Nonlinear Alleviation of Edge Effects in the Context of Minimizing Prediction Errors

Monika Hadaś-Dyduch[1]

[1] Department of Mathematical and Statistical Methods in Economics, University of Economics in Katowice, Katowice, Poland

Correspondence: Monika Hadaś-Dyduch, Department of Mathematical and Statistical Methods in Economics, University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland. E-mail: monika.dyduch@ue.katowiice.pl

Abstract

The aim of this article is to describe so-called "edge effects" in the context of wavelet analysis. The problem of "edge effects" is displayed in cases where the filter length is greater than 2. This is due to the fact that the calculation of the wavelet coefficients for the development of the last signal of finite elements, the filter - should theoretically move beyond the signal. The article describes different ways to solve this problem using an authored approach. One of the ways presented in the article is an innovative approach in terms of "edge effects". The author's proposal is based on the nonlinear function of the trend after the division of the series into smaller units. The results obtained show that in comparison with other methods, the author's method reduces errors. In this article, the Daubechies wavelet was used for the study. The Daubechies wavelets are a family of orthogonal wavelets, characterized by a maximum number of vanishing moments for a given support.

Keywords: wavelets, multi-resolution analysis, wavelet transform, edge effects

1. Introduction

Wavelet analysis is becoming more and more applicable every year. It is used, among others, to denoise, compress data, recognize features, detect system non-linearity and so on (Daubechies, 1992; Meyer, 1998; Mallat, 2008; Luisier, Blu, & Unser, 2007; Melgani & Bazi, 2008; Addison, 2017). Due to the growing application possibilities and the expanding base of the wavelet family, it is becoming an ever more important issue.

Waves are functions that oscillate over time, in space, or both. Wavelets are small waves of energy concentrated in a relatively short period of time. Wavelet analysis consists of decomposing the signal using wavelets obtained through translation and dilation of the mother wavelet. The wavelets are particularly useful for time-frequency analyzes of non-stationary series.

Wavelet transformation is a transformation similar to the Fourier transform. Both transformations are based on the use of the scalar product of the examined signals (t) and the remaining part, called the "nucleus of transformation." The main difference between these transformations is the nucleus (Istas & Lang, 1997; Kahane & Lemarié, 1995).

In the case of wavelet transformation, the kernel is a wavelet - a special function limited by certain requirements that must be met in order to be used for so-called multiresolution analysis (eg. it must have a scaling function). There are infinitely many such functions, so there are infinitely many wavelet transformations.

Wavelet transformation is suitable for the analysis of non-stationary signals because it provides time-frequency information. Often, it is used interchangeably with the Fourier transformation. The reason for this is that as a result of the wavelet transformation, we obtain information about the frequencies of only the individual spectral components of the signal with finite precision. This is because the transformation kernel (or function) does not represent an infinitely narrow frequency range, but a frequency range that is inversely proportional to the duration of the wavelet. "The wavelet transform is calculated as shifting the wavelet function in time along the input signal and calculating the convolution of them. In most practical applications, the signals of interest have finite support. As the wavelet gets closer to the edge of the signal, computing the convolution requires non-existent values beyond the boundary. This creates boundary effects caused by incomplete information in the

boundary regions. Thus, the results of wavelet transform in these boundary effects regions have questionable accuracy." (Su, Liu, & Li, 2012) See also (Queiroz, 1992; Cohen & Daubechies 1993).

Of course we can ignore inaccurate results at the edges, thus accepting the loss of data by cutting them down. However, this is a risky approach, because the data around the edges may contain relevant information. Therefore, a more favorable solution to the situation is to extend a series of data before processing it (Mallat, 2008; Strang & Nguyen, 1998; Jiménez & Prelcic, 2004).

