



Monika Hadaś-Dyduch

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

PREDYKCJA SZEREGÓW CZASOWYCH ALGORYTMEM UWZGLĘDNIAJĄCYM PRZESUWNE OKNO CZASOWE I PODZIAŁ JEDNOSTKOWY SZEREGÓW

Streszczenie: Celem artykułu jest przedstawienie autorskiego algorytmu do predykcji szeregów czasowych. Algorytm oparto na sztucznych sieciach neuronowych oraz analizie wielorozdzielczej. Jednakże główną cechą algorytmu, dającą dobrą jakość prognozy, jest podział wszystkich uwzględnionych w analizie szeregów na kilkuelementowe podszeregi oraz uzależnienie predykcji danego szeregu od innych szeregów ekonomicznych. Aplikację algorytmu przeprowadzono na szeregu prezentującym WIG. Prognozę WIG uzależniono od notowań indeksów Dow Jones, DAX, Nikkei, Hang Seng, z uwzględnieniem przesuwne okna czasowego. Wyznaczono, jako przykładową aplikację autorską, prognozę WIG na okres 10, 20 i 30 dni.

Słowa kluczowe: predykcja, analiza falkowa, analiza wielorozdzielcza, sztuczne sieci neuronowe, falka Daubechies, predykcja.

Wprowadzenie

Algorytm do predykcji szeregów czasowych prezentujących wskaźniki makroekonomiczne oraz indeksy giełdowe oparto na sztucznych sieciach neuronowych oraz analizie falkowej, falką Daubechies. Główną cechą algorytmu jest jednak podział analizowanych szeregów na kilkuelementowe podszeregi oraz uzależnienie predykcji danego szeregu od innych szeregów ekonomicznych z odpowiednim przesuwne oknem czasowym. Horyzont predykcji przyjęto jako decydujący parametr o przesunięciu czasowym danych szeregu prognozowanego.

Oparcie prognozy szeregu prezentującego indeks giełdowy bądź wskaźnik makroekonomiczny na innych indeksach giełdowych ma szerokie uzasadnienie. Należy wspomnieć, że na dokładność otrzymanych prognoz ma również wpływ zastosowanie dodatkowego rozszerzenia szeregu [zob. Hadaś, 2005; Hadaś-Dyduch, 2013, 2014a, 2014b].

1. Przegląd literatury

Predykcja szeregu czasowego z matematycznego punktu widzenia polega na wyznaczeniu jego warunkowej wartości oczekiwanej dla chwili wyprzedzającej bieżący czas o ustaloną liczbę obserwacji zwaną horyzontem predykcji. Wykorzystuje się do tego celu formuły matematyczne. Wśród modeli opartych na formułach matematycznych wyrażonych w sposób jawny można wymienić m.in. modele regresyjne parametryczne i modele w przestrzeni stanu. Natomiast wśród modeli opartych na formułach matematycznych wyrażonych w sposób niejawny można wypunktować m.in. estymatory nieparametryczne oraz predyktory neuronowe. Modele służące do predykcji szeregów czasowych można podzielić na jedno- i wielowymiarowe. W modelach jednowymiarowych szereg czasowy traktowany jest jako proces stochastyczny o nieznanym wejściu losowym. Wówczas predykcję wyznacza się na podstawie modeli sygnałowych typu ARMA Boxa-Jenkinsa lub modeli ekstrapolacyjnych. Właściwości szeregów jedno- i wielowymiarowych wraz z metodami ich identyfikacji są szeroko omówione w: [Box i Jenkins, 1976; Otnes i Enochson, 2006].

