NEURAL NETWORKS WITH WAVELET BASED DENOISING LAYER: APPLICATION TO CENTRAL EUROPEAN STOCK MARKET FORECASTING*

Jozef Barunik, Lukas Vacha

Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

Dept. of Econometrics, barunik@utia.cas.cz, vachal@utia.cas.cz

Extended Abstract

Traditional prediction methods for time series often restrict on linear regression analysis, exponential smoothing, and ARMA. These methods generally produce reasonable prediction results for stationary random time series of linear systems. In the recent decades, development in econometrics resulted also in methods which are capable of forecasting more complex systems, such as Wavelet decomposition or Neural Networks. These methods proved to better explain the complex stock market behavior. In this paper we apply neural network with wavelet denoising layer method for forecasting of Central European Stock Exchanges, namely Prague, Budapest and Warsaw. Hard threshold denoising with Daubechies 6 wavelet filter and three level decomposition is used to denoise the stock index returns, and two-layer feed-forward neural network with Levenberg-Marquardt learning algorithm is used for modeling. The results show that wavelet network structure is able to approximate the underlying process of considered stock markets better that multilayered neural network architecture without using wavelets. Further on we discuss the impact of structural changes of the market on forecasting accuracy on the daily stock market data. Stock markets change their structure rapidly with changing agent sentiment structure. These changes then have great impact on the prediction accuracy.

Keywords: *neural networks, hard threshold denoising, time series prediction, wavelets.*

1. Introduction

Traditional prediction methods for time series often restrict on linear regression analysis, exponential smoothing, and ARMA. These methods generally produce reasonable prediction results for stationary random time series of linear systems. In the recent decades, development in econometrics brought also methods which are capable of forecasting more complex systems, such as stock markets. These mainly include Wavelet Decomposition [10], [4], [1] and Neural Networks analysis, [5], [9]. Further on, the idea of combining both methods into single wavelets and neural networks has resulted in the formulation of wavelet networks. This research area is new, and there is tremendous potential for its development and application to various fields. This paper is one of the first attempts to fit this exciting method to Central European Stock markets, namely Prague Stock Exchange, Budapest Stock Exchange and Warsaw Stock Exchange. Our main expectation will be that this method will improve the single neural network approach.

2. Denoising with Wavelets

Classical time series denoising approaches are rooted in Fourier analysis where noise is assumed to be represented mainly as high frequency oscillations. The wavelet based denoising, assumes that analysis of time series at different resolutions might improve the separation of the true underlying signal from noise. Let us begin with description of the basics of Wavelet decomposition theory.

* A support from the Grant Agency of Charles University (GAUK) under grant 46108 and Czech Science Foundation (GAČR) under grants 402/08/P207 and 402/06/1417

2.1. Wavelets

There are two types of wavelets: father wavelets ϕ and mother wavelets ψ. The father wavelet integrates to unity and the mother wavelet integrates to zero. The father wavelet, also called scaling function, essentially represents the smooth, trend, i.e., the low frequency part of the signal, on the other hand the mother wavelet represents the details, i.e., the high frequency part of the signal. The mother wavelet is compressed or dilated in time domain, to generate cycles to fit the actual time series. The formal definition of the father ϕ and mother ψ wavelet is

$$\phi_{j,k} = 2^{-\frac{j}{2}}\phi\left(\frac{t-2^j k}{2^j}\right), \quad \psi_{j,k} = 2^{-\frac{j}{2}}\psi\left(\frac{t-2^j k}{2^j}\right)$$

where j is the scale (or dilatation) and k is the translation (or shift). Commonly, many types of wavelets can be possibly used, including Haar wavelet, Mexican hat, Morlet wavelet, Daubechies wavelet etc. In the empirical part, we will use Daubechies wavelet db6.

Any time series $x(t)$ can be built up as a sequence of projections onto father and mother wavelets indexed by both j, the scale, and k, the number of translations of the wavelet for any given scale. Usually k is assumed to be dyadic. The wavelet coefficients are approximated by integrals

$$s_{J,k} \approx \int_{-\infty}^{\infty} x(t)\phi_{J,k}(t)dt, \quad d_{j,k} \approx \int_{-\infty}^{\infty} x(t)\psi_{j,k}(t)dt$$

$j = 1, 2, \ldots, J$, where J is the maximum scale. The wavelet representation of the time series $x(t)$ in $L^2(R)$ [1] can be given by

$$x(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) + \sum_k d_{J-1,k}\psi_{J-1,k}(t) + \ldots + \sum_k d_{1,k}\psi_{1,k}(t)$$

where the basis functions $\phi_{J,k}(t)$ and $\psi_{j,k}(t)$ are assumed to be orthogonal. When the number of observations is dyadic, the number of the wavelet coefficients of each type at the finest scale 2^1 is $N/2$, labeled $d_{1,k}$. The next scale 2^2 has $N/2^2$ coefficients, labeled $d_{2,k}$. At the coarsest scale 2^J there are $N/2^J$ coefficients $d_{J,k}$ and $s_{J,k}$.