2. Border Effect

The edge effect in spectral estimation is well known. It hampers interpretation of the spectral estimates, and poses problems for situations where, for example, low frequency signals are to be analyzed from relatively short time serie, or real-time applications where the parts of data that are close to the edges, in particular the newest observations, are critical. Classical numerical methods have been devised to detal with the edge effect, for example data windowing. While these methods are often effective, there are always trade-offs in the form of a certain loss of information. The way the edge effect appears in the time-frequency wavelet spectrum is as follows. If one plots the time-frequency wavelet spectrum with the horizontal axis indicating time and the vertical axis indicating frequency, then only the spectral values within a U-shaped domain are not affected by the edge effect (...)" (Zheng, Chao, Zhou, & Yu, 2000). For signal processing with wavelet transform, a pair of filters forming the so-called analysis filter bank is used. In the case of filters whose length is greater than two, there is the problem of edge effects. This problem results from the fact that for the calculation of the wavelet coefficients of the development wavelets for the latest elements of the finite signal, the filter should move beyond the signal. This is not possible. Therefore various methods of signal extension are proposed (see chapter 3). The filtering part of the reconstruction process is also subject to some discussion as it is the choice of filters that is crucial in achieving perfect reconstruction of the original signal. The downsampling of the signal components performed during the decomposition phase introduces a distortion called aliasing. It turns out that by carefully choosing filters for the decomposition and reconstruction phases, which are closely related (but not identical), we can "cancel out" the effects of aliasing. The low- and high-pass decomposition filters (L and H), together with their associated reconstruction filters (L' and H'), form a system of what is called quadrature mirror filters (see more in: Cohen, Daubechies, & Vial, 1993).

3. Extending of the Time Series

Various methods of extending the finite series are given in the literature (Hadaś-Dyduch, 2016b; Ferretti, Rizzo, 2000; Cohen, Daubechies, & Jawerth, 1993; Strang & Nguyen, 1996). Each method has its advantages and disadvantages. The degree of emphasizing the advantages and disadvantages of the extension methods depends on many factors, one of them being the specificity of a research series.

In the literature are suggests the following extension method the series (Hadaś-Dyduch, 2014b, 2017; Brislawn, 1996; Cohen, Daubechies, & Jawerth, 1993; Strang & Nguyen, 1996; Nason & Silverman, 1995):

• Zero-padding - this method is used in the version of the DWT given in the previous sections and assumes that the signal is zero outside the original support. The disadvantage of zero-padding is that discontinuities are artificially created at the border.

• Symmetrization - this method assumes that signals or images can be recovered outside their original support by symmetric boundary value replication. Symmetrization has the disadvantage of artificially creating discontinuities of the first derivative at the border, but this method works well in general for images.

• Smooth padding of order 1 - this method assumes that signals or images can be recovered outside their original support by a simple first-order derivative extrapolation: padding using a linear extension fit to the first two and last two values. Smooth padding works well in general for smooth signals.

• Smooth padding of order 0 - this method assumes that signals or images can be recovered outside their original support by a simple constant extrapolation. For a signal extension this is the repetition of the first value on the left and last value on the right.

• Periodic-padding (1) - this method assumes that signals or images can be recovered outside their original support by periodic extension.

• Periodic-padding (2) - if the signal length is odd, the signal is first extended by adding an extra-sample equal to the last value on the right. Then a minimal periodic extension is performed on each side. The same kind of   rule   exists   for   images   (Matlab; *https://ch.mathworks.com/help/wavelet/ug/dealing-with-border-distortion.html?searchHighlight=extensions%20*

dwt&s_tid=doc_srchtitle).

In this article, as a method to mitigate the edge effect, the following approach is suggested:

1) Division of a series into sub-series. In the case of a series including seasonal variations, division of the main series according to seasonality is preferred.

2) Estimating the non-linear trend function of each sub-series. Choose a line with the best fit from among the estimated trend lines .

3) Selection of the dominant trend function among the estimated trend functions for the sub-series.

4) The general form of the new series.

5) Determination of the theoretical values based on the selected function.

6) Save the new series:

   i.
   $$\underbrace{p_{2^n-1}\cdots, p_0}_{extension},\ \underbrace{p_0, p_1, p_2, \cdots, p_{2^n-2}, p_{2^n-1}}_{series},\ \underbrace{p_{2^n-1}, \cdots, p_0}_{extension}$$

7) Estimation of wavelet coefficients.

8) Estimation of the approximating function.

9) In the case of a small sample there is no need to divide the sample into smaller units. In this case, the next stage is the designation of the trend line.