Wśród znaczących metod prognozowania można wymienić predykcje oparte na sztucznych sieciach neuronowych oraz analizie falkowej. Predyktory wieloczynnikowe oparte na wykorzystaniu sztucznych sieci neuronowych zostały opisane szeroko w: [Freisleben i Ripper, 1997; Palit i Propovic, 2005]. W ten sposób „(...) implementuje się modele o strukturze zbliżonej do ARMA, ale z wykorzystaniem niejawnych, nieliniowych przekształceń zmiennych objaśniających, z parametrami wyznaczanymi metodą uczenia na danych historycznych (...)” [Pełech-Pilichowski i Duda, 2008]. Natomiast „(...) analiza falkowa jest rodzajem analizy częstotliwościowej pozwalającym efektywnie badać zmienne w czasie charakterystyki spektralnej procesów. Chociaż nie jest ona techniką prognozy styczną *per se*, jej cechy wyróżniające, takie jak dekompozycja procesów według pasm częstości, dobre własności lokalizacyjne w czasie, efektywność obliczeniowa [jest] (...) użyteczna w prognozowaniu ekonomicznych szeregów czasowych, szczególnie szeregów charakteryzujących się niestacjonarnością,

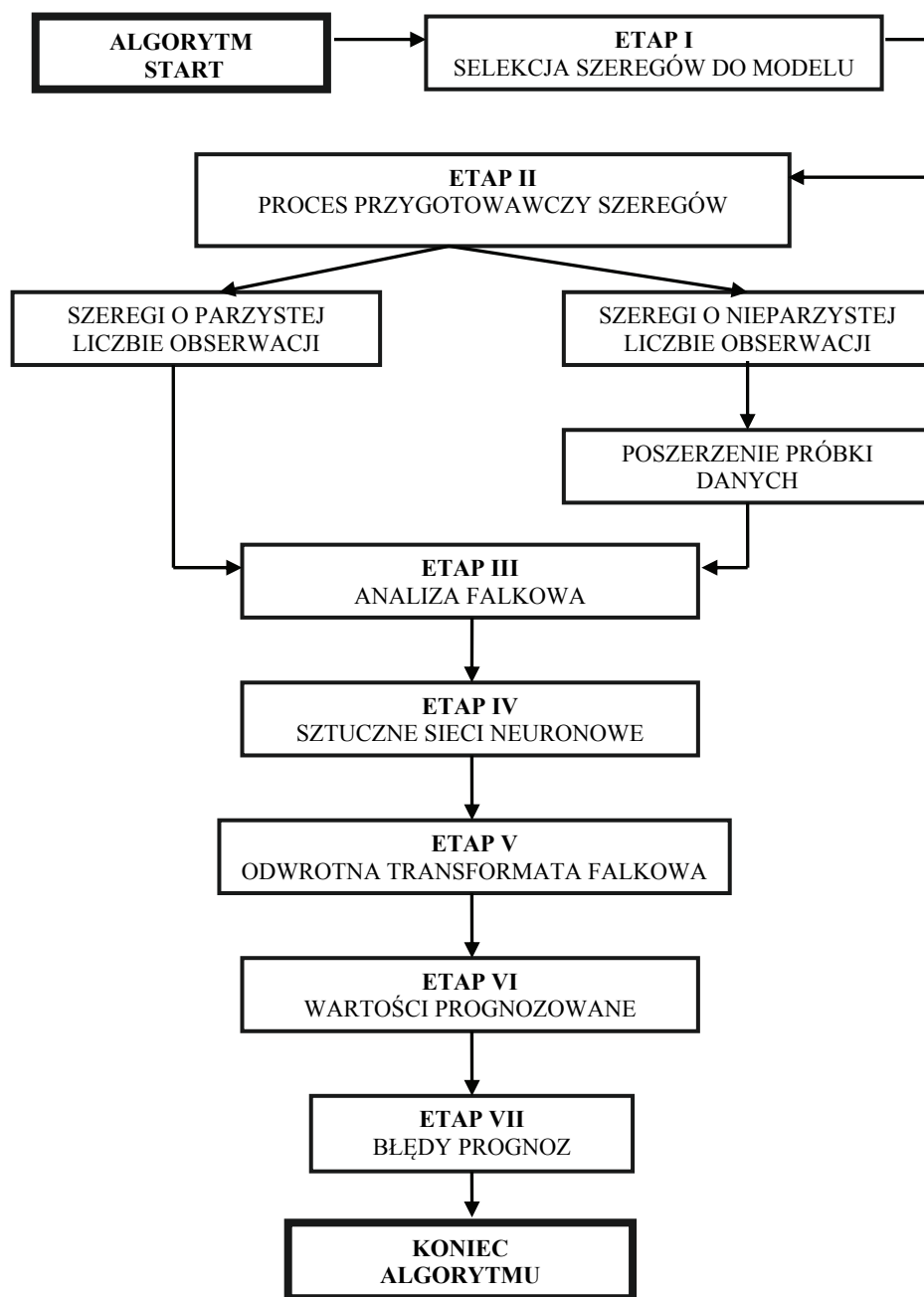
przejawiających krótkookresowe oscylacje o zmiennej amplitudzie, dla których ogniwa poprzedzające w łańcuchach przyczynowych zależą od skali czasu (horyzontu decyzyjnego)” [Bruzda, 2013].

