Uniwersytet
Ekonomiczny
w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 289 · 2016

Monika Hadaś-Dyduch

Uniwersytet Ekonomiczny w Katowicach
Wydział Ekonomii
Katedra Metod Statystyczno-Matematycznych w Ekonomii
monika.dyduch@ue.katowice.pl

WIELOMIANOWA GENERACJA DANYCH W ANALIZIE FALKOWEJ

Streszczenie: Celem badania jest ocena wpływu zaproponowanej metody generacji dodatkowych elementów szeregu na dokładność prognozy. Generowane, dodatkowe elementy szeregu służą do wyznaczenia współczynników, z których wyznacza się współczynniki transformaty falkowej na pierwszym poziomie rozdzielczości falki.

Celem oceny wielomianowej metody rozszerzenia danych wykonano predykcję szeregu, prezentującego stopę bezrobocia państw strefy euro.

Otrzymane wyniki zestawiono z bardziej trywialnymi metodami generacji dodatkowych elementów w transformacie falkowej

Słowa kluczowe: falki, analiza falkowa, transformata falkowa, predykcja.

Wprowadzenie

Konieczność rozszerzenia szeregu danych wejściowych do wyznaczenia współczynników falkowych pojawia się w przypadku filtrów, których długość L jest większa od 2. Wynika to z tego, że przy obliczaniu współczynników rozwinięcia falkowego, dla ostatnich elementów sygnału skończonego, filtr – teoretycznie – powinien wyjść poza sygnał. Nie jest to możliwe. Istnieją różnorodne sposoby rozwiązania tego problemu. Jednym ze sposobów jest rozszerzenie szeregu. W zależności od zastosowanej metody rozszerzenia szeregów, uzyskujemy różne błędy predykcji. Celem artykułu jest ocena wpływu zaproponowanej metody generacji dodatkowych elementów szeregu na dokładność końcowej wartości prognozy szeregu.

1. Zdefiniowanie falki

Falka to funkcja *f* o następujących właściwościach [Przelaskowski, 2002]:

- $f \in L^2(R)$, czyli energia *f* jest skończona: $\int \left| f(t) \right|^2 dt < \infty$,
- wartość średnia *f* wynosi zero, tj. $\int f(t) dt = 0$,

  warunki te wymuszają co najmniej kilka oscylacji,
- alternatywnie do a) i b): $\int \frac{\left| F(\omega)^2 \right|}{\omega} d\omega < \infty$.

Warunki a) i b) oraz c) są równoważne, jeśli *f* zanika szybciej niż $\left| t \right|^{-1}$ dla $t \to \infty$.

Wśród podstawowych cech falek możemy wyróżnić:

- silnie wyróżniona jest lokalizacja w czasie, tj. funkcja jest „lokalna”,
- nośnik (zbiór niezerowych wartości) jest zwarty (czyli domknięty i ograniczony) i niepusty,
- nośnik jest „prawie zwarty” (widmo częstotliwościowe ma zwarty nośnik),
- kształt przypomina gasnące pobudzenie ośrodka, tj. falę z gasnącymi amplitudami kolejnych oscylacji oddalających się od zaburzenia centralnego [Przelaskowski, 2002].

2. Rozszerzenie szeregu danych

Szereg wejściowy danych zapisujemy jako:

$$s_0, s_1, s_2, \cdots, s_{2^n-2}, s_{2^n-1} \tag{1}$$

2.1. Metoda wielomianowa

Główne rozszerzenie szeregu danych metodą wielomianową możemy zapisać następująco:

$$s_{2^n}, \cdots, s_{2^{n+1}-1}, \tag{2}$$

a krótkie rozszerzenie zapisujemy jako kopię dwóch pierwszych elementów szeregu wejściowego, tj.:

$$s_0, s_1$$

Całkowite rozszerzenie szeregu wówczas ma postać:

$$s_{2^n}, \cdots, s_{2^{n+1}-1}, s_0, s_1$$

Zatem otrzymujemy nowy szereg, rozszerzony odpowiednio w stosunku do wejściowego szeregu:

$$s_0, s_1, s_2, \cdots, s_{2^n-2}, s_{2^n-1}\ s_{2^n}, \cdots, s_{2^{n+1}-1}\ s_0, s_1 \tag{3}$$

