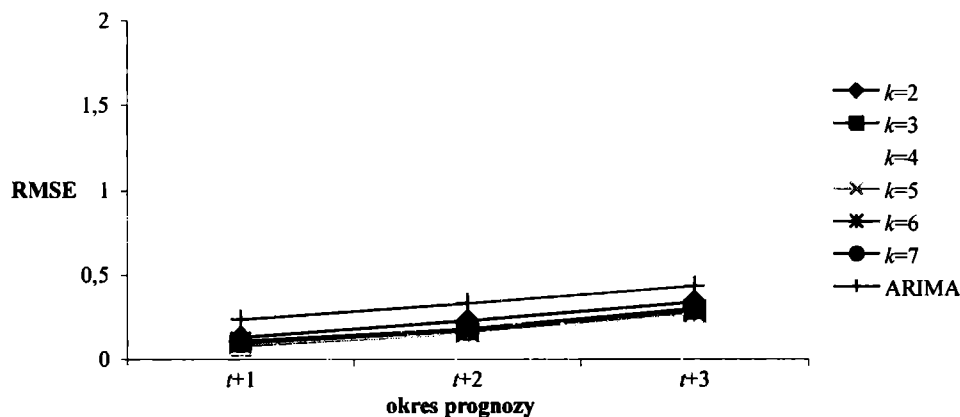
Źródło: opracowanie własne.

Rysunki 10-15 przedstawiają błędy prognoz dla wyprzedzeń od 1 do 3 wyznaczanych dla poszczególnych szeregów przy użyciu różnych modeli. W przypadku szeregu A oraz B prognozy wyznaczone za pomocą metody najmniejszego sympleksu obarczone były mniejszymi błędami niż prognozy wyznaczone modelami ARIMA niezależnie od wyboru przestrzeni k oraz okresu wyprzedzenia. Dla szeregu C, prognozując na jeden okres do przodu, uzyskano również bardziej trafne prognozy w przypadku metody najmniejszego sympleksu. Wzrost wartości wyprzedzenia powodował w tym przypadku zacieranie się różnic. Wprowadzenie do szeregów szumu spowodowało relatywne pogorszenie wyników uzyskiwanych za pomocą prezentowanej metody w porównaniu z modelami ARIMA. W przypadku szeregu B + szum wyniki uzyskane dla modelu ARIMA są najlepsze dla wszystkich wyprzedzeń.



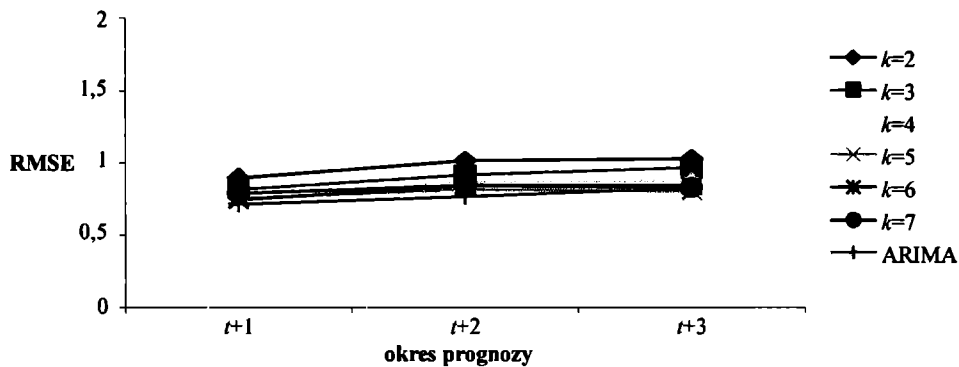
Rys. 11. Szereg A + szum: błędy prognoz wygaśłych

Źródło: opracowanie własne.



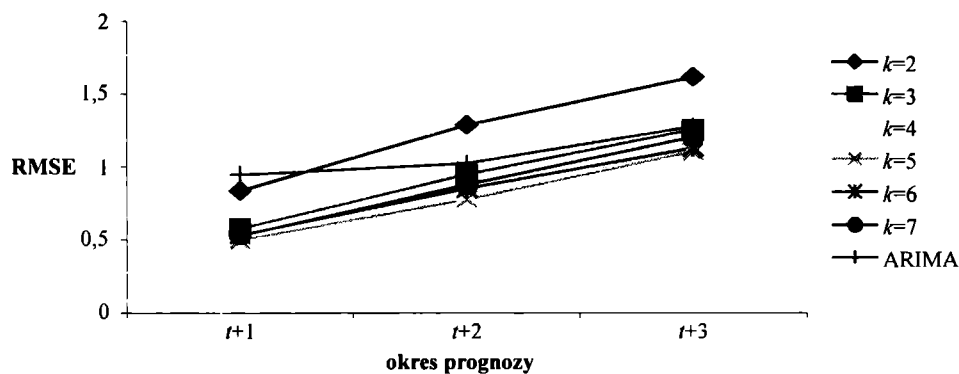
Rys. 12. Szereg B: błędy prognoz wygaśłych

Źródło: opracowanie własne.



