Presented in the 5th International Conference on Advances in Applied Financial Economics (AFE). Samos, Greece (2008).

FORECASTING CASH MONEY WITHDRAWALS USING WAVELET ANALYSIS AND WAVELT NEURAL NETWORKS

Achilleas Zapranis
Department of Accounting and Finance,
University of Macedonia of Economics and Social Studies,
156 Egnatia St.,
Thessaloniki, Greece.
P.O. 54006
Tel: +30 2310 891 690
E-mail: zapranis@uom.gr

Antonis Alexandridis
Department of Accounting and Finance,
University of Macedonia of Economics and Social Studies,
156 Egnatia St.,
Thessaloniki, Greece.
P.O. 54006
Tel: +30 2310 891 631
E-mail: aalex@uom.gr

Abstract

In this paper we use wavelet neural networks to forecast cash money withdrawals in different locations in the UK. Cash demand needs to be forecasted accurately similarly to other products in vending machines, as an inventory of cash money needs to be ordered and replenished for a set period of time beforehand. If the forecasts are flawed, they induce costs: if the forecast is too high unused money is stored in the ATM incurring costs to the institution, similarly, if the ATM runs out of cash, profit is lost and customers are dissatisfied. Cash money demand represents a non-stationary, heteroscedastic process. The time series exhibits trends, singularities, seasonal and irregular structural components of the data as well as causal forces impacting on the data generating process. Having limited domain knowledge and no information on the causal forces we use wavelet analysis to extract the dynamics of the process. In order to evaluate our method we produce in-sample and out-of-sample forecasts in 11 different time series.

Keywords: Wavelets, Wavelet Networks, Forecast

1. Introduction

Wavelets analysis proved to be a valuable tool for analyzing a wide range of time-series and they have already been used with success in image processing, signal de-noising, density estimation, signal and image compression and time-scale decomposition (Donoho & Johnstone; 1994, 1998). Wavelet techniques are also used in finance, for detecting the properties of quick variation of values (Zapranis & Alexandridis, 2007b). Wavelet decomposition is considered a powerful tool for approximation. However, the wavelet analysis is limited to applications of small input dimension since its constructing wavelet basis of a large input dimension is computationally expensive (Zhang, 1997).

On the other hand neural networks can be used at large input dimension problems. Neural networks have the ability to approximate any non-linear process with no knowledge and assumptions of the nature of the process. However, in neural network framework the initial values of the weights are randomly chosen. Random initialization leads to large training times. Additionally the network usually converges to a local minimum of a specified loss function. Finally, the use of sigmoid networks does not provide any information about the network construction.

Wavelet networks were proposed by Zhang & Benveniste (1992) as an alternative to feedforward neural networks hoping to elevate the weakness of each method. Wavelet networks are one hidden layer networks that use a wavelet as an activation function instead of the classic sigmoid function. The activation function can be a wavenet (orthogonal wavelets) or a wave frame (continuous wavelets). Wavelet networks are performing excellent in predicting nonlinear behaviours (Gao & Tsoukalas, 2001). Wavelets show local characteristics hence the hidden units of the wavelet network affect the prediction of the network only in a local range. (Postalcioglou & Becerikli, 2007).

Wavelet networks have been used in a variety of applications so far. They first have been used in static and dynamic input-output modelling (Zhang & Benveniste, 1992; Postalcioglu & Becerikli, 2007) and proved that wavelet networks need less training iterations. Szu et al. (1992) used wavelet networks for classification of phonemes and speaker recognition. Gao & Tsoukalas (2001) consider wavelet networks one of the most promising tools to solve electricity load prediction problems. Subasi *et al*. (2005) used wavelet networks for classification of electroencephalography (EEG) signals while Khayamian *et al*. (2005) used wavelet networks as a multivariate calibration method for simultaneous determination of test samples of copper, iron and aluminium.

In contrast to sigmoid neural networks, wavelet networks allow constructive procedures that efficiently initialize the parameters of the network. Using wavelet decomposition a wavelet library can be constructed. Each wavelon can be constructed using the best wavelet of the wavelet library. These procedures allow the wavelet network to converge to a global minimum of the cost function. Also starting the network training very close to the solution leads to smaller training times. Finally, wavelet networks provide information of the participation of each wavelon to the approximation and the dynamics of the generating process.

In this paper, in the context of a mean reverting process, we use wavelet neural networks to forecast cash money withdrawals in different locations in the UK. Cash demand needs to be forecasted accurately similarly to other products in vending machines, as an inventory of cash money needs to be ordered and replenished for a set period of time beforehand. If the forecasts are flawed, they induce costs: if the forecast is too high unused money is stored in the ATM incurring costs to the institution, similar, if the ATM runs out of cash, profit is lost and customers are dissatisfied. Cash money demand represents a non-stationary, heteroscedastic process. The time series exhibits, trends, singularities, seasonal and irregular structural components of the data as well as causal forces impacting on the data generating process.