Aby można było przystąpić do wyznaczenia współczynników transformaty falkowej, należy w pierwszej kolejności wyznaczyć brakujące elementy szeregu (3). Przyjmując, że:

$$s_0 - s_{2^{n+1}-1} = s_1 - s_0$$

otrzymujemy równanie, z którego można wyznaczyć ostatni element głównego rozszerzenia szeregu:

$$s_{2^{n+1}-1} = 2s_0 - s_1$$

Z wielomianu [zob. Nievergelt, 2001]:

$$\begin{aligned} p(r) = p_0 + p_1(r - [2^n - 1]) + p_2(r - [2^n - 1])(r - [2^n]) + \\ p_3(r - [2^n - 1])(r - [2^n])(r - [2^{n+1} - 1]) \end{aligned} \tag{4}$$

przy warunkach:

$$p(2^n - 1) = s_{2^n-1} \tag{5}$$

$$p(2^n) = s_{2^n} \tag{6}$$

$$p(2^n) = 2s_{2^n-1} - s_{2^n-2} \tag{7}$$

$$p(2^{n+1} - 1) = s_{2^{n+1}-1} \tag{8}$$

$$p(2^{n+1} - 1) = 2s_0 - s_1 \tag{9}$$

$$p(2^{n+1}) = s_0 \tag{10}$$

mamy:

$$p_0 = s_{2^n-1} \tag{11}$$

$$p_1 = s_{2^n-1} - s_{2^n-2} \tag{12}$$

$$p_2 = \frac{2s_0 - s_1 - s_{2^n-1} - 2^n p_1}{2^n\left(2^n - 1\right)} \tag{13}$$

$$p_2 = \frac{s_0 - s_{2^n-1} - (2^n + 1)p_1}{2^n\left(2^n + 1\right)} - p_2 \tag{14}$$

Zatem:

$$s_k := p(k),\ k \in \left[2^n + 1,\ 2^{n+1} - 2\right]$$

2.2. Metody alternatywne

Przyjmujemy, że:

$$s_0, s_1, s_2, \cdots, s_{2^n-2}, s_{2^n-1} = p_0, p_1, p_2, \cdots, p_{2^n-2}, p_{2^n-1}$$

Wśród alternatywnych, zdecydowanie bardziej naiwnych metod rozszerzenia szeregu danych wejściowych do wyznaczenia współczynników falkowych można, m.in. zaproponować następujące sposoby:

- metoda 1:

$$\underbrace{0,0,0,\cdots,0,}_{rozszerzenie}\ \underbrace{p_0, p_1, p_2, \cdots, p_{2^n-2}, p_{2^n-1},}_{szereg\ danych}\ \underbrace{0,0,0,\cdots,0}_{rozszerzenie} \tag{15}$$

- metoda 2:

$$\underbrace{p_{2^n-1} \cdots, p_0,}_{rozszerzenie}\ \underbrace{p_0, p_1, p_2, \cdots, p_{2^n-2}, p_{2^n-1},}_{szereg\ danych}\ \underbrace{p_{2^n-1}, \cdots, p_0}_{rozszerzenie} \tag{16}$$

- metoda 3:

$$\underbrace{p_0, p_1, p_2, \cdots, p_{2^n-2}, p_{2^n-1},}_{rozszerzeie}\ \underbrace{p_0, p_1, p_2, \cdots, p_{2^n-2}, p_{2^n-1},}_{szeregdanych}\ \underbrace{p_0, p_1, p_2, \cdots, p_{2^n-2}, p_{2^n-1}}_{rozszerzeie} \tag{17}$$

