FORECASTING CASH MONEY WITHDRAWALS USING WAVELET ANALYSIS AND WAVELET 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

The increasing demand for easily accessible cash drives banks to expand their Automatic Teller Machine networks. As the network increase it becomes more difficult to supervise it while the operating costs rise significantly. Cash demand needs to be forecasted accurately so that banks can avoid storing extra cash money and can profit by mobilizing the idle cash. This paper is motivated by the Neural Network Association and the NN5 competition. The objective of the paper is to describe a unique, non-supervising method for forecasting cash money withdrawals in different ATMs. More precisely, the data consists of 2 years of daily cash money demand at various ATMs at different randomly selected locations across England. The only available information is the total cash withdrawals in each ATM at the end of each day. Having limited domain knowledge and no information on the causal forces we use wavelet analysis to extract the dynamics of the underlying process of each ATM. Next wavelet neural networks were used in order to find the true generating process of each ATM and to forecast the cash money demand up to 56 day ahead. The performance of the proposed technique is evaluated using various error and fitting criteria.

Keywords: Wavelet Networks, Forecasting

1. Introduction

The increasing demand for easily accessible cash drives banks to expand their Automatic Teller Machine (ATM) networks. As a result the number of ATMs has increased steadily after the early 1980s, (Snellman and Virén, 2006). However the larger an ATM network becomes, the more difficult becomes to supervise it while the operating costs rise significantly. Using optimization methods, banks can avoid storing extra cash money and can profit by mobilizing the idle cash, (Simutis *et al.*, 2007). Cash demand needs to be forecasted accurately similarly to other products in vending machines. If the forecasts are flawed, they induce costs. If the forecasts are too high unused money is stored in the ATM incurring costs to the bank, on the other hand if the ATM runs out of cash customers will be dissatisfied while significant profit will be lost.

This paper is motivated by the Neural Network Association and the NN5 competition[1]. The objective of the paper is to describe a unique, non-supervising method for forecasting cash money withdrawals in different ATMs. More precisely, the data consists of only 2 years of daily cash money demand at various ATMs at different locations in England and the objective is to forecast the cash money demand for the next 56 days. The only available information to competitors was the cash demand in each ATM at the end of each day. In order to avoid overfitting, the out-of-sample data was not known until the end of the competition.

As it shown in the next section cash money demand represents a non-stationary process. The generating process of each ATM is unique while the cash demand of each ATM depends on its physical location. The time-series exhibits trends, singularities and seasonal, periodical and irregular structural components of the data while missing values and outliers are common among the time-series. ATM withdrawals affected by reoccurring holiday periods, regional events of different size and impact and bank holidays of different lead and lag effects. The data provided originate by randomly selected ATMs in unknown locations in England making impossible to identify all the above parameters.

Having limited domain knowledge and no information on the causal forces we use wavelet analysis to extract the dynamics of each process of each ATM. 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 wavelet analysis can be used to denoise the original time-series (Donoho and Johnstone, 1994 and Donoho and Johnstone, 1998) while the assumption of stationarity can be avoided, (Mallat, 1999). While in other frameworks like the Kalman-Filtering (KF) setting the stationarity assumption can be avoided, KF is applicable only to linear or nearly linear problems, (Brown and Hwang, 1992).

Next wavelet networks were used in order to identify the true underlying process of each ATM and produce 1- to 56-step ahead forecasts. Wavelet networks allow constructive procedures that efficiently initialize the parameters of the network. These procedures allow the wavelet network to converge to a global minimum of the cost

[1] http://www.neural-forecasting-competition.com

function and lead to smaller network topologies and smaller training times. Using wavelet analysis and wavelet networks model assumptions are not needed and the framework is not necessarily based on Gaussian errors.

In order to evaluate our method we produce out-of-sample forecasts in 11 different time series. The rest of the paper is organized as follows. In section 2, the data is 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 from the original time-series. The independent variables used for the network training were extracted from wavelet analysis. Then 1- to 56- step ahead out-of-sample forecasts presented. The proposed methodology was compared with a linear approach. Finally, in section 4, we conclude.

2. Time-series description

In this section the available dataset is described and the main statistics of the cash money withdrawals are presented. The time series provided originate from 11 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 only two years of daily data resulting to 735 values. The data provided by the Neural Network Association and first presented in the NN5 competition. The aim of the competition is to forecast 1 to 56 step ahead. The competition focus on 11 "difficult" time series and no other information was given beyond the daily withdrawals for the two years. The out of sample was provided after the end of the competition in order to avoid overfitting and all the data was linearly scaled to ensure the anonymity of the time series.

