TIME SERIES FORECASTING USING NON-DECIMATED WAVELET TRANSFORM

E. Mirotin
Belorussian State University
Minsk, BELARUS
e-mail: emirotin@gmail.com

Abstract

The non-decimated wavelet transform (NDWT) is widely used in data forecasting during last decades, though majority of proposed algorithms do not have complete vindication, and are based on intuitive assumptions. This paper tries to put some criteria for choosing the algorithm filter, and shows how some of these criteria can be sufficed in special case of polynomial trend.

1 Non-decimated Wavelet-Transform

Let us have time series $Y(t), t \in \mathbb{N}$.

Non-decimated wavelet coefficients are computed in the following way [5].

The number of decomposition levels M is chosen. Then the family of iteratively "smoothed" time series is computed by the following rules:

$$c_0(t) = Y(t), t \in \mathbb{N}$$

$$c_i(t) = \sum_{k=-n}^{n} \phi(k) c_{i-1}(t + 2^{i-1}k), t \in \mathbb{N}, i = 1, \ldots, M, \tag{1}$$

where $\phi(\cdot)$ filtering vector of length $2n+1$, $\sum_{k=-n}^{n} \phi(k) = 1$.

After that, the differences are calculated:

$$w_i(t) = c_{i-1}(t) - c_i(t), t \in \mathbb{N}.$$

The initial time series can be recomposed as:

$$Y(t) = c_M(t) + \sum_{j=1}^{M} w_j(t).$$

The component c_M is the "smoothed" series and considered to be an estimation of the trend, and components w_j, $j = 1, \ldots, M$, called detalization components, are refinement components. The bigger M is, the smoother c_M should be, and w_j for greater j corresponds to smaller details.

Note. In practice t usually changes between finite bounds (most often from 1 to some $T \in \mathbb{N}$). In this case the formula (1) is inappropriate for t close to 1 or T. In order

to fix it the formula is corrected by base reduction, which allows eliminating indices less that 1 or bigger than T.

Generally algorithms using NDWT are based on assumption that extracted components w_j have stationarity properties [2]. For investigation of these properties let us create the analogue of NDWT for continuous case.

Let us have the probability process $c_t = c(t) = c(t, \omega),\ t \in \mathbb{R}$.

Consider we chosen a number $p > 0$ and function φ with compact support such that $\int_{-\infty}^{\infty} \varphi(t)dt = 1$.

We define the smoothing operator:

$$W_A c(t) = \int_{\mathbb{R}} c\,(t + pl)\,\varphi(l)dl. \tag{2}$$

We will call the resultant probability process the "approximation component" of the initial process.

We also define the detalization operator as:

$$W_D c(t) = c(t) - W_A c(t). \tag{3}$$

The resultant process will be called the "detalization component".

Thus, the continuous NDWT splits the initial process into two components. Empirically (analogically to [1–5]) the approximation component is considered as trend estimation, and the detalization component - a stochastic part of the initial process.

The reverse transform has the trivial formula:

$$c(t) = W_A c(t) + W_D c(t).$$

Theorem 1. *Let $c(t)$ be the probability process with mean $m(t)$ and covariance function $f(t, s)$, both considered to be measurable, and let $d(t) = W_A c(t), w(t) = W_D c(t)$. We denote $h(t) = \frac{1}{p}\varphi\left(\frac{t}{p}\right)$. Then $d(t)$ has the mean equal to*

$$E\{d(t)\} = \int_{\mathbb{R}} m(t + l)h(l)dl, \tag{4}$$

and covariance function

$$cov\{d(t), d(s)\} = \iint_{\mathbb{R}^2} f(t + l, s + l')h(l)h(l')dldl'. \tag{5}$$

And for $w(t)$ we have

$$E\{w(t)\} = m(t) - \int_{\mathbb{R}} m(t + l)h(l)dl, \tag{6}$$

$$cov\{w(t), w(s)\} = f\,(t, s) - \int_{\mathbb{R}} f\,(t, l)\,h\,(l - s)\,dl - \int_{\mathbb{R}} f\,(k, s)\,h\,(k - t)\,dk + \iint_{\mathbb{R}^2} f\,(k, l)\,h\,(k - t)\,h\,(l - s)\,dkdl. \tag{7}$$

Definition 1. *We call the non-stationarity rate of the probability process $c(t)$ (in terms of mean) of size K the value $N_K^{mean}(c(t)) = \|m(t) - m_C\|_{L_2(\mathbb{R}, \frac{1}{2K} I_{[-K,K]})}$, where m_C is the projection of $m(t)$ to the subspace $C = \{g(t) : g(t) = const\} \subseteq L_2(\mathbb{R}, \frac{1}{2K} I_{[-K,K]})$.*

In a similar way the non-stationarity rate in terms of covariance can be defined.