Having limited domain knowledge and no information on the causal forces we use wavelet analysis to extract the dynamics of the process. Wavelet Transform (WT) is localized in both time and frequency and overcomes the fixed time-frequency partitioning, (Daubechies, 1992). The time-frequency partition is long in time in low- frequencies and long in frequency in high-frequencies. This means that the WT has good frequency resolution for low-frequency events and good time resolution for high-frequency events. Also, the WT adapts itself to capture features across a wide range of frequencies. Consequently the assumption of stationarity can be avoided, (Mallat, 1999).

In order to evaluate our method we produce in-sample and out-of-sample forecasts in 11 different time series.

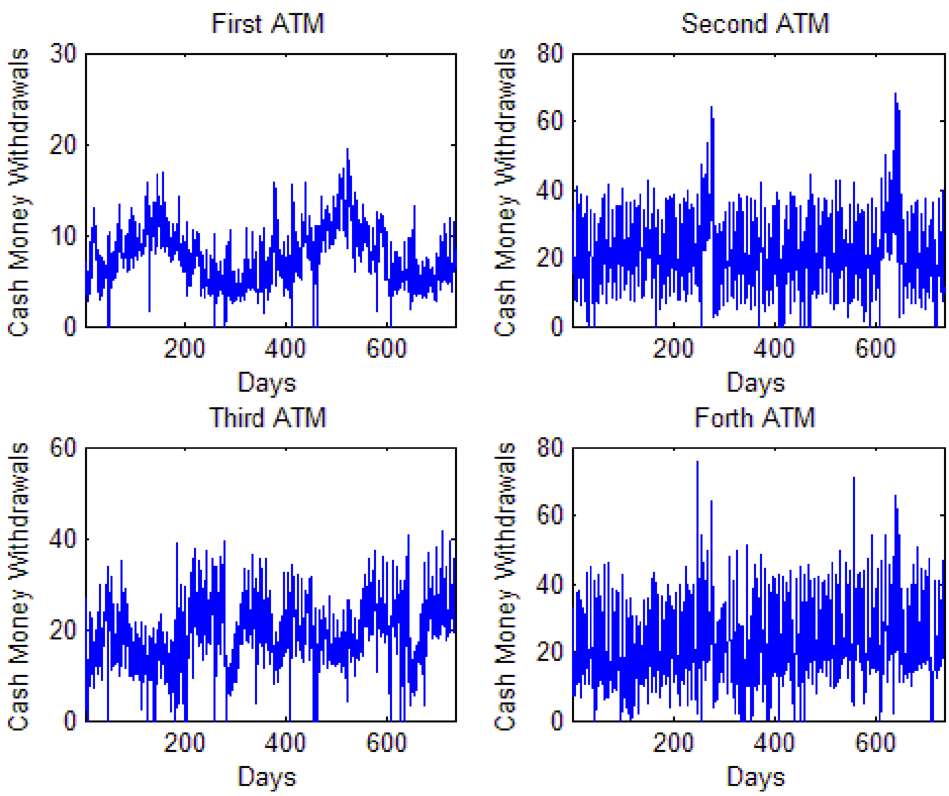
The rest of the paper is organized as follows. In section 2, the data are described. In section 3 the cash money withdrawals modeled non-parametrically using a wavelet neural network. A wavelet analysis is used in order to remove the noise and the outliers from the original time-series. The independent variables used for the network training were extracted from wavelet analysis. Then in-sample and out-of-sample forecasts presented. Finally, in section 4, we conclude.

2. Time-series description

The time series provided originate from different cash machines at different, randomly selected locations within England and are not related. All time series start on March 18, 1996 and run until March 22, 1998, providing two years of daily data. The data provided by the Neural Network Association and first presented in the NN5 competition.

In this section the properties of the time-series are discussed. Observing Figure 1 someone can conclude that time evolution of cash money withdrawals are similar to temperature simulation used in weather derivative pricing and electricity load. Cash money demand represents a non-stationary, heteroscedastic process. The volume of cash money withdrawals shows strong evidence of seasonality. As it is shown in next section there are multiple overlying seasonalities, local trends and structural breaks. Other parameters that affect cash money withdrawals are the reoccurring holiday periods, regional special events that occur in different time periods and with different magnitude and bank holidays which affect differently each ATM. A closer inspection on the data reveals outliers and missing values. The dynamics of the time-series, such as seasonality, trends or volume, are changing over time.

Figure 1: Cash demand from four different ATMs.