4. Empirical Research

Details of the procedure for the proposed method for alleviating edge effects will be presented as applied to a randomly chosen series, shown in Figure 1.
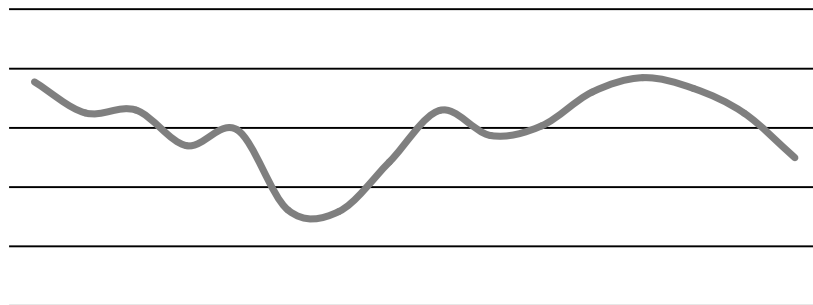

Figure 1. Data for analysis

Source: Own elaboration.

1) The series shown in Fig. 1 is not a very large sample. Thus, there is no need to divide the series into sub-series.

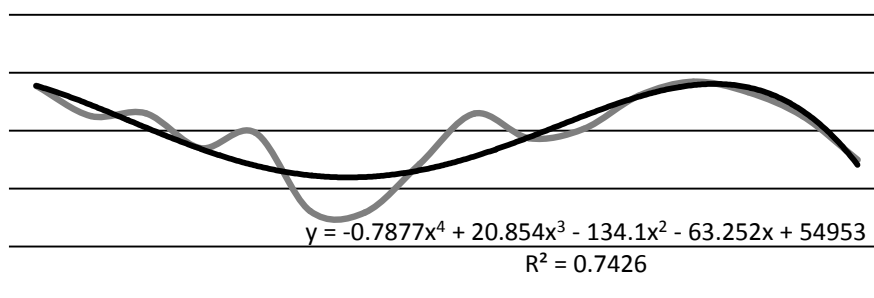2) The best approximating function for the series shown in Fig. 1 is a polynomial fourth degree function (Fig. 2).

$y = -0.7877x^4 + 20.854x^3 - 134.1x^2 - 63.252x + 54953$

$R^2 = 0.7426$

Figure 2. Time series and polynomial trend function

Source: Own.

$$\underbrace{\hat{f}_{2^n-1}\cdots,\hat{f}_0}_{extension}, \underbrace{p_0,p_1,p_2,\cdots,p_{2^n-2},p_{2^n-1}}_{primary\ series}, \underbrace{\hat{f}_{2^n-1},\cdots,\hat{f}_0}_{extension}$$

$$\underbrace{\hphantom{\hat{f}_{2^n-1}\cdots,\hat{f}_0, p_0,p_1,p_2,\cdots,p_{2^n-2},p_{2^n-1}, \hat{f}_{2^n-1},\cdots,\hat{f}_0}}_{new\ series}$$

3) New series:

4) Determination of theoretical values based on the estimated trend function. The theoretical values are determined consecutively for *t=0,1,2,3,4,...,N-1*. We obtain:

t	0	1	2	...	N-1
data	54774,6	54252,21	54300,47	...	53496,38
theoretical values	54953	54775,71	54444,32	...	54336,66

5) New series (p_r) has the form:

r	-N	-N-1	...	-1	0	1	...	N-1	N	...	2N-1
p_r	54953	54775,71	...	54336,66	54774,6	54252,21	...	53496,38	54953	...	54336,66

6) Determine the coefficients of wavelet, Daubechies wavelets (see more about Daubechies wavelet and Daubechies wavelet coefficients in (Cohen, Daubechies, & Vial, 1993; Daubechies, 1992; Hadaś-Dyduch 2014a, 2015a, 2015b, 2015c, 2016a; Hadaś-Dyduch & Hadaś, 2017)) from:

$$a_k = \sum_{r=k+0}^{k+3} \varphi(r-k)p_r, \quad k \in \{0,1,2,\ldots,2^n-1\}.$$

where:

$$\varphi(r) = \frac{1+\sqrt{3}}{4}\varphi(2r) + \frac{3+\sqrt{3}}{4}\varphi(2r-1) + \frac{3-\sqrt{3}}{4}\varphi(2r-2) + \frac{1-\sqrt{3}}{4}\varphi(2r-3)$$

wherein:

