ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 225, 2009

*Joanna Trzęsiok**

THE WAVELET TRANSFORM IN REGRESSION

Abstract

The wavelet transform was introduced in the 1980's and it was developed as an alternative to the short time Fourier transform. The wavelets theory is very popular in signal processing and pattern recognition and its applications are still growing.

This paper presents the wavelet transform in nonparametric regression. The use of wavelets in statistical applications was pioneered by D. Donoho and I. Johnstone. Here we discuss their methodology – wavelet shrinkage. The wavelet transform is compared with another nonparametric regression method – splines.

Key words: wavelets, wavelet transform, wavelet thresholding, nonparametric regression.

1. Introduction

The subject of the regression analysis is a set of observations:

$$U = \{(x_i, y_i) : i = 1, \dots, N\}$$

We look for a function f which describes the connection between the response Y and the predictor X:

$$Y = f(X) + \varepsilon \tag{1}$$

where ε is an error rate (noise).

There are many ideas for solving the problem. Among them there is a fast developing group of methods called nonparametric methods of regression. In these methods we do not have to make any assumptions about

* M. Sc., Department of Mathematics, Karol Adamiecki University of Economics, Katowice.

the distribution of a variable X. They produce models which are often better fitted to the data than functions obtained by the least squares method. Nonparametric models are more robust and resistant against outliers. The wavelet estimation methodology is one of the nonparametric methods of regression.

Wavelets are applied in a diverse set of fields, such as signal processing, pattern recognition, data compression, and numerical analysis. This methodology includes a wide range of tools, such as the wavelet transform, multiresolution analysis or wavelet decomposition.

Wavelet methods were introduced to statistics by D. Donoho and I. Johnstone in 1994. They developed the procedure based on the wavelet transform and thresholding for approximating an unknown function f.

2. Orthonormal basis function

In signal processing, a popular approach for approximating a univariate function is to use orthonormal basis functions $g_i(x)$, i.e. functions satisfying following condition:

$$\int g_i(x)\cdot g_j(x)\,dx = \begin{cases} 1, & \text{if } i=j \\ 0, & \text{if } i\neq j \end{cases} \tag{2}$$

We seek a function $\hat{f}$ in an additive form:

$$\hat{f}(x)=\sum_{j=1}^{m} w_j\cdot g_j(x) \tag{3}$$

Finding this function is equivalent to estimating values of parameters w_j. We get the values of parameters w_j by minimizing theoretical risk:

$$R(\mathbf{w})=\sigma^2+\int\left[f(x)-\sum_{j=1}^{m} w_j\cdot g_j(x)\right]^2 dx \tag{4}$$

where f is an unknown function in (1), and σ^2 denotes the noise variance. Solving this problem leads to:

$$w_j=\int f(x)\cdot g_j(x)\,dx \text{ for } j=1,\dots,m \tag{5}$$

We cannot evaluate (5), because the target function f is unknown. We estimate w_j using the given training set U:

$$\hat{w}_j = \frac{1}{N}\sum_{i=1}^{N} y_i \cdot g_j(x_i) \text{ for } j = 1, ..., m \tag{6}$$

The wavelet transform is defined as a decomposition of the function f using a specified set of orthonormal basis functions.

3. Wavelets

In this section we present the construction of the set of orthonormal basis functions – wavelet functions. We start with defining the *mother wavelet* as a function ψ satisfying the following conditions:

1) $\int \psi(x)\,dx = 0$,
2) $\int \psi^2(x)\,dx < \infty, \quad (\psi \in L^2(\mathbf{R}))$.

The examples of mother wavelet functions are:

- Haar wavelet – used only in theoretical examples and illustrations.