It is clear that the driving forces are different for each ATM. Since the only available information is the level of the cash money withdrawal each day, we use wavelet analysis to extract the underlying process of each ATM. Wavelets have the ability to decompose a signal or a time-series in different levels. As a result, this decomposition brings out the structure of the underlying signal as well as trends, periodicities, singularities or jumps that cannot be observed originally. In this study we use the Daubechies wavelet family that proved to perform better than other wavelet families, (Daubechies, 1992).

In (Zapranis & Alexandridis; 2007a,b) we give a concise treatment of wavelet theory. Here the emphasis is in presenting the theory and mathematics of wavelet neural networks and thus we give only the very basic notions of wavelets. Very briefly, a *family* of wavelets is constructed by translations and dilations performed on a single fixed function called the *mother wavelet*. A wavelet ψ_j is derived from its mother wavelet ψ by the relation:

$$\psi_j(x) = \psi\left(\frac{x - m_j}{d_j}\right) = \psi(z_j) \tag{2.1}$$

where its translation factor m_j and its dilation factor d_j are real numbers (d_j > 0).

In the context of process modeling applications, the wavelets are determined either by orthogonal wavelet decomposition or according to space-frequency analysis of the data. Another approach is to construct a feedforward wavelet neural network, which serves as a representation of a family of parameterized non-linear wavelet functions. The translation and dilation factors are real numbers and they are considered as network weights.

In bibliography two mother wavelets are suggested the Gaussian derivative

$$\psi(a) = -ae^{-\frac{1}{2}a^2} \tag{2.2}$$

and the second derivative of the Gaussian the so-called “Mexican Hat”

$$\psi(\alpha) = (1-a^2)e^{-\frac{1}{2}a^2} \tag{2.3}$$

Following Zhang (1994) we use as a mother wavelet the Mexican Hat function. Other mother wavelets can also be used.

The structure of a single-hidden-layer feedforward wavelet network is given in Figure 2. The network output is given by the following expression:

$$\hat{y}(x) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(x) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i \tag{2.4}$$

In that expression, $\Psi_j(\mathbf{x})$ is a multidimensional wavelet which is constructed by the product of m scalar wavelets, x is the input vector, m is the number of network inputs, λ is the number of hidden units and w stands for a network weight. The multidimensional wavelets are computed as follows:

$$\Psi_j(x) = \prod_{i=1}^{m} \psi(z_{ij}) \tag{2.5}$$

where

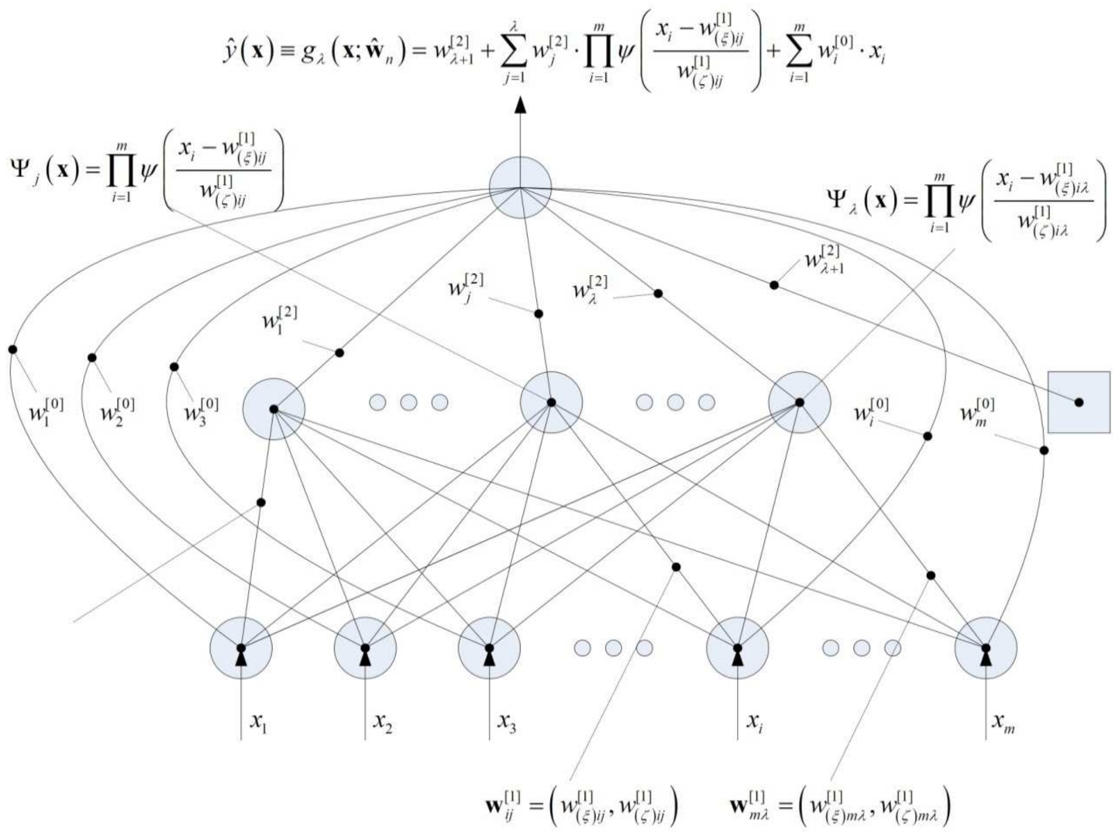
$$z_{ij} = \frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}} \tag{2.6}$$