Table 1 shows the descriptive statistics of all ATM's. The standard deviation is large for all 11 ATM's while in the nine of total eleven time-series values of zero appear. A closer inspection of the data reveals zeroes, outliers and missing values. Moreover missing values and outliers appear in periods corresponding to the forecasting horizon. Observing Figure 1 someone can conclude that the volume of cash money withdrawals shows strong evidence of seasonality and periodicities. In addition, Figure 2 shows the autocorrelation and partial autocorrelation function of the first ATM where a periodicity of seven days is clear. The results for the rest of the time-series are similar

The cash money demand represents a non-stationary process. Performing an Augmented Dickey-Fuller unit root test it can be shown that non-stationarity was a common problem among the original time-series. The results can be found on Table 2. From Figure 1 a linear trend is clear in several time-series. Table 3 shows the two parameters of the linear fitting for each ATM and the corresponding p-values. Next, the mean and the upward trend were removed from the data.

3. Methodology

As it was shown in the previous section and from Figure 2 a strong periodicity of seven days is present. Cash withdrawals depend on the day of the withdrawal. In other words the level of withdrawals each Monday is similar but different from the level of withdrawals each Friday for example. Having only two years of data the noise levels are high. In order to smooth out the data we split the data 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 was split in two vectors y_1 and y_2 where y_1 contains the observations from Tuesday 25 March 1996 until Monday 23 March 1997 and y_2 contains the observations from Tuesday 24 March 1997 until Monday 22 March 1998. Both y_1 and y_2 have 364 data points. Next the vector y is formed where y is the average of y_1 and y_2. Since we want two equal samples, the first seven observations were not used.

The missing data and the zeroes lead to misleading average values. If a missing value appears in y_1 then this value and the corresponding one in y_2 are removed. Hence the corresponding observation removed from vector y. Figure 3 shows the detrended average values of the first ATM. For simplicity we will refer only to the first ATM. The analysis and results for the rest of the time-series are similar.

Next outliers on the data were indentified. Outliers indified using the leverage value of each observation in vector y. The leverage value, h_t of each observation is calcualted by the t^{th} diagonal element of the 'hat' matrix $H = y(y'y)^{-1}y$ and its value lie between 0 and 1. An observation is regarded as an outlier when its leverage exceeds three time the average leverage, p/n, where p is the number of parameters in the model and n the sample size. In Table 4 the number of outliers, missing values, zeroes as well as the number of final observations in each ATM are presented.

Next we use wavelet analysis in order to extract the underlying dynamics. Figure 4 shows the wavelet decomposition. A wavelet is a mathematical function used to divide a given function or continuous-time signal into different frequency components and study each component with a resolution that matches its scale. A wavelet transform is the representation of a function by wavelets. (Mallat, 1999).

In Zapranis & Alexandridis (2008) and Zapranis & Alexandridis (2007) 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:

$$\psi_j(x) = \psi\left(\frac{x - m_j}{d_j}\right) = \psi(z_j) \tag{3.1}$$

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

In this study the Daubechies 7 at level 7 wavelet was used since sevel levels of decomposition needed to extract the one year cycle of the data. The Daubechie family wavelets have many good properties and proved to perfrom very well in various problems, (Daubechies, 1992). Other wavelet families can be used. The wavelet transform decomposes the original signal into seven details and one approximation. It is clear that the approximation (a_7) captures the periodicity of one year that is also clear in Figure 4. The lower detail, d_1 catpures the noise part of the original signal that must be removed. In d_2 counting the distance between two spikes reveals a weekly periodicity as it was expected from Figure 2. Details 3, 4, 5 and 6 capture periodicities that originaly cannot be observed. For example, at d_6 a periodicity of two months is captured. However its effect is stronger in the beginning of the year and fades later on. If the two months periodicity is ignored forecasts in the beginning of the year will be understimated. Similarly, if the periodicity considered constant, forecasts at the end of the year will be overestimated.

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, except the lower detail d_1, of each time-series is used as an input to the wavelet neural network. Wavelet Network acts as a second filter which infers the true underlying function of the cash withdrawals in each ATM.

In this study we use a multidimensional wavelet neural network with a linear connection of the wavelons (hidden units) to the output. Moreover in order for the model to perform well in linear cases we use direct connections from the input layer to the output layer. A network with zero hidden units (*HU*) is the linear model.

The structure of a single hidden-layer feedforward wavelet network is given in Figure 5. As it is shown in Figure 5 the wavelet network is separated in three layers. The lower level is called the input layer where the input units receive information outside the network. The middle layer is called the hidden layer where the multidimensional wavelets are calculated and each node is called a wavelon or hidden unit. The upper layer is called the output layer. The network output is given by the following expression:

$$\hat{y}(\mathbf{x}) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i \tag{3.2}$$

In that expression, $\Psi_j(\mathbf{x})$ is a multidimensional wavelet (wavelon) which is constructed by the product of m scalar wavelets, $\mathbf{x}$ 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 parameter m is known and is equal to the number of the input variables. The parameter λ is very crucial to the network performance and is estimated using the cross-validation criterion described later. The multidimensional wavelets are computed as follows:

$$\Psi_j(\mathbf{x}) = \prod_{i=1}^{m} \psi(z_{ij}) \tag{3.3}$$

where ψ is the mother wavelet and

$$z_{ij} = \frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}} \tag{3.4}$$

In the above expression, $i = 1, \ldots, m$, $j = 1, \ldots, \lambda+1$ 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:

$$w = \left(w_i^{[0]}, w_j^{[2]}, w_{\lambda+1}^{[2]}, w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]} \right) \tag{3.5}$$