3. Nonlinear Wavelet Denoising

The simplest method of nonlinear wavelet denoising is via thresholding. The procedure sets all wavelet coefficients that has lower value than some fixed constant to zero. Two thresholding rules were instrumental in the initial development of wavelet denoising, both for their simplicity and performance: hard and soft thresholding [4].

---

[1] Square integrable real-valued function, $\int_{-\infty}^{\infty} x^2(t)dt < \infty$.

3.1. Threshold Selection

Optimal thresholding occurs when the threshold is set to the noise level, i.e., $\eta = \sigma_{\in}$. Setting $\eta < \sigma_{\in}$ will allow unwanted noise to enter the estimate while setting $\eta > \sigma_{\in}$ will destroy information that belongs to the underlying signal. Following [3] we can set a universal thresholding as

$$\eta^U = \hat{\sigma}_{\in}\sqrt{2\log N},$$

where N is the sample size. In practical situations the standard deviations of noise $\sigma_{\in}$ is not known. The most commonly used estimator of $\sigma_{\in}$ is the maximum absolute deviation (MAD) standard deviation [10].

$$\hat{\sigma}_{MAD} = \frac{median(|d_{1,1}|, |d_{1,2}|, \ldots, |d_{1,N/2-1}|)}{0.6745}.$$

The denominator is needed to rescale the numerator so that $\hat{\sigma}_{MAD}$ is tuned to estimating the standard deviation for Gaussian white noise [4].

3.2. Hard Thresholding

In our paper we use a hard thresholding. The hard thresholding rule on the wavelet coefficients o_t is given by

$$\delta_\eta^H(o_t) = \{ \begin{array}{l} o_t \text{ if } |o_t| > \eta \\ 0 \text{ otherwise} \end{array},$$

where η is the threshold value. The operation is not a continuous mapping, it only affects input coefficients that are less or equal to the threshold η. After obtaining the thresholded wavelet coefficients using δ_η^H we compose the denoised times series via an inverse wavelet transform (IDWT) so we get $\hat{x}_{den}(t)$. For a more detailed treatment see [10].

4. Wavelet Network Structure

Wavelet Network is a network combining the ideas of the feed-forward neural networks and the wavelet decomposition. Wavelet networks use simple wavelets and wavelet network learning is performed by the standard type algorithm such as Conjugate-Gradient, or more efficient Levenberg-Marquardt [6], [8]. Neural networks can be viewed as universal approximation tools for fitting linear or nonlinear models, as [5] showed. Limiting space of this paper do not allow us to explore Neural Networks estimation methodology in detail, but reader is advised to follow i.e. [12], or [9] for very good explanation.

There are basically two main approaches to form wavelet networks. In the first approach, the wavelet decomposition is decoupled from the learning component of neural network architecture. In other words, the series are firstly decomposed / denoised using wavelets, and then fed to the neural network. In the second approach, the wavelet theory and neural networks are combined into a single method, where the inputs $x_1, \ldots, x_k$ with weights $\omega_1, \ldots, \omega_n$ are combined to estimated output in Multilayer feed-forward network (MPL):

$$\widehat{x}_{DWNN}(t) = \sum_{i=1}^{N} \omega_i f(\gamma_i x(t) + \beta_i),$$

where f is an activation function, $\gamma_i, \beta_i, \omega_i$ are network weight parameters that are optimized during learning, and N is number of hidden layers. If we feed this network with nonlinear wavelet denoised values $\widehat{x}_{den}(t)$ (see section 3) we will get the form of the wavelet neural network (DWNN) used in this paper.