In the above expression, i=1…m, j = 1…λ+1 and the weights w correspond to the translation ($w_{(\xi)ij}^{[1]}$) and the dilation ($w_{(\zeta)ij}^{[1]}$) factors. The complete vector of the network parameters comprises:

$$w = \left(w_i^{[0]}, w_j^{[2]}, w_{\lambda+1}^{[2]}, w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]} \right) \tag{2.7}$$

Here we have to note that the families of multidimensional wavelets preserve the universal approximation property that characterizes neural networks. For detailed exposition wavelet networks we refer to, for example Zhang *et al*. (1992), Oussar *et al*. (2000), Oussar e*t al*. (1998) and Zhang (1997).

Figure 2: The structure of a Wavelet Neural Network.

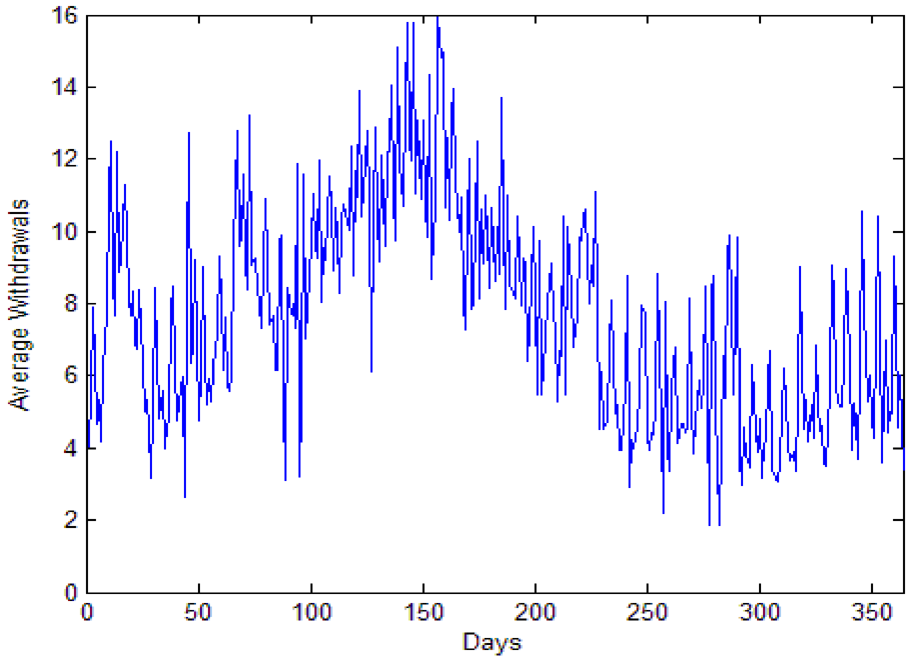


3. Methodology

Each series consists of data ranging from 18 of March 1996 until 22 of March 1998 resulting in 735 values. In order to overcome the data inconsistencies discussed earlier data is splitted in two groups according to the weekday of withdrawals. For example the Thursday, 2nd of January 1997 will be matched with the first Thursday of 1998, 1st of January. Each series is splitted in two vectors y_1 and y_2 where y_1 contains the observations from Tuesday 19 March 1997 until Monday 17 March 1997 and y_2 contains the observations from Tuesday 18 Match 1997 until Monday 16 March 1998. Both y_1 and y_2 have 364 data points. The first and the last six values of each time-series were not used. Next the vector *y* is formed where *y* is the average of y_1 and y_2.

The missing data and the outliers lead to misleading average values. Hence the corresponding observation removed from vector *y*. Figure 3 shows the average values of the first ATM. For simplicity we will refer only to the third ATM. The analysis and results for the rest of the time-series are similar.