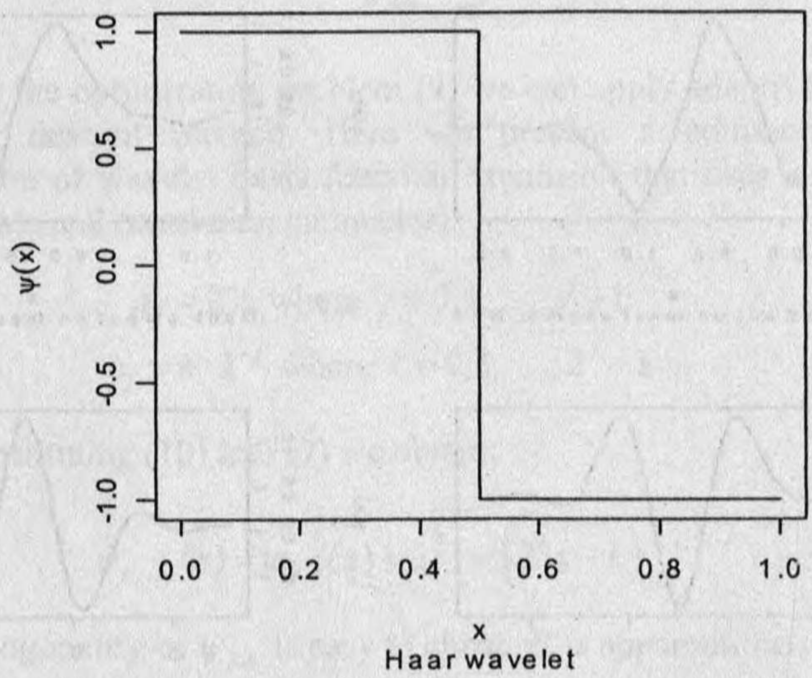

Fig. 1. Haar wavelet

- *Daublets* – the first type of continuous wavelet with compact support introduced by Ingrid Daubechies.