2. Opis autorskiego modelu

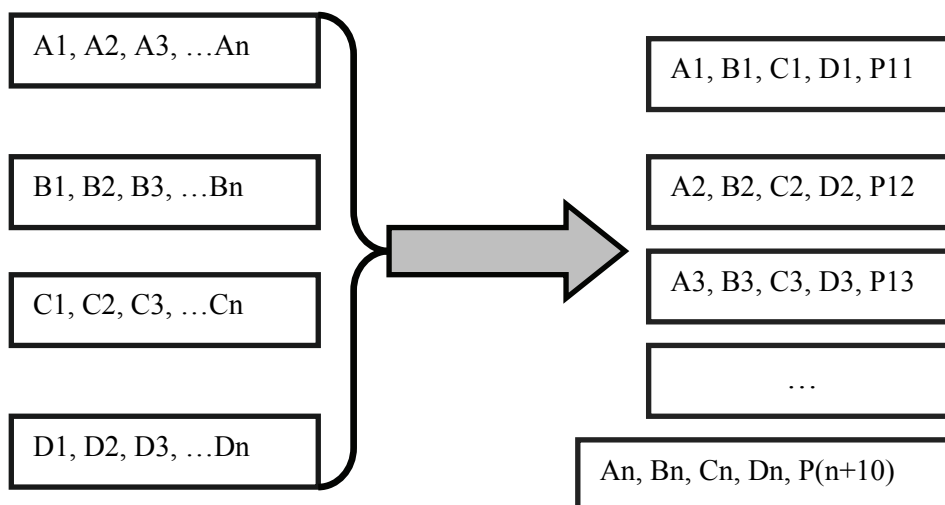
Zaproponowany do predykcji szeregów czasowych algorytm składa się z kilkunastu etapów, jednak kluczowych można wyróżnić siedem (rys. 1).

Pierwszy etap proponowanego algorytmu, właściwie każdego innego również, polega na selekcji szeregów czasowych do predykcji. Do badania należy wybrać szeregi spokrewnione z pewnym opóźnieniem czasowym z szeregiem prognozowanym i szeregi zależne w szerokim tego słowa znaczeniu od szeregu prognozowanego. Szeregi wybrane do badania mają znaczącą rolę w predykcji, ponieważ od nich uzależniona jest prognoza szeregu prognozowanego. Dlatego istotne na tym etapie są metody doboru szeregów spokrewnionych do badania z szeregiem prognozowanym.

Drugi etap (rys. 2) jest głównie etapem przygotowania wybranych szeregów czasowych przewidzianych do badania. Krok ten obejmuje, oprócz standardowych procedur, również przesunięcie szeregu prognozowanego (nazwanego również w dalszej części szeregiem bazowym) o horyzont prognozy w stosunku do pozostałych szeregów uwzględnionych w badaniu. Innymi słowy, aplikujemy w ramach tego etapu tzw. „przesuwne okno czasowe”. Zakładamy, że szeregi spokrewnione z szeregiem bazowym oddziałują na niego z pewnym opóźnieniem czasowym. To opóźnienie czasowe w najprostszej wersji to horyzont predykcji. Zatem obserwacja numer 1 każdego szeregu spokrewnionego jest przypisana do 31. obserwacji szeregu bazowego w przypadku predykcji 30-dniowej. Natomiast w przypadku predykcji o horyzoncie 10 dni, obserwacja pierwsza każdego szeregu spokrewnionego z szeregiem bazowym jest przypisana do 11. obserwacji szeregu bazowego, a każda obserwacja druga – do 12. obserwacji szeregu bazowego itd. (graf 1). Zatem znając w przypadku predykcji 10-dniowej ostatnie dziesięć obserwacji każdego z szeregów spokrewnionych z szeregiem bazowym, możemy na podstawie proponowanego algorytmu wyznaczyć prognozę na dziesięć dni do przodu dla szeregu prognozowanego, przy założeniu, że szeregi spokrewnione oddziałują na szereg bazowy właśnie z 10-dniowym opóźnieniem.