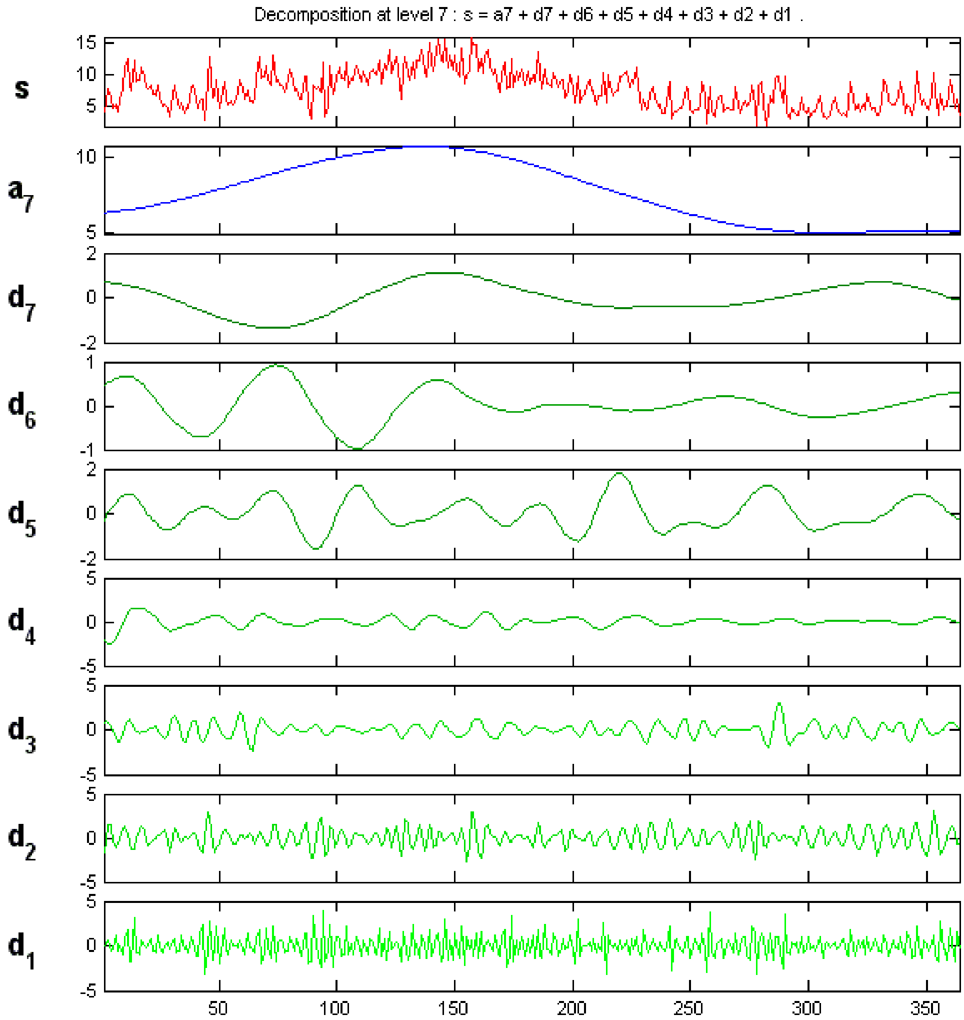
Figure 3: Average cash demand.



Next we use wavelet analysis in order to extract the underlying dynamics. Figure 4 shows the wavelet decomposition. The Daubechies 7 at level 7 wavelet was used. The wavelet transform decomposes the original signal into seven details and one approximation. It is clear that the approximation (a_7) captures the seasonality of one year that is also clear in Figures 1 and 3. The lower detail catpures the noise part of the original signal. Perfoming a single sample Kolmogorov-Smirnov goodness-of-fit hypothesis test at d_1 the *p-value*=0.056 leading to the acceptance of the hypothesis that d_1 follows a standard normal N(0,1) distribution at significanse level of 5%. In d_2 the weekly seasonality is shown which is also clear in Figure 4. Details 3, 4, 5 and 6 capture seasonalities that originaly cannot be observed. For exampple, at d_6 as seasonalityof two months is captured. However its effect is stronger in the beginning of the year and fades later on. If the two months seasonality is ignored forecasts in the beginning of the year will be understimated. Similarly, if the seasonality considered constant, forecasts at the end of the year will be overestimated. In d_2 and d_3 at days 45 and 287 two large spikes are shown representing an outlier or a jump at the time-series. These values if included will affect the network train and consequently the forecasts.

So far wavelets used to denoise the original signal and to extract the dynamics of the underlying cash withdrawals process of each ATM. Next the wavelet decomposition is used as an input to the wavelet neural network. First the data were rescaled to [-1,1] domain and were transformed according to equation (2.6). The vector *y* contains the target values.

Figure 4: Wavelet decomposition and the original signal.



One of the most crucial steps is to identify the correct topology of the network. A network with less hidden units than needed will not be able to learn the underlying function while selecting more hidden units than needed the network will overfit the data – the network will learn part of the noise. In order to select the correct network we use the cross-validation criterion.