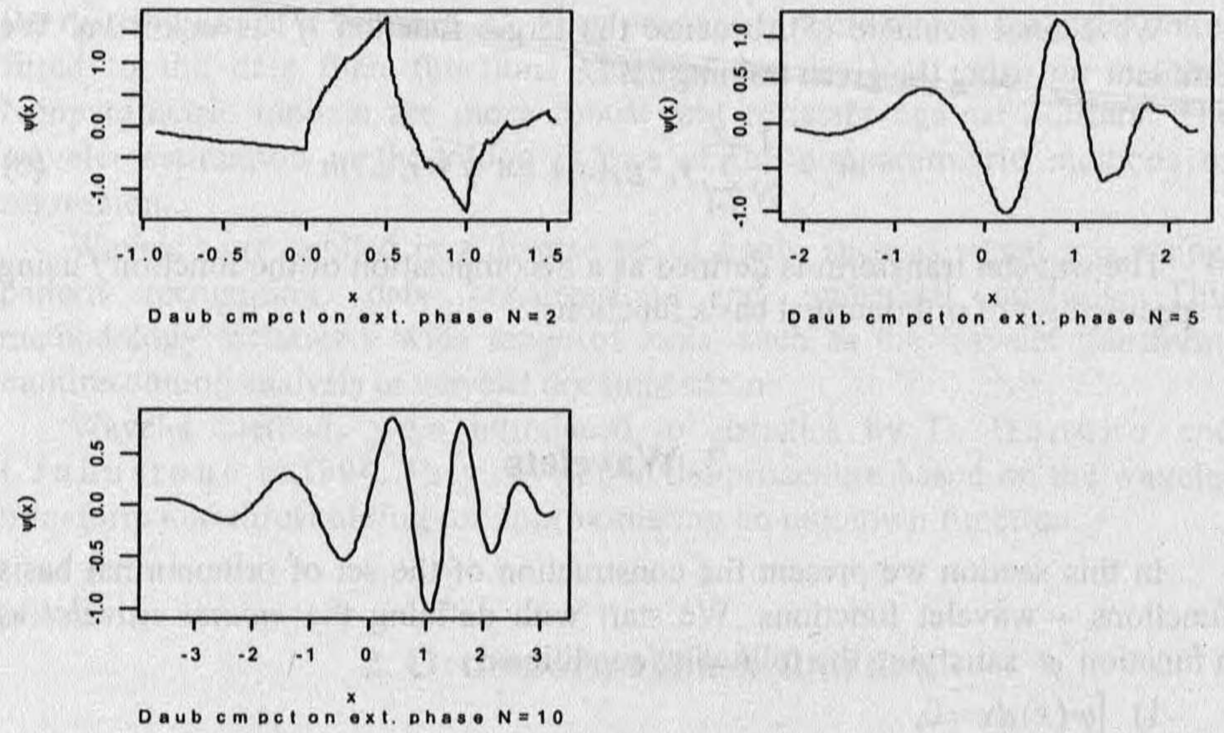

Fig. 2. Three different Daublets (with different parameter settings)

- *Symmlets* – an "nearly symmetric" equivalent of Daublets also constructed by Ingrid Daubechies.

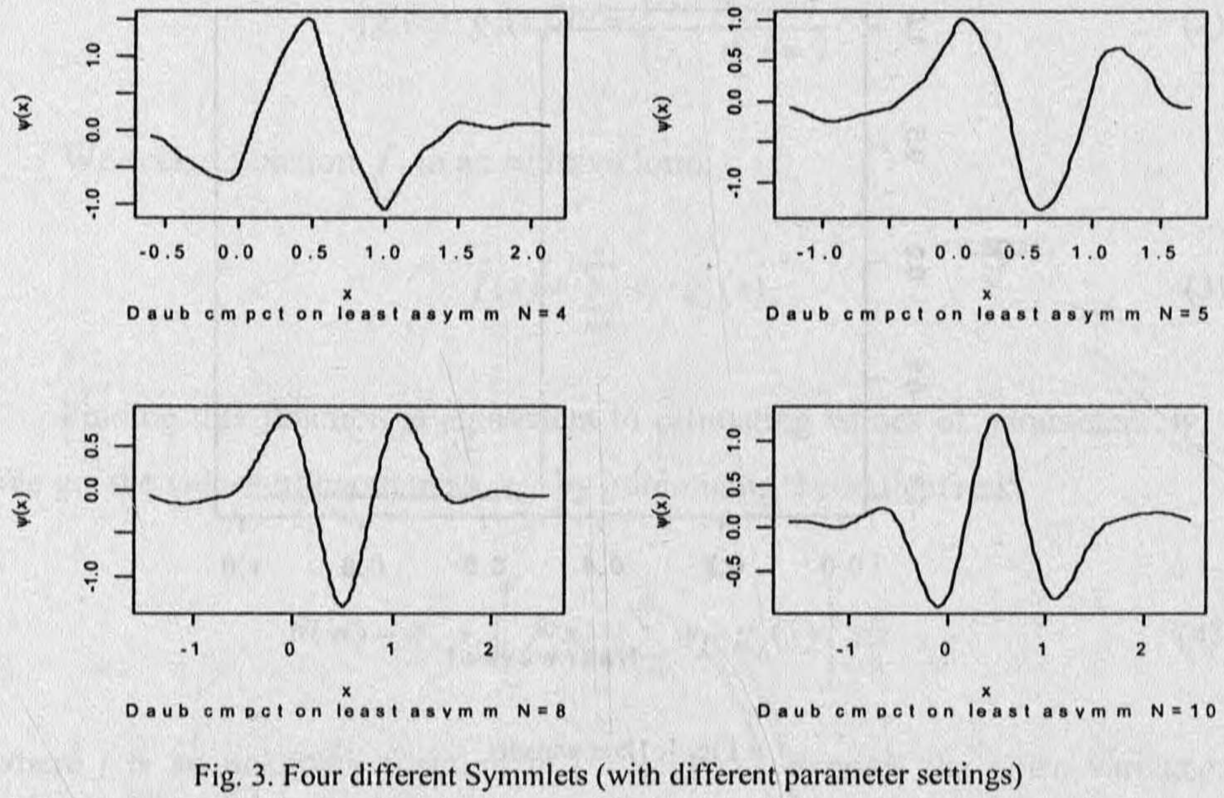

Fig. 3. Four different Symmlets (with different parameter settings)