5. Results

In the testing, we focus on sample of 1050 daily returns from 7.1.2004 to 3.4.2008 of value-weighted indices PX-50, BUX and WIG (Prague, Budapest and Warsaw stock exchanges respectively). The dataset was downloaded from the server www.stocktrading.cz. In the prediction task we start with denoising of the 512 data sample with db6 wavelet filter and a 3-level decomposition. Then two-layer neural network with 5 neurons in each layer and Levenberg-Marquardt learning algorithm is used to learn the sample. To avoid over-fitting of the network we use a window of 50 real out-of-sample data on which we test the estimated model on one day predictions. This algorithm is repeated 10 times with moving window of 50, so the final prediction of 500 data is obtained. Finally we compare this method to neural network approach so we can see if the wavelet denoising layer improves the forecasts. Architecture of the network is again two layers with 5 neurons and Levenberg-Marquardt optimization.

As for evaluation, we focus mainly on out-of-sample performance, as it is most important in financial time series forecasting. We consider Root Mean Square Error statistics (RMSE) to see the performance of out-of-sample prediction. Further on, we use statistics proposed by Pesaran Timmerman – SR (PT) [11], which evaluates the correctness of the signs prediction. Such statistics is often used in financial literature as the predicted positive change predicts buy signal, negative change sell signal which allows evaluating a trading strategy. Pesaran Timmerman statistics is based on the null hypothesis that a given model has no economic value in forecasting direction and is approximately normally distributed. In other words, we test the null hypothesis that the signs of the forecasts and the signs of actual variables are independent. If the prediction of signs is statistically dependent, we approached a good forecasting model with economic significance.

Finally to test the performance of the wavelet network and the neural network approach we use the statistic proposed by Clark and McCracken (CM) [2]. They compare out-of-sample accuracy for two models which are nested. The statistics is normally distributed under the null hypothesis of equal predictive ability of the two models.

Table 1: Prediction results for PX50, BUX and WIG

	PX50		BUX		WIG	
	MPL	DWNN	MPL	DWNN	MPL	DWNN
RMSE	0.8	0.175	0.08	0.11	0.37	0.41
CM	0.78***		0.25***		-7.6	
SR (PT)	0.52	0.56**	0.51	0.53***	0.47	0.49

*, **, ***, 1%, 5% and 10% significance levels

To compare the results, we can see that Wavelet Neural Networks (DWNN) performed best on BUX returns with lowest RMSE, while RMSE of PX50 was little bit higher, and WIG more than double of PX50 and BUX. This would indicate that the DWNN model will forecast the PX50 and BUX returns better than WIG. This is confirmed by Pesaran Timmerman statistics which is significant on 5% level of significance for PX 50, 10% level of significance of BUX but is not significant at all for WIG. Success rate of correct forecasted direction is 0.56 for PX50, 0.53 for BUX and 0.49 for WIG. As to the Multilayer Feedforward architecture (MPL) without denoising the directional accuracy is not significant at all. But if we compare the DWNN and MPL using Clark and McCracken statistics, we can see that DWNN yields significantly better prediction accuracy for PX50 and BUX series. For WIG the results are not significantly different, which would be expected as WIG does not seem to be significantly predictable even using Pesaran Timmerman.

To sum up the achieved results we can say, that DWNN performs significantly better on the PX50 and BUX returns prediction. Although reader can notice, that the success rate is quite low. This is probably caused by quite large data sample which contains large structural changes as recent stock market crash of February 2008, etc. Even though Wavelet Neural Networks are considered as a universal approximation theorem as mentioned before, reader can see that if we feed it with data which simply cannot be approximated, its performance is poor. As the stock market structure changes in time quite quickly, testing the prediction on such a large data samples does not seem to provide reasonable predictions. Thus our last test is focused on the division of the dataset into 3 month moving windows, where we simply look at how does the Success Rate statistics of Pesaran Timmerman evolve in time.