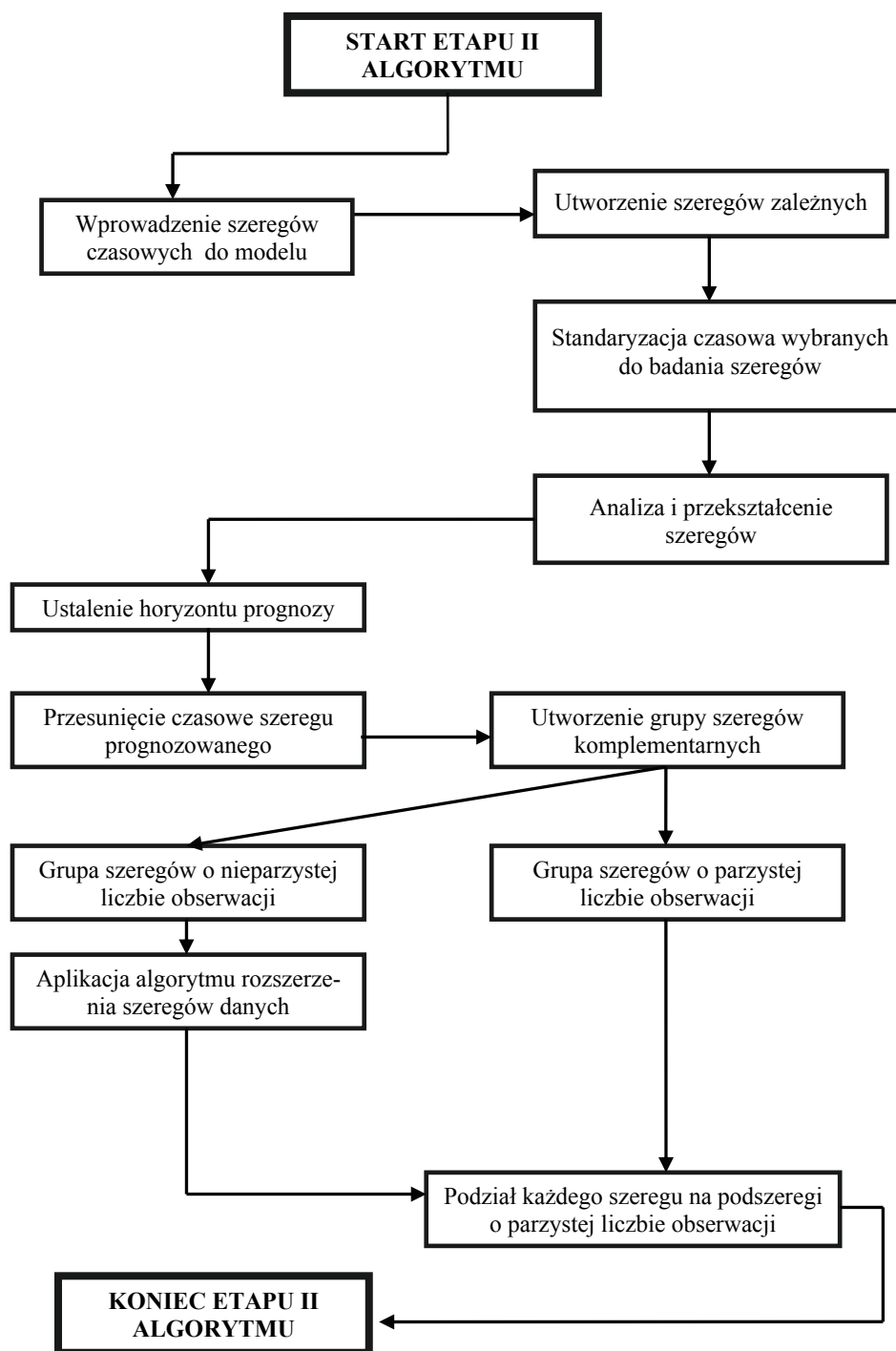
Rys. 1. Uproszczony schemat modelu predykcji