Let us assume that a set of functions $\psi_{a,b}$ is generated from mother wavelet through scaling and translation:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \cdot \psi\left(\frac{x-b}{a}\right) \tag{7}$$

where $a > 0$ is a scale factor and $b > 0$ is a translation parameter. When a gets larger, $\psi_{a,b}$ gets shorter and more spread out. Functions $\psi_{a,b}$ defined by (7) are called *wavelet functions*.

The wavelet functions $\psi_{a,b}$ are orthonormal functions in $L^2(\mathbf{R})$ (Cherkassky *et al.*, 1998). So we can approximate a target function f as in (3) by:

$$\hat{f}(x) = \sum_{j=1}^{m} w_j \cdot \frac{1}{\sqrt{a_j}} \cdot \psi\left(\frac{x-b_j}{a_j}\right) \tag{8}$$

We can estimate the parameters of the function (8) by minimizing the loss function $L(y, f(x))$ over the training sample U:

$$(\hat{\mathbf{w}}, \hat{\mathbf{a}}, \hat{\mathbf{b}}) = \arg\min_{\mathbf{w},\mathbf{a},\mathbf{b}} \sum_{i=1}^{N} L(y_i, f(x_i)) \tag{9}$$

To solve the optimization problem (9) we can apply adaptive methods, e.g. the gradient descent method. Here we present a common nonadaptive implementation of wavelet basis function expansion that uses a basis function with fixed scale and translation parameters:

$$\begin{aligned} a_j &= 2^{-j} \text{ where } j = 0, 1, \ldots, J-1 \\ b_j &= k \cdot 2^{-j} \text{ where } k = 0, 1, \ldots, 2^j - 1 \end{aligned} \tag{10}$$

Then substituting (10) into (7) we obtain:

$$\psi_{a_j,b_j}(x) = \psi_{j,k}(x) = 2^{\frac{j}{2}} \cdot \psi\left(2^j x - k\right) \tag{11}$$

The orthogonality of $\psi_{j,k}$ is easy to check. It is apparent that:

$$\int \psi_{j,k}(x) \cdot \psi_{j',k'}(x) dx = \begin{cases} 1, & \text{if } j = j' \wedge k = k', \\ 0, & \text{if } j \neq j' \vee k \neq k' \end{cases} \tag{12}$$

Thus the set $\{\psi_{j,k} : j \in \mathbf{Z}, k \in \mathbf{Z}\}$ defines an orthonormal basis for $L^2(\mathbf{R})$ (Hastie *et al.*, 2001).

4. The wavelet transform

Given the wavelet functions of the form (11) we obtain approximation of the target function f:

$$\hat{f}(x)=\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} w_{jk}\cdot 2^{\frac{j}{2}}\psi\left(2^j x-k\right) \tag{13}$$

The formula (13) defines *the wavelet transform* of a function $\hat{f}$. Coefficients w_{jk} in (13) have the following form:

$$w_{jk}=\int f(x)\cdot\psi_{j,k}(x)dx \tag{14}$$

Hence the target function f is unknown we estimate the values of parameters w_{jk} by $\hat{w}_{jk}$ using the training set:

$$\hat{w}_{jk}=\frac{1}{N}\sum_{i=1}^{N} y_i\cdot\psi_{j,k}(x_i) \tag{15}$$

5. The wavelet thresholding

The presence of noise in the training data set implies the values of many coefficients $\hat{w}_{jk}$ close to zero. It is connected with the problem of overfitting the data. Donoho and Johnstone addressed the issue with *wavelet thresholding.* There are two popular approaches to it:

a) *"hard" thresholding* where all wavelet coefficients smaller than a certain threshold θ are set to zero:

$$\hat{w}_{jk}^{s}=\hat{w}_{jk}\cdot I(|\,\hat{w}_{jk}\,|>\theta) \tag{16}$$

b) *"soft" thresholding*, where:

$$\hat{w}_{jk}^{s}=\operatorname{sgn}(\hat{w}_{jk})\cdot\max\{0,|\,\hat{w}_{jk}\,|-\theta\} \tag{17}$$

There are many ideas for choosing the value of the threshold θ, e.g. a very popular formula:

$$\theta=\sigma\cdot\sqrt{2\ln N} \tag{18}$$

where N is the number of observations in the data set and σ is the standard deviation of noise (usually estimated from the data).

Summarizing, the regression function has the form of wavelet decomposition:

$$\hat{f}(x) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \hat{w}_{jk}^s \cdot 2^{\frac{j}{2}} \psi\left(2^j x - k\right) \tag{19}$$

where $\hat{w}_{jk}^S$ are adjusted coefficients given by the formula (16) or (17).

6. Example of application of wavelet transform

For the illustration of the wavelet transform and wavelet thresholding we conduct computation on the *bev* data set. This set contains the well-known *Beveridge Wheat Price Index* which gives the annual price data from 1500 to 1869, averaged over many locations in western and central Europe. It is an univariate time series with 370 observations (Fig. 4).

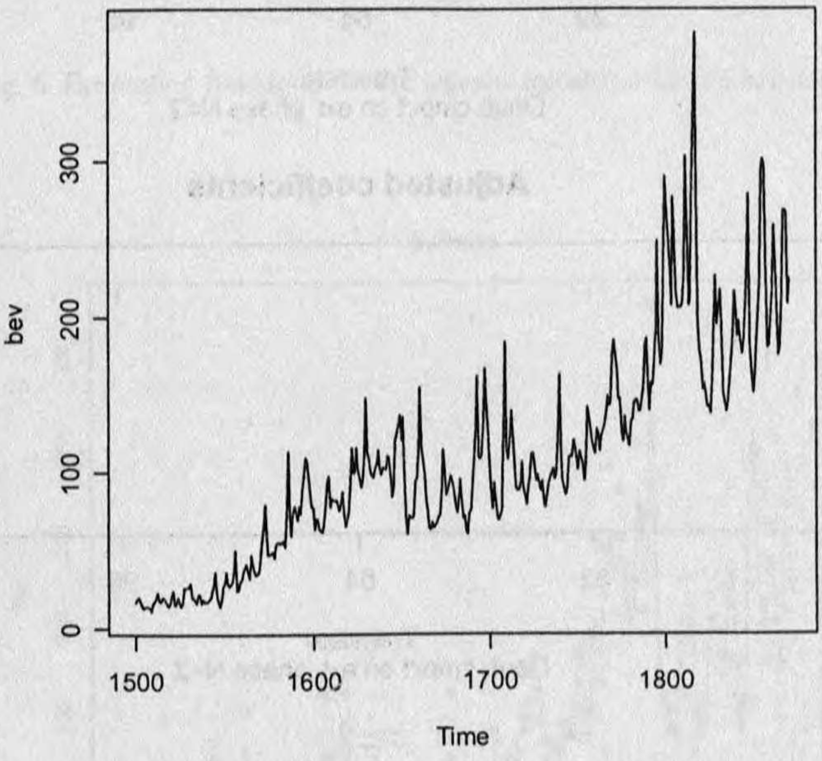

Fig. 4. Plot of the *bev* time series

D. Donoho and I. Johnstone developed the *WaveShrink* procedure estimating an unknown function f. *WaveShrink* is able to remove the noise from the time series while preserving the spike. Traditional noise reduction methods, such as splines, would result in some smoothing of the spike.