2 Criteria of NDWT Applicability to Data Forecasting

No we can define the criteria for considering NDWT with specific filter applicable for forecasting. These criteria in a faithful way correspond to the previously mentioned assumptions of approximation component being trend estimation and detalization component being stochastic part of the process.

We consider continuous NDWT being applied to the probability process $c(t)$ in a chain way like in (1):

$$\begin{aligned} c_0(t) &= c(t), \\ c_j(t) &= W_A c_{j-1}(t), \\ w_j(t) &= W_D c_{j-1}(t), j = 1, \ldots, M. \end{aligned} \tag{8}$$

Then, the criteria would be as follows:

- All w_j have non-stationarity rate by mean non greater than c, and better equal to 0;
- All w_j have non-stationarity rate by covariance non greater than c, and better equal to 0;
- c_M has variance non greater than c, and better converging to 0 when M converges to infinity.

Theorem 2. *If $c(t) = m(t) + \xi_t$, where $m(t)$ is non-stochastic function (trend), and ξ_t is stationary probability process with zero mean, then $d(t) = W_A c(t)$ has the same form.*

Hence supposing such a form of $c(t)$ we can check criteria defined above for only one step of NDWT, i.e. for $w = w_1, d = c_1$.

3 Non-stationarity by Mean for Polynomial Trend

We consider $c(t) = m(t) + \xi_t$, $m(t) = \sum_{i=0}^{s} a_i t^i$ is a polynomial of degree non greater than s and ξ_t is stationary probability process with zero mean.

The following statements apply.

Theorem 3. *If we know that the degree of $m(t)$ is not greater than s, there exist function φ such that $N_K^{mean}\langle w(t)\rangle = 0$ for any p. This function can be found in form*

$$\varphi(k) = \begin{cases} k^{s+1} - \sum_{i=0}^{s} q_i k^i, & k \in [-1,1], \\ 0, & otherwise, \end{cases} \tag{9}$$

where coefficients q_i can be calculated from a linear system

$$\begin{cases} \alpha_0 q_0 & +\alpha_1 q_1 & +\cdots & +\alpha_s q_s & = \alpha_{s+1} - 1 \\ \alpha_1 q_0 & +\alpha_2 q_1 & +\cdots & +\alpha_{s+1} q_s & = \alpha_{s+2} \\ \alpha_2 q_0 & +\alpha_3 q_1 & +\cdots & +\alpha_{s+2} q_s & = \alpha_{s+3} \\ \vdots & & & & \\ \alpha_s q_0 & +\alpha_{s+1} q_1 & +\cdots & +\alpha_{2s} q_s & = \alpha_{2s+1} \end{cases}, \tag{10}$$

whith $\alpha_i = \begin{cases} 2/(i+1), & i \text{ even}, \\ 0. & i \text{ odd}, \end{cases} \; i = 0, 1, 2, \cdots$

And it always applies that the system has the only solution.

Theorem 4. *There is no function φ with compact support such that $N_K^{mean}\langle w(t)\rangle = 0$ for any polynomial $m(t)$ (no matter how we choose p).*

Theorem 5. *There are functions $\varphi(t)$ from the Schwartz space S such that $N_K^{mean}\langle w(t)\rangle = 0$ for any polinomial $m(t)$ for some p.*

References

[1] Aussem A., Campbell J., Murtagh F. (1997). Wavelet-Based Feature Extraction and Decomposition Strategies for Financial Forecasting. *Journal of Computational Intelligence in France.*

[2] Conejo A.J., Plazas M.A., Espinola R., Molina A.B. (2005). Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models. *IEEE Transactions on Power Systems.* Vol. **20**, No. **2**.

[3] Craigmile P.F., Percival D.B. (2001). Wavelet-Based Trend Detection and Estimation. *Encyclopedia of Environmetrics.* Vol. **4**, edited by A. H. El-Shaarawi and W.W. Piegorsch. John Wiley and Sons, Ltd, Chichester, England.

[4] Fryzlewicz P., Van Bellegem S., von Sachs R. (2003). Forecasting Non-stationary Time Series by Wavelet Process Modeling. *Annals of the Institute of Statistical Mathematics.* No. **55**.

[5] Renaud O., Starck J.-L., Murtagh F. (2004). Wavelet-based Forecasting of Short and Long Memory Time Seriesr. *Cahiers du department econometrie Faculte des sciences economiques et sociales Universite de Geneve.* No. **2004/04**.