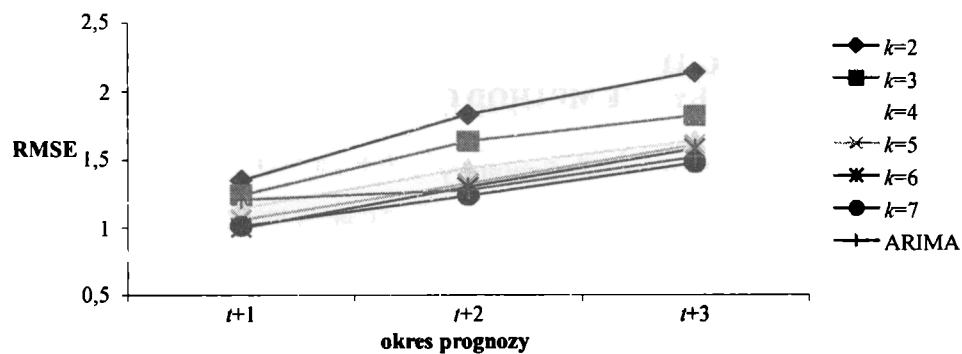
Rys. 13. Szereg B + szum: błędy prognoz wygasłych

Źródło: opracowanie własne.



Rys. 14. Szereg C: błędy prognoz wygasłych

Źródło: opracowanie własne.



Rys. 15. Szereg C + szum: błędy prognoz wygasłych

Źródło: opracowanie własne.

4. Wnioski

Przeprowadzony eksperyment pozwala stwierdzić, że metoda najmniejszego sympleksu może służyć prognozowaniu szeregów czasowych z nieregularnymi wahaniami okresowymi. Uzyskane wyniki nie są wystarczające, aby stwierdzić, że metoda ta jest wyraźnie lepsza lub gorsza od modeli ARIMA. W tym celu należałoby poddać badaniu znaczącą statystycznie liczbę szeregów czasowych spośród rozpatrywanych tu typów. Wskazane jest również przeprowadzenie eksperymentu dla wyższych wymiarów przestrzeni sympleksów. W dalszych badaniach należałoby się również skupić nad wpływem szumu na przydatność metody najmniejszego sympleksu.

Możliwości zastosowania prezentowanej metody ograniczają się do szeregów stacjonarnych w średniej. Wynika to z konieczności znalezienia sympleksu otaczającego punkt odniesienia. W przypadku tendencji rozwojowej kolejne fragmenty szeregu mogą tworzyć wielowymiarowe punkty, dla których nie da się znaleźć żadnego sympleksu. Należy także mieć na uwadze fakt, że taka sytuacja może się zdarzyć również w przypadku szeregów stacjonarnych, zwłaszcza gdy wymiar przestrzeni sympleksu jest duży.

Literatura

- Box G.E.P., Jenkins G.M., *Analiza szeregów czasowych*, PWN, Warszawa 1983.
 Cieślak, M., *Prognozowanie analogowe*, [w:] *Prognozowanie gospodarcze*, red. M. Cieślak, Wydawnictwo Naukowe PWN, Warszawa 2005.
 Drosdowsky W., *Analogue (non – linear) Forecasts of the Southern Oscillation Index Time Series*, „Weather Forecasting” 1994, 78-84.
 Sugihara, G., May, R.M., *Nonlinear Forecasting as a Way of Distinguishing Chaos from Measurement Error in Time Series*, „Nature” 1990, 344, 734-741.

FORECASTING TIME SERIES WITH IRREGULAR PERIODIC FLUCTUATIONS BY THE METHOD OF THE SMALLEST SIMPLEX

Summary

The paper presents the conception of using the Sugihara and May's method of the smallest simplex for forecasting time series with irregular periodic fluctuations. The idea of the method has been summarized. The method has been tested on generated time series:

- a) with constant cycle length and random amplitude,
- b) with constant amplitude and random cycle length,
- c) with random both amplitude and cycle length.

The forecasts have been verified. The results of the method have been compared with ARIMA models.