In *v-fold* cross-validation from our initial training sample, of length *n*, we create *v* random sub-samples without replacement, D_i, of size *m*, where $i = 1,...,v$ and *m*<*n*. Here a 10% fold of the original training sample was used. Next the sub-samples D_i are removed one by one from the original sample D_n and a network is trained on the remaining data. Then the trained network is evaluated, on the removed sample, using the prediction risk measure. The network is evaluated using the averaged square errors function. The procedure is repeated for each hidden unit and the network with the smallest prediction risk is selected. Figure 5 shows optimal the number of hidden units selection and the prediction risk for the selected time-series.

Figure 5: Cross-Validation

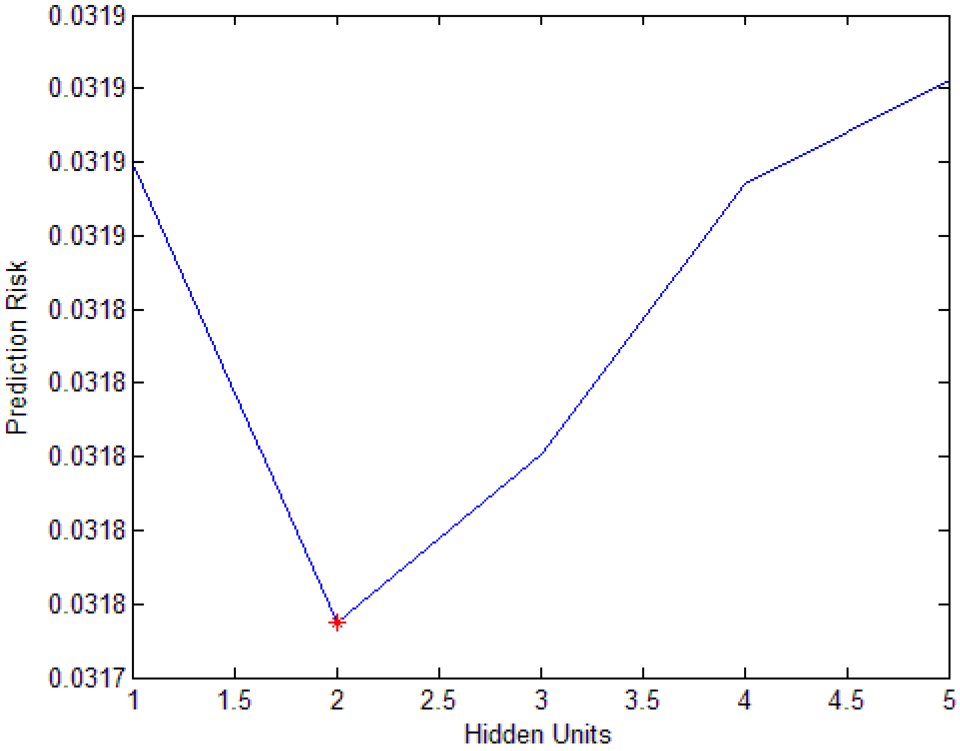


Table 1 shows the topology for all 11 time series while Figure 6 shows the original and the fitted data. It is clear that the network was able to learn the underlying process very well. Moreover as presented in Table 2 the POCID and IPOCID are 83.47% and 69.14% meaning that the network can predict the movement in changes of the cash money withdrawals. The MAE is only 0.97 when the maximum observation is 15.98 and the minimum is 1.86. Finally, performing a Kolmogorov-Smirnov goodness-of-fit hypothesis test we accept the hypothesis that the error follows a standard normal N(0,1) distribution. This means that the wavelet neural network acted as a second filter producing a denoised forecast.