➢ $\sum_{k \in Z} \varphi(k) = 1$,

➢ $\varphi : D \to R$

➢ $\varphi(r) = 0$ *for* $r \leq 0 \vee r \geq 3$.

$$D_j = \{k2^j : k \in Z\}, \quad D = \bigcup_{j \in Z} D_j = \bigcup_{j=0}^{\infty} D_j.$$

Daubechies is a wavelet with a compact carrier. Changing the row of wavelets affects the change in coefficients describing it. From a practical point of view, it is not always the best solution to use higher order wavelets as the increase in the order implies that the system needs more calculations. Their most important feature is that the lower- and upper-pass filters have a finite length. Daubechies are available for every even filter length. The Haar wavelet is the simplest element of this family, it has the shortest filters, but it is discontinuous. As the length of the filters increases, the smoothness of the wavelets increases.

7) Save the wavelet function - approximates, according to the formula:

$$\tilde{f}(r) = a_{-2}\varphi(r+2) + a_{-1}\varphi(r+1) + \ldots + a_{2^n-1}\varphi(r-[2^n-1])$$

Therefore, we have:

$$\tilde{f}(r) = 54176,3594\varphi(r+2) + 54965,808\varphi(r+1) + \ldots + 54520,4293\varphi(r-[2^n-1])$$
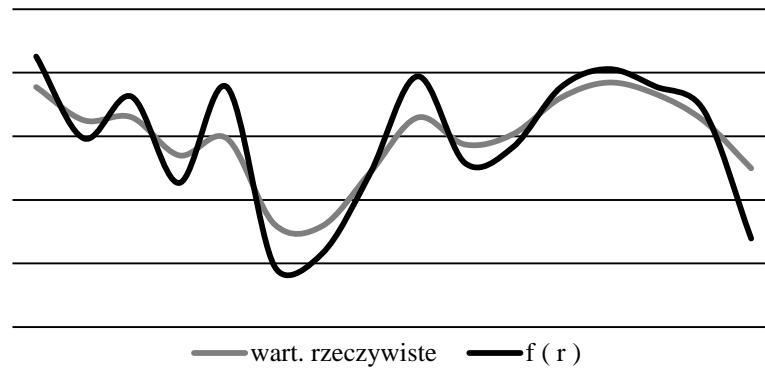
Figure 3. Time series and wavelet function approximating

Source: Own.

5. Result of Empirical Research

Expired error forecasts based on the author's methods are as follows: ME: -31.8337; MAE: 397.2213; MAPE: 0.7359%; RMSPE: 0.8988%. The coefficient of determination is 78.1053%.

For comparison, using the Method 1 extension, the coefficient of determination is 2%, and the extinct forecast errors are the following: ME: 3383,4681; MAE: 3674,5250; MAPE: 6,8072%; RMSPE: 17,7102%.

Using the authored algorithm for other time series as shown in Fig. 5, 6, 7, we obtain results which confirm that the author's proposed method is correct (see Tab. 1, 2).
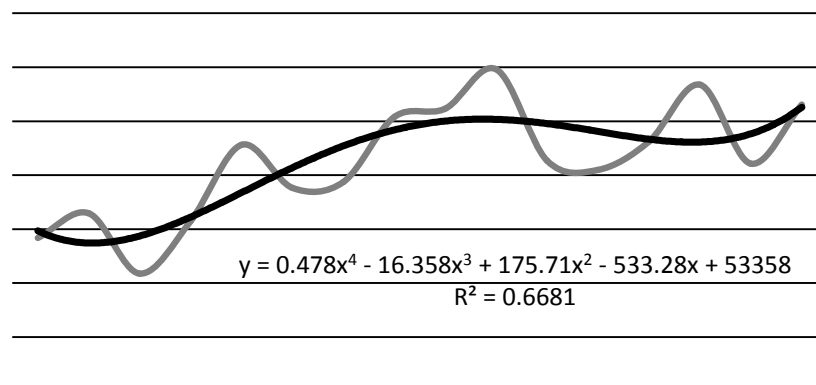

$$y = 0.478x^4 - 16.358x^3 + 175.71x^2 - 533.28x + 53358$$
$$R^2 = 0.6681$$

Figure 4. Series 2 and trend function

Source: Own.