The *WaveShrink* procedure can be presented as follows:

1) Apply the wavelet transform (decomposition) of observations from the *bev* set.
2) Threshold the wavelet coefficients towards zero.
3) Use the wavelet reconstruction as an estimate $\hat{f}$.

The process of shrinking coefficients is much like the process of keeping only important coefficients of wavelet decomposition.

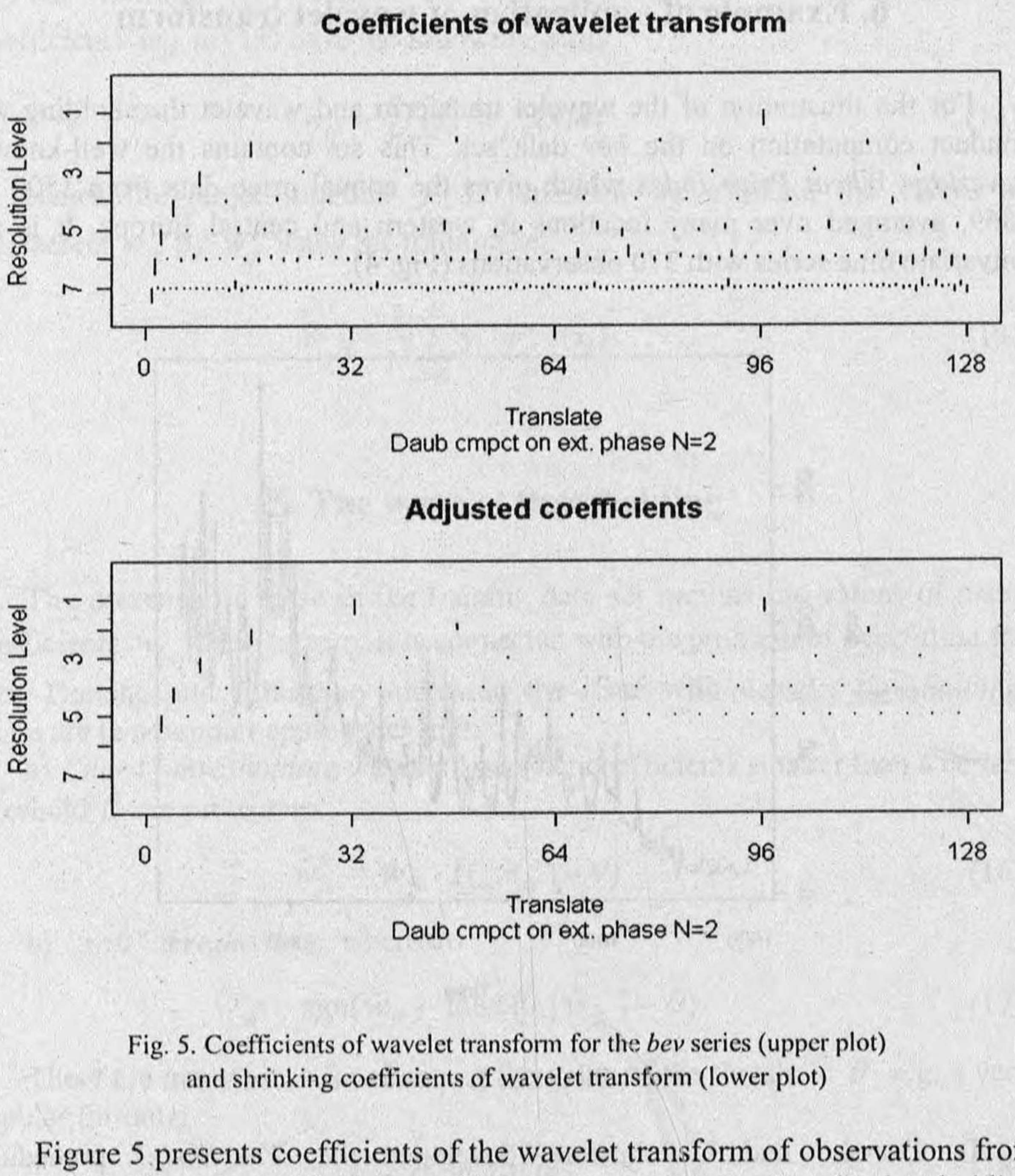

Fig. 5. Coefficients of wavelet transform for the *bev* series (upper plot) and shrinking coefficients of wavelet transform (lower plot)

Figure 5 presents coefficients of the wavelet transform of observations from the *bev* data set and adjusted coefficients given by the "hard" thresholding procedure. Figure 6 displays the estimating function $\hat{f}$ via the *WaveShrink* procedure for the *bev* time series.