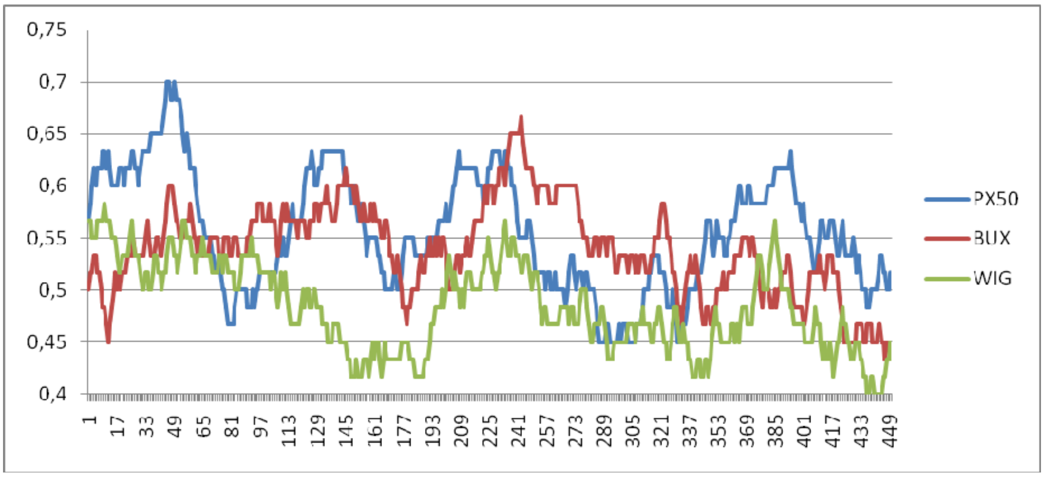

Figure 1: Success Rate statistics of Pesaran Timmerman of PX50, BUX and WIG in time.

Figure 1 demonstrates that our assumption was right. For some periods, the statistics for PX50 and BUX returns is 0.6 to 0.7, which means that 60% to 70% of the one day sign change is predicted correctly. For WIG returns the predictability does not excess 55%, which can lead us to suspicion that WIG simply does not contain strong predictable patterns. As to other two tested series, we can see that the pattern strongly evolves over time, and that it would make sense to adjust appropriate method for forecasting each 3-month. This would be done using adjusting wavelets and their levels, number of hidden layer of neural network, etc. which we leave for further research.

Conclusion

Our results indicate that the Wavelet Neural Network might outperform simple Neural Networks while forecasting Central European stock exchanges. More concretely, PX50 and BUX returns were predicted using DWNN structures significantly better than using MPL Network structure. We also conclude that the stock market structural changes affected the final stock market direction prediction greatly. Dataset used for testing was quite large and contained structural changes such as large market crash of February 2008 which leads to significantly lower prediction accuracy of the used methods. For this reason we also used the three month moving window using which we have showed, that for some periods the prediction accuracy reached sustainable 60% to 70%, meaning 60% to 70%

future directions of the stock markets were predicted correctly using PX50 and BUX data. On the other hand, the prediction accuracy using WIG data did not improve a lot. Thus we conclude that this market does not simply contain stronger predictable patterns. Further research should concentrate on the exploring of dynamically adjusted wavelet types, number of hidden layers of neural networks or other parameters specific to structural changes in the forecasted underlying series.

References:

[1] Abramovich F., Bailey T.C., Sapatinas T. (1999), *Wavelet Analysis and its Statistical Applications*, J. R. Statist. Soc. D, 48.

[2] Clark, Tod E., and Michael W. McCracken (2001): *Tests of Forecast Accuracy and Encompassing for Nested Models*, Journal of Econometrics 105: 85-110.

[3] Donoho, D. L. and Johnstone, I. M. (1994) *Ideal Spatial Adaptation by Wavelet Shrinkage.* Biometrica, 81, 422-455.

[4] Gençay, R., Selçuk, F., Whitcher, B. (2002), *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press.

[5] Hornik, K., (1989): *Multilayer Feedforward Networks are universal Approximators,* Neural Networks 2, 359-366.

[6] Levenberg, K. (1944): *A Method for the Solution of Certain Problems in Least Squares.* Quart. Appl. Math. 2, 164-168

[7] Mallat S. (1998), *A Wavelet Tour of Signal Processing,* Academic Press

[8] Marquardt, D. (1963): *An Algorithm for Least-Squares Estimation of nonlinear Parameters.* SIAM J.Appl.Math. 11, 431-441.

[9] McNelis, P.D. (2005): *Neural Networks in Finance: Gaining predictive edge in the markets,* Elsevier Academic Press, ISBN 0-12-485967-4

[10] Percival, D. B., Walden, A. T. (2000), *Wavelet Methods for Time series Analysis.* Cambridge University Press.

[11] Pesaran, M.H., and A.Timmermann (1992): *A Simple Nonparametric test of Predictive Performance,* Journal of Business and Economic Statistics 10: pp. 461-465

[12] Poggio, T. and F. Girosi (1990): *Networks for Approximation and Learning.* Proc. of the IEEE, vol. 78, no.9, pp. 1481-1497.