- metoda 4:

$$\underbrace{p_0, p_1, p_2, \cdots, p_{2^n-2}, p_{2^n-1},}_{szereg\ danych}\ \underbrace{p_{2^n-1}, \cdots, p_0,}_{rozszerzenie}\ \underbrace{p_0, p_1}_{krótkie\ rozszerzenie} \tag{18}$$

Przykładowo, dla szeregu cztero-elementowego:

$$s_0, s_1, s_2, s_3 := p_0, p_1, p_2, p_3$$

mamy:

- metoda 1:

$$\underbrace{0,0,0,0,}_{rozszerzenie}\ \underbrace{p_0, p_1, p_2, p_3,}_{szereg\ danych}\ \underbrace{0,0,0,0}_{rozszerzenie}$$

$$\underbrace{s_{-4}, s_{-3}, s_{-2}, s_{-1},}_{rozszerzenie}\ \underbrace{s_0, s_1, s_2, s_3,}_{szereg\ danych}\ \underbrace{s_4, s_5, s_6, s_7}_{rozszerzenie}$$

- metoda 2:

$$\underbrace{p_3, p_2, p_1, p_0,}_{rozszerzenie}\ \underbrace{p_0, p_1, p_2, p_3,}_{szereg\ danych}\ \underbrace{p_3, p_2, p_1, p_0}_{rozszerzenie}$$

$$\underbrace{s_{-4}, s_{-3}, s_{-2}, s_{-1},}_{rozszerzenie}\ \underbrace{s_0, s_1, s_2, s_3,}_{szereg\ danych}\ \underbrace{s_4, s_5, s_6, s_7}_{rozszerzenie}$$

- metoda 3:

$$\underbrace{p_0, p_1, p_2, p_3}_{rozszerzenie}, \underbrace{p_0, p_1, p_2, p_3}_{szereg\ danych}, \underbrace{p_0, p_1, p_2, p_3}_{rozszerzenie}$$

$$\underbrace{s_{-4}, s_{-3}, s_{-2}, s_{-1}}_{rozszerzenie}, \underbrace{s_0, s_1, s_2, s_3}_{szereg\ danych}, \underbrace{s_4, s_5, s_6, s_7}_{rozszerzenie}$$

- metoda 4:

$$\underbrace{p_0, p_1, p_2, p_3}_{szereg\ danych}, \underbrace{p_3, p_2, p_1, p_0}_{rozszerzenie}, \underbrace{p_0, p_1}_{kr\acute{o}tkie\ rozszerzenie}$$

$$\underbrace{s_0, s_1, s_2, s_3}_{szereg\ danych}, \underbrace{s_4, s_5, s_6, s_7}_{rozszerzenie}, \underbrace{s_8, s_9}_{kr\acute{o}tkie\ rozszerzenie}$$

3. Analiza empiryczna

Badanie przeprowadzono dla szeregu czasowego, prezentującego bezrobocie państw strefy euro. Dane dotyczą lat 1997-2014 ((EA11-2000, EA12-2006, EA13-2007, EA15-2008, EA16-2010, EA17-2013, EA18-2014, EA19), średnia roczna, tysiąc osób). Dane dotyczące liczby osób bezrobotnych w strefie euro, wykorzystane do implementacji, pobrano z bazy Eurostat.

3.1. Dane i rozszerzenie danych

Jak wyżej wspomniano, badanie przeprowadzono dla danych rocznych, czyli na szeregu 17-elementowym. Analiza przedstawiona w tym artykule ma na celu wskazanie najlepszego rozszerzenia szeregu, na małej próbie, która daje najmniejszy błąd predykcji.

Celem uproszczenia opisu analizy, dzielimy szereg na krótsze, zachodzące na siebie szeregi. Przyjmujemy subiektywnie podział na szeregi 4-elementowe. Wówczas otrzymujemy 15 nowych szeregów.