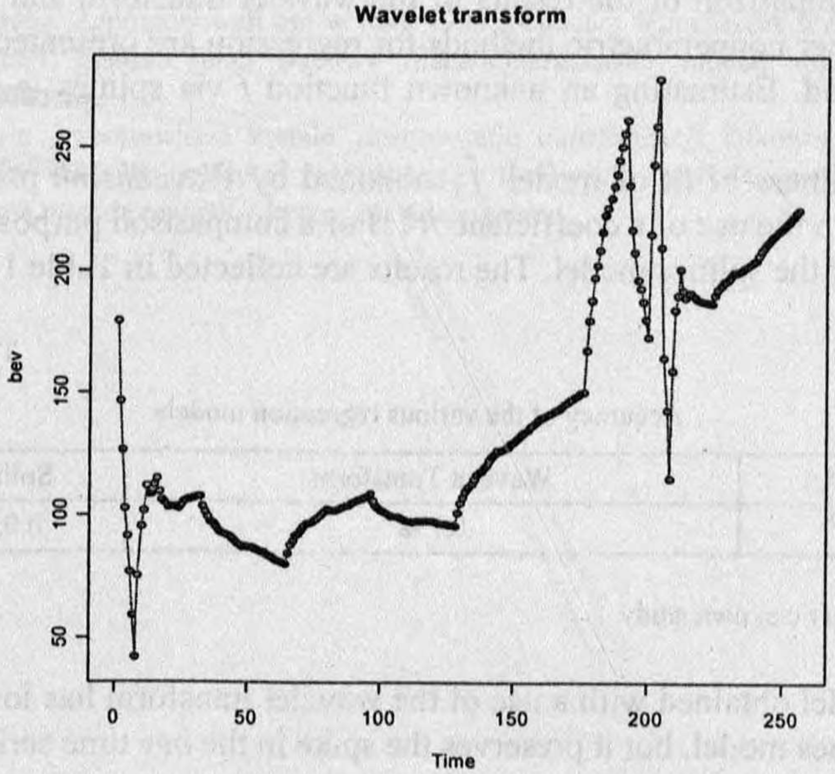

Fig. 6. Estimating function *f* via the wavelet transform for the *bev* series

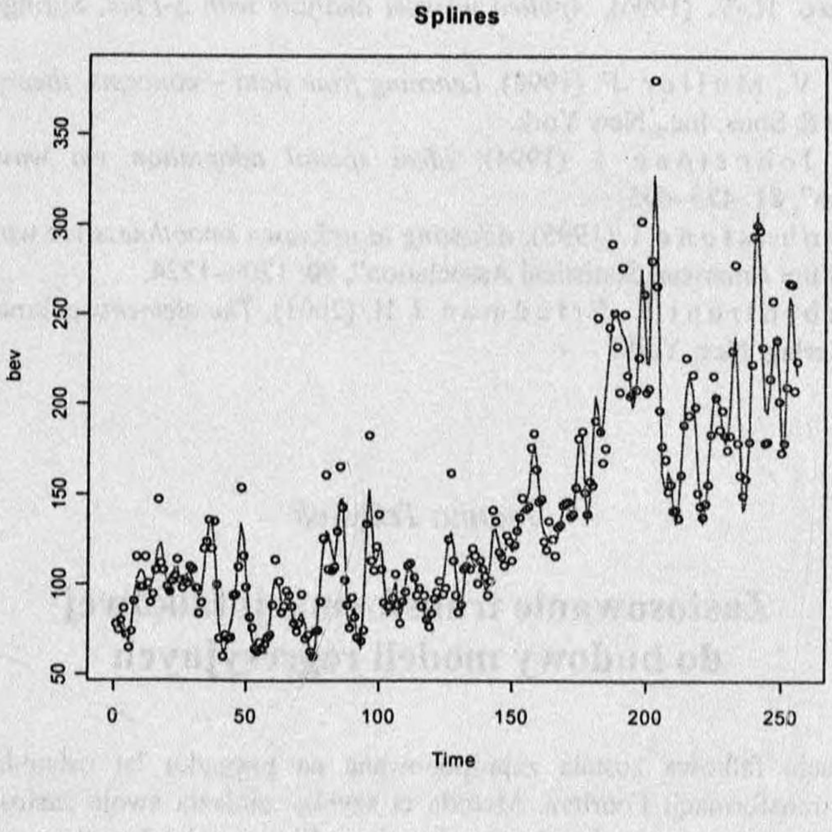

Fig. 7. Estimating function *f* via splines for the *bev* series

For a comparison of the results of the wavelet transform and thresholding procedure other nonparametric methods for regression are presented. Here is the splines method. Estimating an unknown function f via splines is illustrated in Fig. 7.

The goodness of fit of model $\hat{f}$, obtained by *WaveShrink* procedure, was measured with the use of a coefficient R^2. For a comparison purpose, R^2 is also calculated for the splines model. The results are collected in Table 1.

Table 1

Accuracy of the various regression models

Model	Wavelet Transform	Splines
R^2	0.738	0.956

Source: own study.

The model obtained with a use of the wavelet transform has lower accuracy than the splines model, but it preserves the spike in the *bev* time series.

References

Bruce A., Gao H.-Y. (1996), *Applied wavelet analysis with S-Plus*, Springer-Verlag, New York.

Cherkassky V., Mulier F. (1998), *Learning from data – concepts, theory, and methods*, John Wiley & Sons, Inc., New York.