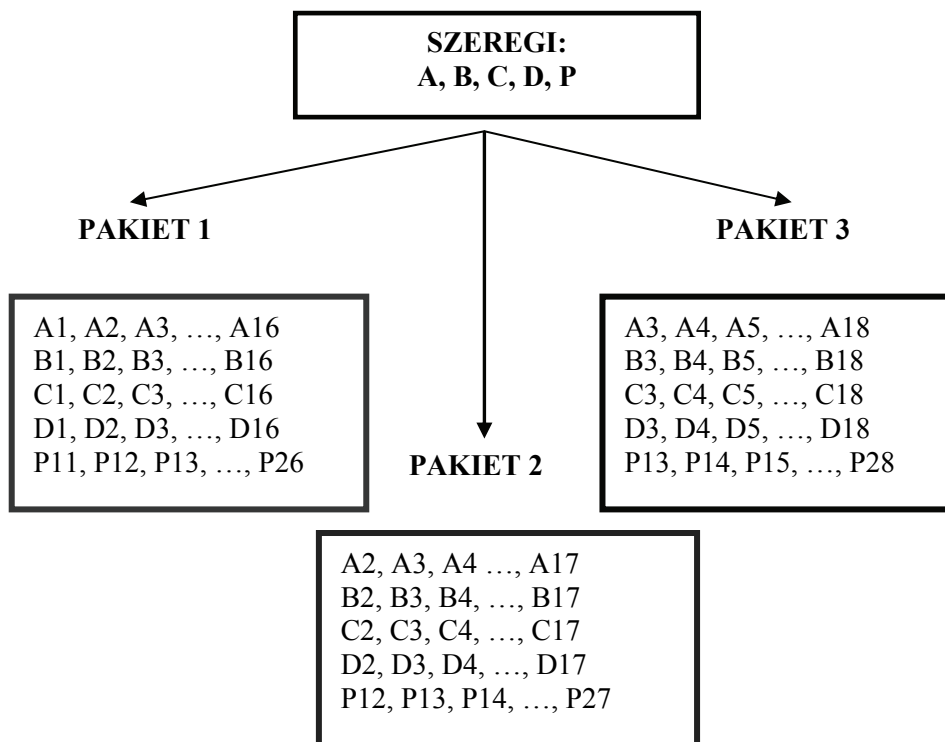
Graf 1. Prezentacja przesunięcia czasowego dla horyzontu predykcji 10-dniowej, gdzie szeregi A, B, C, D to szeregi spokrewnione z szeregiem bazowym, a szereg P to szereg prognozowany, czyli bazowy (rozpatrujemy 4 szeregi spokrewnione z szeregiem bazowym, jeden szereg prognozowany)

Ponadto, celem uzyskania dokładniejszych wyników (tak pokazują wcześniejsze badania), każdy szereg uwzględniony w badaniu (zarówno bazowy, jak i spokrewniony z szeregiem prognozowanym) dzielimy na podszeregi, tzw. próbki o parzystej liczbie obserwacji, będące wielokrotnością liczby dwa. Tworzymy więc pakiety złożone z szeregów np. 16-elementowych bądź 32-elementowych (graf 2).

Każdy utworzony tzw. pakiet zawiera obserwacje szeregów spokrewnionych oraz odpowiadające im obserwacje szeregu bazowego, przesuniętego o odpowiedni horyzont predykcji (graf 1).