Pierwszy szereg 4-elementowy składa się z następujących wartości:

$$14\,122{,}0;\ 13\,503{,}0;\ 12\,428{,}0;\ 11\,262{,}0 \tag{19}$$

Zatem rozszerzenie w postaci ogólnej dla szeregu (19) możemy zapisać:

$$\underbrace{s_0, s_1, s_2, s_3}_{szereg\ danych}, \underbrace{s_4, s_5, s_6, s_7}_{rozszerzenie}, \underbrace{s_8, s_9}_{kr\acute{o}tkie\ rozszerzenie}$$

$$\underbrace{14\,122{,}0,\ 13\,503{,}0, 12\,428{,}0,\ 11\,262{,}0}_{\text{Dane}}\ \underbrace{s_4, s_5, s_6, s_7}_{rozszerzenie}, \underbrace{s_8, s_9}_{kr\acute{o}tkie\ rozszerzenie}$$

Wartości s_8, s_9 są równe odpowiednio s_0, s_1, czyli 14 1220,13 503,0. Z warunków (6) i (7) mamy:

$$s_{2^n} = 2s_{2^n-1} - s_{2^n-2}\,.$$

Zatem wartość s_4 wyznaczona z powyższego wzoru, wynosi odpowiednio:

$$s_4 = 2s_3 - s_2$$

$$s_4 = 2 \cdot 11262 - 12428$$

$$s_4 = 10096$$

Z przyrównania warunków (8), (9) mamy:

$$s_{2^{n+1}-1} = 2s_0 - s_1$$

Zatem wartość s_7 wyznaczona z powyższego wzoru, wynosi odpowiednio:

$$s_7 = 2s_0 - s_1$$

$$s_7 = 2 \cdot 14122{,}0 - 13503{,}0$$

$$s_7 = 14741$$

Pozostaje wyznaczyć brakujące wartości rozszerzenie, tj. s_6 i s_5.

Podstawiając za *n* liczbę 2 oraz za *k* odpowiednio liczbę 5 oraz 6 wyznaczamy z wielomian opisanego równaniem (4) brakujące wartości s_6 i s_5, przyjmując przy tym:

$$p_0 = s_3$$

$$p_1 = s_3 - s_2$$

$$p_2 = \frac{2s_0 - s_1 - 5s_3 + 4s_2}{12}$$

$$p_3 = \frac{-7s_0 + 5s_1 + 7s_3 - 5s_2}{60}$$

Zatem: s_5 11263,5 oraz s_6 = 13300.

Postępując analogicznie, wyznaczamy rozszerzenia pozostałych szeregów. Następnie wyznaczamy odpowiednie współczynniki falkowe.

3.2. Predykcja

Analizę i predykcję szeregów czasowych można sporządzić różnymi niekonwencjonalnymi metodami [Barczak, 2013; Hadaś-Dyduch, 2013, 2015a, 2015b, 2016a, 2016b; Przybylska-Mazur, 2013; Janiga-Ćmiel, 2010]. Można również do analizy szeregów aplikować metody dotychczas stosowane w innych dziedzinach, np. metody proponowane przez A. Biernackiego [2009]. Testy porównujące dokładność prognoz, wyznaczonych na podstawie różnych modeli omówiono w pracy A. Przybylskiej-Mazur [2015]. W obecnej pracy do predykcji aplikowano dość prosty model, którego główne zasady opisano poniżej.

Autorski model zastosowany do badania składa się z kilku zasadniczych etapów. W pierwszej kolejności celem uzyskania dokładniejszych prognoz, wyjściowy szereg czasowy dzieli się na mniejsze jednostki szeregowe. Podział szeregu na mniejsze jednostki jest subiektywny. Jednakże wcześniejsze badania dowodzą, że najlepszy jest następujący podział:

$$\underbrace{y_1,\ y_2,\ y_3,\ y_4,}_{\textit{nowy szereg } 1}\ \underbrace{y_2,\ y_3,\ y_4,\ y_5,}_{\textit{nowy szereg } 2}\ \underbrace{y_3,\ y_4,\ y_5, y_6,}_{\textit{nowy szereg } 3}$$