Donoho D., Johnstone I. (1994), *Ideal spatial adaptation via wavelet shrinkage*, „Biometrika", **81**: 425–455.

Donoho D., Johnstone I. (1995), *Adapting to unknown smoothness via wavelet shrinkage*, „Journal of the American Statistical Association", **90**: 1200–1224.

Hastie T., Tibshirani R., Friedman J. H. (2001), *The elements of statistical learning*, Springer-Verlag, New York.

Joanna Trzęsiok

Zastosowanie transformacji falkowej do budowy modeli regresyjnych

Transformacja falkowa została zaproponowana na początku lat osiemdziesiątych, jako alternatywa do transformacji Fouriera. Metoda ta szybko znalazła swoje zastosowanie w teorii sygnałów oraz w rozpoznawaniu obrazów, a zakres jej aplikacji nadal dynamicznie się rozwija.

Autorami pionierskich prac z zakresu zastosowań teorii falek w statystyce są David Donoho and Iain Johnstone. Zaproponowali oni w roku 1994 procedurę *WaveShrink* wykorzystywaną do estymacji funkcji gęstości oraz budowy nieparametrycznych modeli regresji opartą na transformacji falkowej.

W artykule przedstawione zostało zastosowanie transformacji falkowej oraz procedury *WaveShrink* do budowy modelu regresyjnego. Omawianą metodę porównano z inną nieparametryczną metodą regresji – krzywymi sklejanymi.