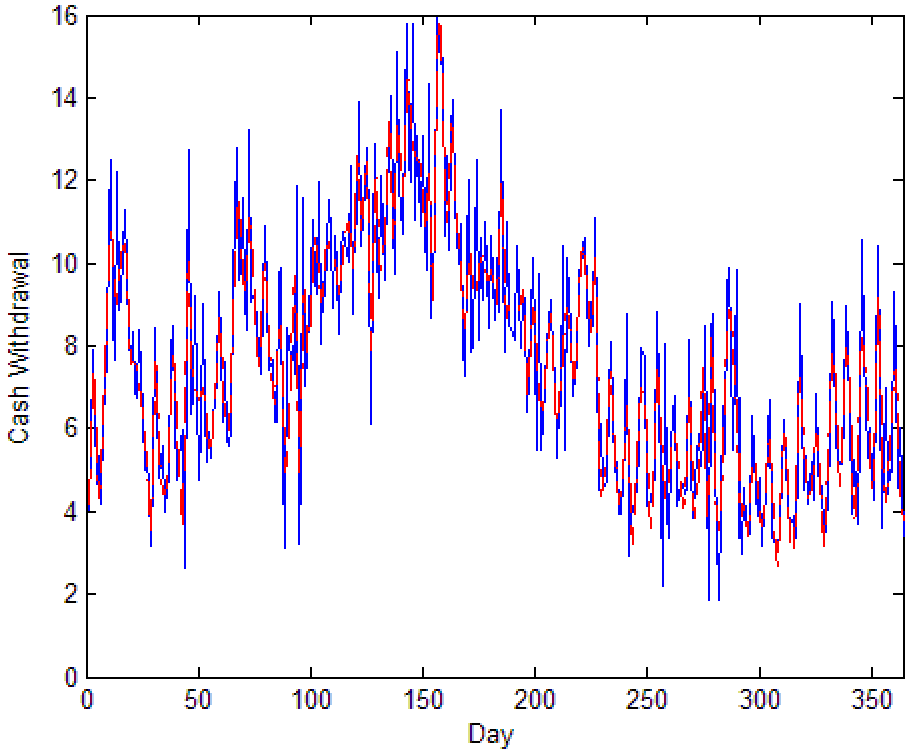
The next step is to produce out-of-sample forecasts. However the real data will be not available since the fall of 2008, hence we are not able to evaluate our method yet. Figure () shows the out-of-sample forecasts for the next 62 days.

Table 1: Networks Topology

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}
H.U.[1]	2	2	1	1	5	1	1	1	6	6	2

[1] H.U.=Hidden Units.

Figure 6: Original and fitted data.



4. Conclusions

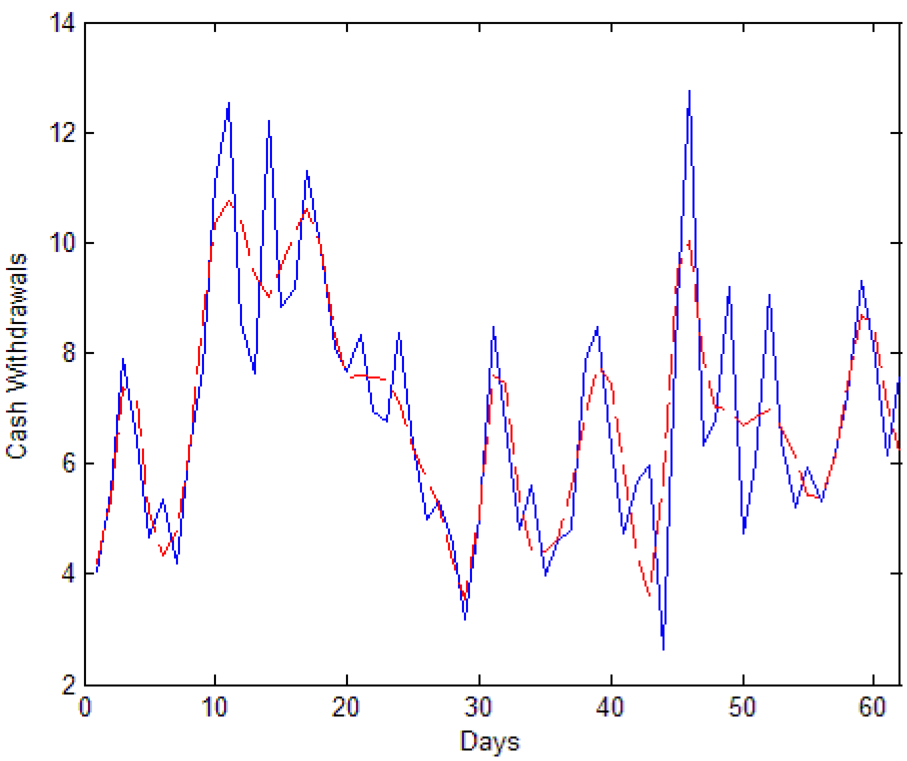
In this study a novel forecasting method presented. We used a wavelet analysis in order to decompose 11 different time series. The decomposition extracted the driving dynamics of the underlying process that leads the cash money withdrawals in the form of details and an approximation. Wavelet analysis was able to successfully capture and remove the noise from the original signal. In addition wavelet analysis indicated observations that should be removed from the training sample of the network. The remaining details and the approximation comprised the training sample to the wavelet network. Using the driving dynamics as inputs, results to a smaller network topology and less training time. Finally, in-sample and out-of-sample forecasts presented.

Table 2: Error Criteria.