$$y_4,\ y_5, y_6,\ y_7,\quad y_5, y_6,\ \ y_7,\ y_8,\quad y_6,\ y_7,\ y_8,\ y_9,\quad y_7,\ y_8,\ y_9,\ y_{10},$$

$$y_8,\ y_9,\ y_{10}, y_{11},\quad y_9,\ y_{10}, y_{11},\ y_{12},$$

$$y_{10},\ y_{11},\ y_{12},\ y_{13},\quad y_{11},\ y_{12},\ y_{13},\ y_{14},\quad y_{12},\ y_{13},\ y_{14}, y_{15}$$

$$y_{13},\ y_{14}, y_{15}, y_{16}\ ,\ \underbrace{y_{14}, y_{15}, y_{16}, y_{17}}_{\textit{nowy szereg } 14}\ \underbrace{y_{15}, y_{16}, y_{17}, y_{18}}_{\textit{nowy szereg } 15}$$

Każdy szereg utworzony z szeregu głównego jest traktowany jako oddzielny szereg czasowy. W tym przypadku dla każdego utworzonego 4-elementowego szeregu, aplikujemy jedną z metod rozszerzenia szeregów. Następnie obliczamy odpowiednie elementy rozszerzeń, a w kolejnym etapie wyznaczamy współczynniki falkowe pierwszego poziomu rozdzielczości, a jeżeli w badaniach zachodzi taka potrzeba to drugiego poziomu rozdzielczości itd. Skupienie współczynników falkowych każdego 4-elementowego szeregu i odpowiadających im wartości rzeczywistych oraz współczynników odwrotnej transformaty falkowej pozwala, poprzez uczenie sztucznej sieci neuronowej, wygenerować współczynniki transformaty falkowej dla ostatniego 4-elemntowego szeregu, czyli szeregu zawierającego prognozowaną wartość. Poprzez aplikację algorytmu odwrotnej transformaty falkowej otrzymuje się wartości szukanej prognozy.

4. Wyniki badania

Dane uwzględnione w badaniu są danymi rocznymi, zatem predykcja może być wykonywana tylko i wyłącznie z częstotliwością roczną. Z uwagi na liczebność zbioru wejściowego – 18 obserwacji – predykcję wykonano tylko na jeden i dwa okresy do przodu, gdyż wydaje się nieuzasadnione prognozowanie na tak małej próbie, bez szeregów skointegrowanych z szeregiem prognozowanym, na dłuższy okres. Predykcja miałaby sens przy uwzględnieniu dodatkowych czynników, wpływających na poziom zmiennej prognozowanej. Jednakże w badaniu takiego zagadnienia nie podjęto, gdyż celem było wyłonienie najlepszej metody generacji dodatkowych elementów szeregu w sensie minimalizacji błędu prognozy.

W zależności od zastosowanej metody rozszerzenia szeregów, uzyskujemy różne błędy predykcji. Jednakże jak pokazują dane zawarte w tab. 1 największym błędem są obarczone prognozy z zastosowanymi metodami rozszerzeniem od 1 do 4. Najmniejszym błędem jest obarczona predykcja z aplikacją metody wielomianowej.

Tabela 1. Wyniki badania

Okres predykcji	**Metoda rozszerzenia**				
	I	II	III	IV	WIELOMIANOWA
1 rok	2,11%	1,95%	2,05%	0,9%	0,44%
2 lata	2,41%	2,25%	2,45%	1,2%	0,45%

Podsumowanie

Analiza przedstawiona w tym artykule miała na celu wskazanie najlepszego rozszerzenia szeregu pod względem wpływu na wynik predykcji z wykorzystaniem analizy falkowej. Przeprowadzone badania w sposób jednoznaczny pokazują, że najlepsze wyniki w sensie minimalizacji błędów predykcji uzyskuje się poprzez zastosowanie do generacji dodatkowych danych, metody wielomianowe.