$$y = -1.0013x^4 + 35.994x^3 - 401.17x^2 + 1244.6x + 52821$$
$$R^2 = 0.7671$$
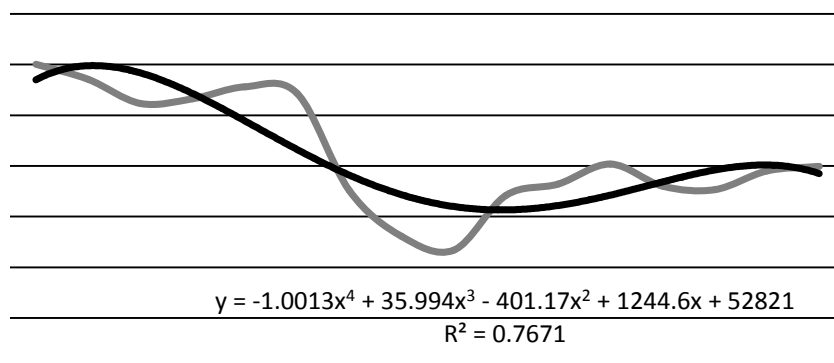
Figure 5. Series 3 and trend function

Source: Own.

$$y = -1.1323x^3 + 3.8294x^2 + 165.34x + 51770$$
$$R^2 = 0.9231$$
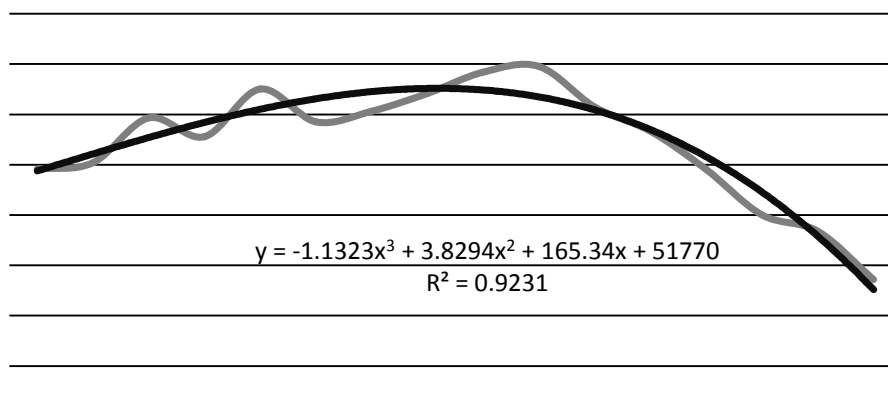
Figure 6. Series 4 and trend function

Source: Own.

$$y = 938.26\ln(x) + 50101$$
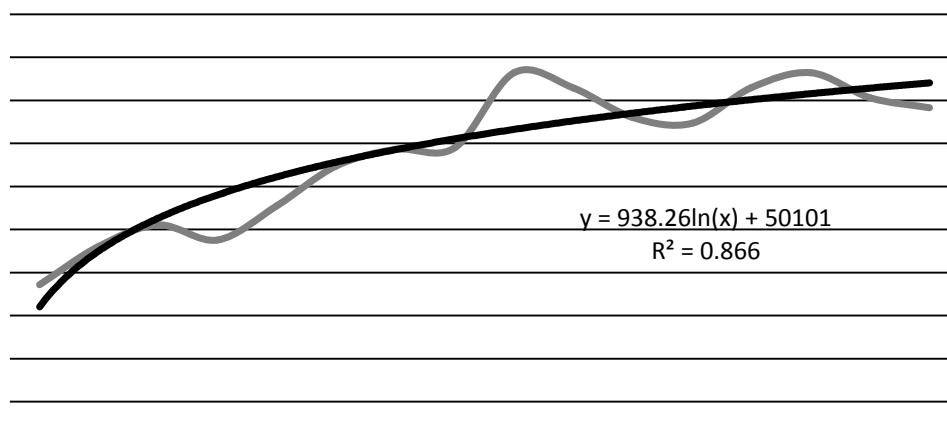$$R^2 = 0.866$$

Figure 7. Series 5 and trend function

Source: Own.

Table 1. Results approximation for the series 2, 3, 4, 5 with copyright extension.