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}
MdAE[1]	0.8293	3.4819	2.3326	4.6460	1.6819	3.0509	2.7363	2.2972	3.7196	1.9585	2.2251
MAE[2]	0.9741	4.5818	3.0056	5.1684	2.1778	3.9237	3.4423	2.8428	4.2804	2.4858	2.6070
MaxAE[3]	3.9331	20.7510	12.400	19.1574	13.3056	15.5776	17.7972	12.5829	20.8330	11.1435	17.1659
RMSE[4]	1.2245	0.3021	3.8119	6.2624	2.8759	5.0256	4.3710	3.6277	5.4708	3.2974	3.3644
NMSE[5]	0.1740	0.3021	0.2912	0.3104	0.2418	0.3528	0.3860	0.3319	0.2285	0.1222	0.2654
MSE[6]	1.4994	33.8928	14.5312	39.2181	8.2709	25.2575	19.1060	13.1603	29.9304	10.8731	11.3198
MAPE[7]	14.62%	33.13%	18.23%	-	14.94%	29.76%	32.62%	19.7130	19.74%	12.86%	-
SMAPE[8]	13.81%	25.87%	16.95%	28.23%	14.17%	23.68%	24.75%	17.7732	18.28%	12.17%	21.63%
POCID[9]	83.47%	86.22%	79.33%	80.71%	85.67%	83.74%	87.05%	89.80%	89.80%	88.70%	83.19%
IPOCID[10]	69.14%	67.&6%	66.66%	66.66%	79.06%	65.84%	65.84%	72.72%	77.13%	68.87%	72.72%

[1] Median Absolute Error
[2] Mean Absolute Error
[3] Maximum Absolute Error
[4] Root Mean Square Error
[5] Normalized Mean Square Error
[6] Mean Square Error
[7] Mean Absolute Percentage Error
[8] Symmetric Mean Absolute Percentage Error
[9] Prediction of Change in Direction
[10] Independent Prediction of Change in Direction

Figure 7: Forecasts



References

1. Daubechies, I. (1992). *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics.
2. Donoho, D. L., & Johnstone, I. M. (1994). Ideal Spatial Adaption by Wavelet Shrinkage. *Biometrika , 81*, pp. 425-455.
3. Donoho, D. L., & Johnstone, I. M. (1998). Minimax Estimation Via Wavelet Shrinkage. *Annals of Statistics , 26*, pp. 879-921.
4. Gao, R., & Tsoukalas, H. I. (2001). Neural-wavelet Methodology for Load Forecasting. *Journal of Intelligent & Robotic Systems , 31*, pp. 149-157.
5. Khayamian, T., Ensafi, A. A., Tabaraki, R., & Esteki, M. (2005). Principal Component-Wavelet Networks as a New Multivariate Calibration Model. *Analytical Letters , 38* (9), pp. 1447-1489.
6. Mallat, S. G. (1999). *A Wavelet Tour of Signal Processing.* San Diego: Academic Press.
7. Oussar, Y., & Dreyfus, G. (2000). Initialization by Selection for Wavelet Network Training. *Neurocomputing , 34*, pp. 131-143.
8. Oussar, Y., Rivals, I., Presonnaz, L., & Dreyfus, G. (1998). Trainning Wavelet Networks for Nonlinear Dynamic Input Output Modelling. *Neurocomputing , 20*, pp. 173-188.
9. Postalcioglu, S., & Becerikli, Y. (2007). Wavelet Networks for Nonlinear System Modelling. *Neural Computing & Applications , 16*, pp. 434-441.
10. Subasi, A., Alkan, A., Koklukaya, E., & Kiymik, M. K. (2005). Wavelet Neural Network Classification of EEG Signals by Using AR Model With MLE Pre-processing. *Neural Networks , 18*, pp. 985-997.
11. Szu, H., Telfer, B., & Kadambe, S. (1992). Neural Network Adaptive Wavelets for Signal Representation and Classification. *Opt. Engineering , 31*, pp. 1907-1916.
12. Thuillard, M. (2000, September 14-15). A Review of Wavelet Networks, Wavenets, Fuzzy Wavenets and Their Applications. *In proc. of ESIT* .
13. Zapranis, A., & Alexandridis, A. (2007b). Modelling Temperature Time Dependent Speed of Mean Reversion in the Context of Weather Derivetive Pricing. *Applied Mathematical Finance* (accepted, to appear).
14. Zapranis, A., & Alexandridis, A. (2007a). Weather Derivatives Pricing: Modelling the Seasonal Residuals Variance of an Ornstein-Uhlenbeck Temperature Process With Neural Networks. *Neurocomputing* (accepted, to appear).
15. Zhang, Q. (1997). Using Wavelet Network in Nonparametric Estimation. *IEEE Trans. Neural Networks , 8* (2), pp. 227-236.
16. Zhang, Q., & Benveniste, A. (1992). Wavelet Networks. *IEEE Trans. Neural Networks , 3* (6), pp. 889-898.