W artykule nie dokonano porównania wyników predykcji z innymi modelami, ponieważ celem badania nie jest ocena i wybór najlepszego modelu predykcji, ale ocena oraz wybór najlepszej metody generacji dodatkowych danych w procesie transformaty falkowej.

Literatura

Barczak S. (2013), *Zastosowanie teorii szarych systemów do przewidywania przyszłych ofert składanych na aukcjach pierwszej ceny poprzez pryzmat modelu szarego GM(1,1)*, Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, nr 146.

Biernacki A. (2009), *Numerical Evaluation of the Random Walk Search Algorithm* [w:] *Man-Machine Interactions*, Springer, Berlin Heidelberg, s. 533-540.

Hadaś-Dyduch M. (2013), *Prognozowanie szeregów czasowych w oparciu o współczynniki transformaty falkowej, optymalizowane przez sztuczną sieć neuronową* [w:] A.S. Barczak (red.), *Metody matematyczne, ekonometryczne i komputerowe w finansach i ubezpieczeniach 2009*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, s. 59-69.

Hadaś-Dyduch M. (2015a), *Prognozy instrumentów finansowych generowane współczynnikami falkowymi z rozszerzeniem*, Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, nr 227, s. 5-15.

Hadaś-Dyduch M. (2016a), *Econometric-wavelet Prediction in Spatial Aspect* [w:] M. Papież, S. Śmiech (eds.), *The 10th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*, Foundation of the Cracow University of Economics, Cracow, s. 45-52.

Hadaś-Dyduch M. (2015b), *Prediction of Wavelets Analysis* [w:] *Financial management of Firms and Financial Institutions, Proceedings (Part I.) 10th International Scientific Conference*, VSB-Technical University of Ostrava, Faculty of Economics, Department of Finance, Ostrava, Czech Republic, s. 341-348.

Hadaś-Dyduch M. (2016b), *Wygładzanie falkowe jako kluczowy instrument w predykcji krótkookresowej/Alignment Waveletes as Main Instrument in the Short-Time Term Prediction*, Hradec Economic Days. Double-blind peer reviewed proceedings of the international scientific conference Hradec Economic Days 2016, University of Hradec Králové, Executive department, Faculty of Informatics and Management Department, s. 62-68.

Janiga-Ćmiel A. (2010), *Prognoza fluktuacji koniunktury gospodarczej Polski i wybranych krajów Unii Europejskiej w latach 2007-2020* [w:] J. Mika (red.), *Metody i modele analiz ilościowych w ekonomii i zarządzaniu*, cz. 2, s. 94-110.

Nievergelt Y. (1999), *Wavelets Made Simple*, Birkhauser, Boston, MA.

Przelaskowski A. (2002), *Falkowe metody kompresji danych obrazowych*, rozprawa habilitacyjna, Oficyna Wydawnicza, Warszawa, s. 226.

Przybylska-Mazur A. (2013), *Wybrane metody prognozowania wskaźnika inflacji* [w:] W. Szkutnik (red.), *Wybrane aspekty modelowania statystycznego i analiz zagadnień rynku kapitałowego oraz rynku pracy w koncepcji zarządzania ryzykiem*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice.

Przybylska-Mazur A. (2015), *Selected Tests Comparing the Accuracy of Inflation Rate Forecasts Constructed by Different Methods*, „Statistics in Transition", Vol. 15, No. 2, s. 299-308.

POLYNOMIAL GENERATION DATA WAVELET ANALYSIS

Summary: The aim of the study is to assess the impact of the proposed method for the generation of additional elements series on the accuracy of the forecast. Generated a number of additional elements are used to determine the coefficients of which are determined coefficients of wavelet transform on the first level of resolution wavelets. In order to assess the polynomial method of data extension made prediction series, presenting the unemployment rate of the euro area. The results obtained with the more trivial methods of generation of additional elements in the wavelet transform.

Keywords: wavelet, wavelet analysis, wavelet transform, prediction.