Error	Series 2	Series 3	Series 4	Series 5
MAPE	0,7365%	0,7197%	0,3952%	0,6258%
RMSPE	0,8256%	0,9164%	0,5126%	0,9554%

Source: Elaboration based on my own calculations.

Table 2. Results approximation for the series 2, 3, 4, 5 with the extension of Method 1

Error	Series 2	Series 3	Series 4	Series 5
MAPE	6.8464%	6.9015%	6.3955%	6.4798%
RMSPE	17.717%	18.0174%	17.147%	17.417%

Source: Elaboration based on my own calculations.

In the research article, only the Daubechies wavelet was used. However, there are many families of orthogonal wavelet functions with different properties. The first is the family of Haar wavelets, historically the oldest since it was proposed as long ago as 1910 by the Hungarian mathematician Alfred Haar. An important advantage of this simplest wavelet system is the ease of demonstrating the idea of a discrete wavelet transform using it. The second of the selected wavelet families are the Daubechies wavelets (Daubechies, 1992; Hadaś-Dyduch, 2017; 2017b). The further two families - also constructed by Daubechies – are modifications of the basic family of Daubechies wavelets: the Coiflet family and the Symlet family. Other wavelets known in the literature are:

Battle-Lamair, Meyer, Morlet and Shannon. In practical applications, biorthogonal wavelets are also very popular.

6. Conclusion

The article presents an innovative approach in terms of "edge effects". The author's proposal is based on the nonlinear function of the trend after the division of series into smaller units. The results obtained show that in comparison with other methods the author's method reduces errors.

Further studies, may apply the proposed method for removing edge effects in wavelet analysis, for the analysis of the relationship between the different regions in the context of a multi-dimensional series, taking into account macroeconomic indicators and determining the best extension for related series. The study will use results from earlier studies, as well as some interesting methods used so far in other tests, for example (Akbalik & Tunay, 2016; Hadaś-Dyduch, 2014a, 2015b).

References

Addison, P. S. (2017). *The illustrated wavelet transform handbook: Introductory theory and applications in science, engineering, medicine and finance*. CRC press. https://doi.org/10.1887/0750306920

Akbalik, M., & Tunay, K. B. (2016). An Analysis of Ramadan Effect by GJR-GARCH Model: Case of Borsa Istanbul. *Oeconomia Copernicana*, *7*(4), 593-612. https://doi.org/10.12775/oec.2016.033

Brislawn, C. M. (1996). Classification of nonexpansive symmetric extension transforms for multirate filter banks. *Applied and Computational Harmonic Analysis*, *3*(4), 337-357. https://doi.org/10.1006/acha.1996.0026

Cohen, A., Daubechies, I., & Vial, P. (1993). Wavelets and fast wavelet transform on the interval. *Appl. Comput. Harmon. Anal*, *1*(1), 54-81. https://doi.org/10.1006/acha.1993.1005

Cohen, A., Daubechies, I., & Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, *1*(1), 54-81. https://doi.org/10.1006/acha.1993.1005

Cohen, A., Daubechies, I., Jawerth, B., & Vial, P. (1993). Multiresolution analysis, wavelets and fast algorithms on an interval. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, *316*(5), 417-421. https://doi.org/10.1006/acha.1993.1005

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for industrial and applied mathematics. https://doi.org/10.1137/1.9781611970104

De Queiroz, R. L. (1992, March). Subband processing of finite length signals without border distortions. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on* (Vol. 4, pp. 613-616). IEEE. https://doi.org/10.1109/icassp.1992.226323

Ferretti, M., & Rizzo, D. (2000). Handling borders in systolic architectures for the 1-D discrete wavelet transform for perfect reconstruction. *IEEE Transactions on Signal Processing*, *48*(5), 1365-1378. https://doi.org/10.1109/78.839983

Hadas-Dyduch, M. (2014a). Non-classical algorithm for time series prediction of the range of economic phenomena with regard to the interaction of financial market indicators. *Chinese Business Review*, *13*(4). https://doi.org/10.17265/1537-1506/2014.04.002

Hadas-Dyduch, M. (2016a). Wavelets as Basis Functions in the Adaptation's Methods: Author's Model for Forecasting Short-Term. *Chinese Business Review*, *15*(1), 8-17. https://doi.org/10.17265/1537-1506/2016.01.002

Hadaś – Dyduch, M. (2015a). Prediction of wavelets analysis. Financial management of Firms and Financial Institutions, *Proceedings (Part I.) 10th International Scientific Conference* (pp. 341-348). VSB-Technical University of Ostrava, Faculty of Economics, Department of Finance, Ostrava, Czech Republic. 7-8.09.2015.