Rys. 2. Schemat drugiego etapu algorytmu



Graf 2. Przykładowe trzy pierwsze pakiety, tj. trzy pierwsze grupy podszeręgów, przy podziale szeregów uwzględnionych w badaniu na podszeregi 16-elementowe i 10-dniowym horyzoncie predykcji (rozpatrujemy 4 szeregi spokrewnione z szeregiem bazowym, jeden szereg prognozowany, gdzie szeregi A, B, C, D to szeregi spokrewnione z szeregiem bazowym, a szereg P to szereg prognozowany, czyli bazowy)

Kolejny etap algorytmu ma na celu wygenerowanie współczynników falkowych, falki Daubechies. Odpowiednio przygotowane szeregi, a właściwie pakiety podszeręgów szeregów wybranych do badania, podlegają operacji transformacji falkowej. W ramach tej operacji dokonujemy również rozszerzenia podszeręgów zgrupowanych w pakietach jedną z metod dostępnych w programie Matlab. Wybór metody rozszerzenia szeregów ma wpływ na dokładność predykcji, czego dowodzą przeprowadzone w tym zakresie badania.

Z uwagi na fakt, że szereg bazowy jest przesunięty o horyzont predykcji w stosunku do szeregów spokrewnionych, ostatni pakiet podszeręgów zawiera tylko zestaw szeregów spokrewnionych. Przykładowo, dla predykcji o horyzoncie 10 dni i podszeręgach długości 16, przy liczebności 100 obserwacji każdego

szeregu spokrewnionego (rozpatrujemy 4 szeregi spokrewnione), ostatni pakiet podszeregów ma postać:

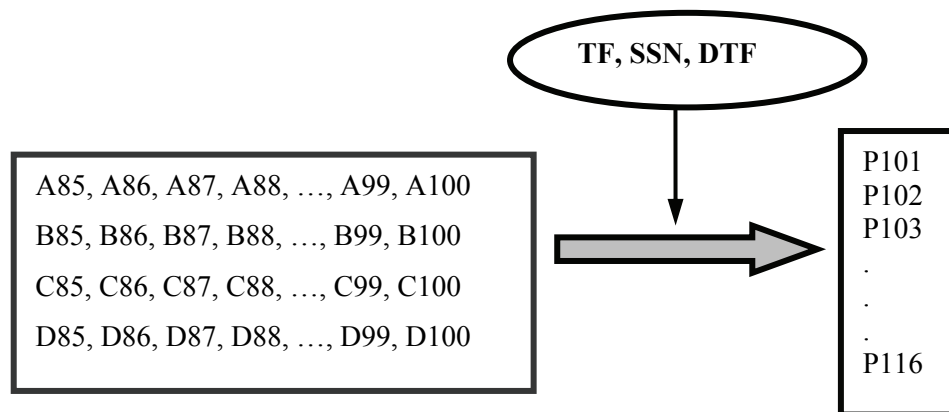
A85, A86, A87, A88, ..., A99, A100

B85, B86, B87, B88, ..., B99, B100

C85, C86, C87, C88, ..., C99, C100

D85, D86, D87, D88, ..., D99, D100

Właściwie ostatni pakiet współczynników falkowych służy do wyznaczenia szukanych wartości szeregu bazowego:



Objaśnienia:

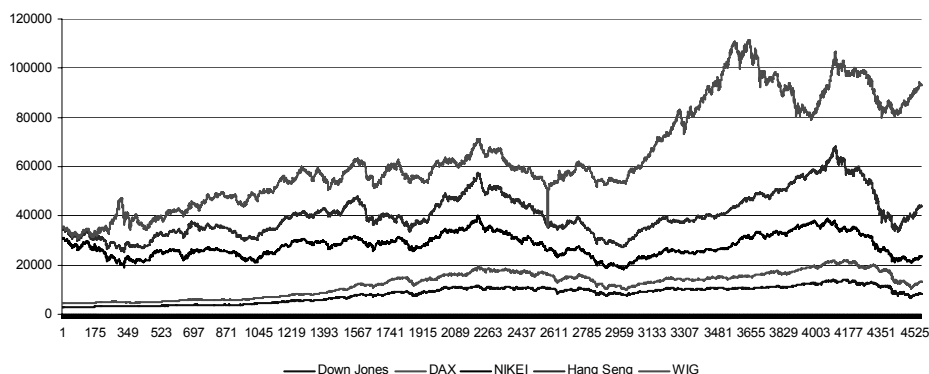
TF – transformata falkowa, *SSN* – sztuczne sieci neuronowe, *DTF* – odwrotna transformata Falkowa.