Hadaś-Dyduch, M. (2014b). Wpływ rozszerzenia próbki przy generowaniu współczynników falkowych szeregu na trafność prognozy. *Ekonometria, 4*(46). https://doi.org/10.15611/ekt.2014.4.06

Hadaś-Dyduch, M. (2015b). Wavelets in the prediction of short-time series. *Mathematical Economics, 11*(18), 43-54. https://doi.org/10.15611/me.2015.11.04

Hadaś-Dyduch, M. (2016b). Wielomianowa generacja danych w analizie falkowej. *Studia Ekonomiczne*, *289*, 42-50.

Hadaś-Dyduch, M. (2017a). Artificial neural networks as one of the methods to alleviate edge effects in wavelet

analysis of macroeconomic indicators. In Proceeding of *8th International Scientific Conference "Analysis of International Relations 2017. Methods and Models of Regional Development* (pp. 21-27). Publishing House of the University of Economics in Katowice.

Hadaś-Dyduch, M. (2017b). Approximating Financial Time Series with Wavelets. *Argumenta Oeconomica Cracoviensia*, (16), 9-22. https://doi.org/10.15678/aoc.2017.1601

Hadaś-Dyduch, M., & Hadaś, A. (2017). Wavelet Daubechies as a tool supporting stock index prediction in the author's multi-component and multi-stage model. *IOSR Journal of Business and Management, 19*(10), 59-66.

Härdle, W., Kerkyacharian, G., Picard, D., & Tsybakov, A. (2012). *Wavelets, approximation, and statistical applications* (Vol. 129). Springer Science & Business Media. https://doi.org/10.1007/978-1-4612-2222-4_8

Istas, J., & Lang, G. (1997). Quadratic variations and estimation of the local Hölder index of a Gaussian process. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics* (Vol. 33, No. 4, pp. 407-436). No longer published by Elsevier. https://doi.org/10.1016/s0246-0203(97)80099-4

Jiménez, M. E. D., & Prelcic, N. G. (2004). Linear boundary extensions for finite length signals and paraunitary two-channel filterbanks. *IEEE transactions on signal processing*, *52*(11), 3213-3226. https://doi.org/10.1109/tsp.2004.836526

Kahane, J. P., Lemarié, P. G., & Lemarié-Rieusset, P. G. (1995). *Fourier series and wavelets* (Vol. 3). Routledge.

Luisier, F., Blu, T., & Unser, M. (2007). A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Transactions on Image Processing*, *16*(3), 593-606. https://doi.org/10.1109/tip.2007.891064

Mallat, S. (1999). *A wavelet tour of signal processing*. Academic press. https://doi.org/10.1016/b978-012466606-1/50009-x

Melgani, F., & Bazi, Y. (2008). Classification of electrocardiogram signals with support vector machines and particle swarm optimization. *IEEE transactions on information technology in biomedicine*, *12*(5), 667-677. https://doi.org/10.1109/titb.2008.923147

Meyer, Y. (1998). *Wavelets, vibrations, and scalings* (No. 9). American Mathematical Soc. https://doi.org/10.1090/crmm/009

Nason, G. P., & Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. *Lecture notes in Statistics* (p. 281). New York, Springer Verlag. https://doi.org/10.1007/978-1-4612-2544-7_17

Strang, G., & Nguyen, T. (1996). *Wavelets and Filter Banks* (Wellesley, MA: Wellesley-Cambridge). https://doi.org/10.1109/icassp.2002.5744025

Su, H., Liu, Q., & Li, J. (2012). Boundary effects reduction in wavelet transform for time-frequency analysis. *WSEAS Transactions on Signal Processing*, *8*(4), 169-179.

Zheng, D., Chao, B. F., Zhou, Y., & Yu, N. (2000). Improvement of edge effect of the wavelet time–frequency spectrum: Application to the length-of-day series. *Journal of Geodesy*, *74*(2), 249-254. https://doi.org/10.1007/s001900050283

Copyrights