Graf 3. Uproszczona graficzna prezentacja wykonywanego przekształcenia na szeregach

W wyniku transformaty falkowej otrzymuje się dla każdego pakietu zestaw współczynników falkowych na różnych poziomach rozdzielczości, które są niezbędne w procesie uczenia sztucznej sieci neuronowej, gdyż pakiety współczynników falkowych, zawierające w sobie zarówno współczynniki podszeregów szeregów spokrewnionych z szeregiem bazowym, jak i współczynniki podszeregów szeregu bazowego, służą jako zbiór uczący do sztucznej sieci neuronowej. Natomiast pakiet zawierający tylko współczynniki falkowe szeregów spokrewnionych z szeregiem prognozowanym służy do wyznaczenia współczynników falkowych na jeden okres do przodu, w sensie horyzontu predykcji.

3. Aplikacja

Aplikację opisanego powyżej algorytmu wykonano, przyjmując jako szereg prognozowany – szereg WIG. Jako szeregi spokrewnione z szeregiem bazowym, tj. z szeregiem WIG, przyjęto: Dow Jones, DAX, Nikkei, Hang Seng.



Rys. 3. Notowania indeksów Dow Jones, DAX, Nikkei, Hang Seng i WIG (23.04.1991-16.09.2011)

Szeregi indeksów Dow Jones, DAX, Nikkei, Hang Seng i WIG, uwzględnione w badaniach, są notowaniami dziennymi z okresu 23.04.1991-16.09.2011. Szeregi nie są równoliczne, zatem dokonano ich standaryzacji czasowej. Następnie, postępując zgodnie z zaproponowanym w punkcie 2 algorytmem, wygenerowano odpowiednie prognozy szeregu bazowego, tj. szeregu WIG.

W badaniu podział na zbiory uczący i testowy był rozpatrywany procentowo z uwzględnieniem długości oczekiwanej prognozy. Przyjęto trzy strategie (przedstawione w tab. 1) danych wprowadzonych na wejście do sztucznej sieci neuronowej.

Tabela 1. Strategie

Wyszczególnienie	Horyzont predykcji [w dniach]		
	10	20	30
Zbiór uczący	87%	90%	92%
Zbiór testowy	13%	10%	8%

Wartości prognozy szeregu WIG wyznaczono odpowiednio na okres 10, 20 i 30 dni. Wartości predykcji na określony horyzont predykcji otrzymujemy w formie wektora. Średni bezwzględny błąd procentowy poszczególnych prognoz był mniejszy od 1% , co prezentuje szczegółowo tab. 2.

Tabela 2. Wyniki badania

Wyszczególnienie	Horyzont predykcji [w dniach]		
	10	20	30
Średni bezwzględny błąd procentowy	0,092%	0,94%	0,13%

Podobne badanie do opisanego powyżej przeprowadzono dla wskaźników makroekonomicznych [Hadaś-Dyduch, 2015]. Jako szereg bazowy wybrano stopę bezrobocia w Polsce. Jako szeregi spokrewnione z szeregiem bazowym przyjęto: PKB i stopę bezrobocia Czech, Danii, Niemiec, Grecji i Austrii. Każdy szereg zawierał na wejściu do modelu po 28 obserwacji. Predykcję wykonano na okres 12 miesięcy i 6 miesięcy. Średni bezwzględny błąd procentowy był następujący:

- 1% dla prognozy na stopę bezrobocia na jeden rok;
- 0,9% dla prognozy na stopę bezrobocia na pół roku.

Podsumowanie

W artykule przedstawiono całkowicie autorską metodę prognozowania szeregów czasowych, opartą na sztucznych sieciach neuronowych oraz transformacie falkowej – falka Daubechies, z uwzględnieniem przesuwnego okna czasowego oraz podziału analizowanych szeregów na podszeregi n-elementowe. Zaprezentowane wyniki pokazują, że zastosowanie modelu opartego na analizie falkowej i sztucznych sieciach neuronowych jest uzasadnione w świetle analizowanych danych.

Osiągnięte wyniki pokazują, że zaproponowany algorytm może służyć do długookresowej predykcji, ponieważ uzyskane błędy prognoz są stosunkowo małe. Można stwierdzić, że przedstawiony model może być skutecznym narzędziem prognozowania wskaźników makroekonomicznych, których przewidywanie jest bardzo trudne ze względu na złożoność mechanizmu tego rynku, a zwłaszcza czynników oddziałujących na ten rynek.

Literatura

- Box G.E.P., Jenkins G.M., (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Bruzda, J. (2013), *Prognozowanie metodą wyrównywania falkowego*, Acta Universitatis Nicolai Copernici „Zarządzanie”, 39, s. 77-95.
- Calderon A., (1964), *Intermediate Space and Interpolation, the Complex Method*, „Studia Mathematica”, 24, s. 113-190.
- Freisleben B., Ripper K. (1997), *Volatility Estimation with a Neural Network*, IEEE Computational Intelligence for Financial Engineering.
- Hadaś M. (2005), *Falki w kontekście zastosowań ekonomicznych* [w:] T. Trzaskalik (red.), *Zarządzanie-Finanse-Ekonomia*, Warsztaty doktorskie’05, Prace naukowe AE, Wydawnictwo AE, Katowice, s. 107-119.

- Hadaś-Dyduch M. (2013), *Prognozowanie wskaźników makroekonomicznych z uwzględnieniem transformaty falkowej na przykładzie wskaźnika inflacji*, Zeszyty Naukowe Wyższej Szkoły Bankowej we Wrocławiu, 2, s. 175-186.
- Hadaś-Dyduch M. (2014a), *Wykorzystanie transformaty falkowej w analizie i predykcji wskaźników makroekonomicznych*, „Studia Ekonomiczne”, nr 187, s. 124-135.
- Hadaś-Dyduch M. (2014b), *Wpływ rozszerzenia próbki przy generowaniu współczynników falkowych szeregu na trafność prognozy*, „Ekonometria”, Vol. 4, Iss. 46, s. 62-71.
- Hadaś-Dyduch M. (2015), *Polish Macroeconomic Indicators Correlated-prediction with Indicators of Selected Countries* [w:] M. Papież i S. Śmiech (eds.), *Proceedings of the 9th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Conference Proceedings, Foundation of the Cracow University of Economics, Cracow.
- Otnes R.K., Enochson L. (2006), *Analiza numeryczna szeregów czasowych*, WNT, Warszawa.
- Palit A.K., Propovic D. (2005), *Computational Intelligence in Times Series Forecasting Theory and Engineering Applications*, Springer-Verlag, London.
- Pelech-Pilichowski T., Duda J.T. (2008), *Adaptacyjne algorytmy detekcji zdarzeń w szeregach czasowych*, Rozprawa doktorska, WEAliE AGH, Kraków.

TIME SERIES PREDICTION ALGORITHM CONTAINING TIME WINDOW AND DIVISION UNIT SERIES

Summary: This article presents the author's algorithm for time series prediction. The algorithm based on artificial neural networks and multiresolution analysis. However, the main feature of the algorithm, giving a good quality of forecasts, it is all included in the division series analysis on several elements under-series and dependence prediction of a series of other economic ranks. The application of the algorithm was performed on a series of presenting WIG. The forecast WIG made dependent on trading the Dow Jones, DAX, Nikkei, Hang Seng taking into account the shift of the time window. They were, as a sample application copyright forecast WIG for a period of 10, 20 and 30 days.

Keywords: prediction, wavelet analysis, multi-resolution analysis, artificial neural networks, wavelet Daubechies, prediction.