Following Zhang (1997) we use as a mother wavelet the second derivative of the Gaussian the so-called "Mexican Hat" function:

$$\psi(z_{ij}) = (1 - z_{ij}^{\ 2}) e^{-\frac{1}{2} z_{ij}^{\ 2}} \tag{3.6}$$

Other wavelets can be used depending on the application

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

The final approximation and all details, except the noise part, d_1, were used as an input to the wavelet neural network. Since the Mexican Hat's set of values lies in the interval (-1,1) the target values, the vectors y after removing the outliers, were rescaled to (-1,1) domain. By rescaling the input values in the same domain helps to reduce the training times of the network and increase the wavelet network performance.

One of the most crucial steps is to identify the correct topology of the network i.e. to find the correct number λ in equation (3.2) that minimizes the prediction risk. 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 that proved to outperform other techniques (Zapranis and Refenes, 1999, Efron and Tibshirani, 1993).

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 k, where $i=1,\dots,v$ and $k<n$. Here a 2% fold of the original training sample was used forming 50 samples from each time-series. 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. Hence, selecting an appropriate value for λ a stable network that learns the true underlying of function of a signal can be constructed while at the same time the model and prediction risk is minimized. Table 5 shows the number of hidden units needed to model the cash withdrawals in each ATM.

Each network is trained using past data, then the trained network can be used to produce forecasted values for the cash withdrawal in each ATM. Here we forecast 1- to 56- steps ahead. Figure 6 presents the real and forecasted out-of-sample values for all 11 ATMs. Examining Figure 6 it is clear that the real and forecasted values are very close in most cases. Also the weekly periodicity successfully captured from the wavelet network. However the fitting is very poor for 1^{st}, 6^{th} and 10^{th} ATM.

To account for a different number of observations in the individual data sub-samples of training and test set, and the different scale between individual series we propose to use a mean percentage error metric, which is also established best-practice in industry and in previous competitions[2]. The evaluation of each competitor was based on the mean Symmetric Mean Absolute Percent Error (*SMAPE*) across all time series. Table 6 presents the *SMAPE* for each ATM as well as 9 more error criteria for testing the performance of the wavelet networks.

Since the structure of the time-series is irregular and the noise level is high, the error criteria are high as it was expect. The average Symmetric Mean Absolute Percentage Error (*SMAPE*) for the 11 ATMs is 27.92. The Mean Absolute Error (*MAE*) is always less than 5.62 when the spread between the maximum and minimum out-of-sample observation is over 30 for each ATM.

Moreover as presented in Table 6 the Percentage of Change in Direction (*POCID*) and Independent *POCID* (*IPOCID*) are very high. Leaving out the 10th ATM, the POCID criterion is above 60% in all cases with maximum 90.91% in the ninth ATM. Similarly the IPOCID criterion is above 60% in all cases with maximum 83.64% in the ninth ATM meaning that the WN can successfully predict the movement in the changes of the cash money withdrawals.

In order to evaluate our methodology different linear models of the ARMA family were fitted in each time-series. The out-of-sample results of the linear approach can be found on Table 7. Examining Table 7 it is clear that the wavelet networks outperform the linear approach. Only for the 1st and the 3rd ATM the linear approach seems perform better. However, the IPOCID criterion is only 34.54% and 41.82% for the two ATMs. Examining the real and the forecasted values in Figure 7 it is clear that the predictions are not good even if the error criteria are smaller than the ones in the case of the wavelet network. In general from Figure 7 someone can conclude that the performance of the linear models is very poor.

4. Conclusions

In this study a novel forecasting method presented. The particular problem designed by the Neural Network Association for the NN5 Competition. Having in our possession only two years daily cash withdrawals in 11 randomly selected ATM across England we produced 1- to 56-step ahead forecasts. The original data were irregular while missing values, outliers and observations of zero appear, especially in periods corresponding to the forecasting horizon. Moreover the forecasting window was very large. We used 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. The remaining details and the approximation comprised the training sample to the wavelet network. Using the driving dynamics as inputs to the wavelet network, results to a smaller network topology and less training time while the

[2] Evaluation method of the NN5 Competition.

changes in the dynamics of different periodicities can be captured. Finally, out-of-sample forecasts presented. From the results it is clear that the wavelet network can successfully capture different periodicities and can predict the change in direction of the cash money withdrawals even in the cases where the volume of withdrawals were not forecasted satisfactory. Moreover the wavelet network framework outperformed the linear approach which proved to be inappropriate for predicting cash money withdrawals.

The proposed technique proved to be useful in predicting cash money withdrawals. However the limited information and the design of the problem lead to large out-of-sample errors. Having more information such as the number of withdrawals, more data, the physical location of each ATM and the different regional effects in each ATM, the results can significantly improved.

Acknowledgements

We would like to thank the anonymous referee for the constructive comments that improved substantially the final version of this paper.

Figure 1: Cash demand from 11 different ATM's.

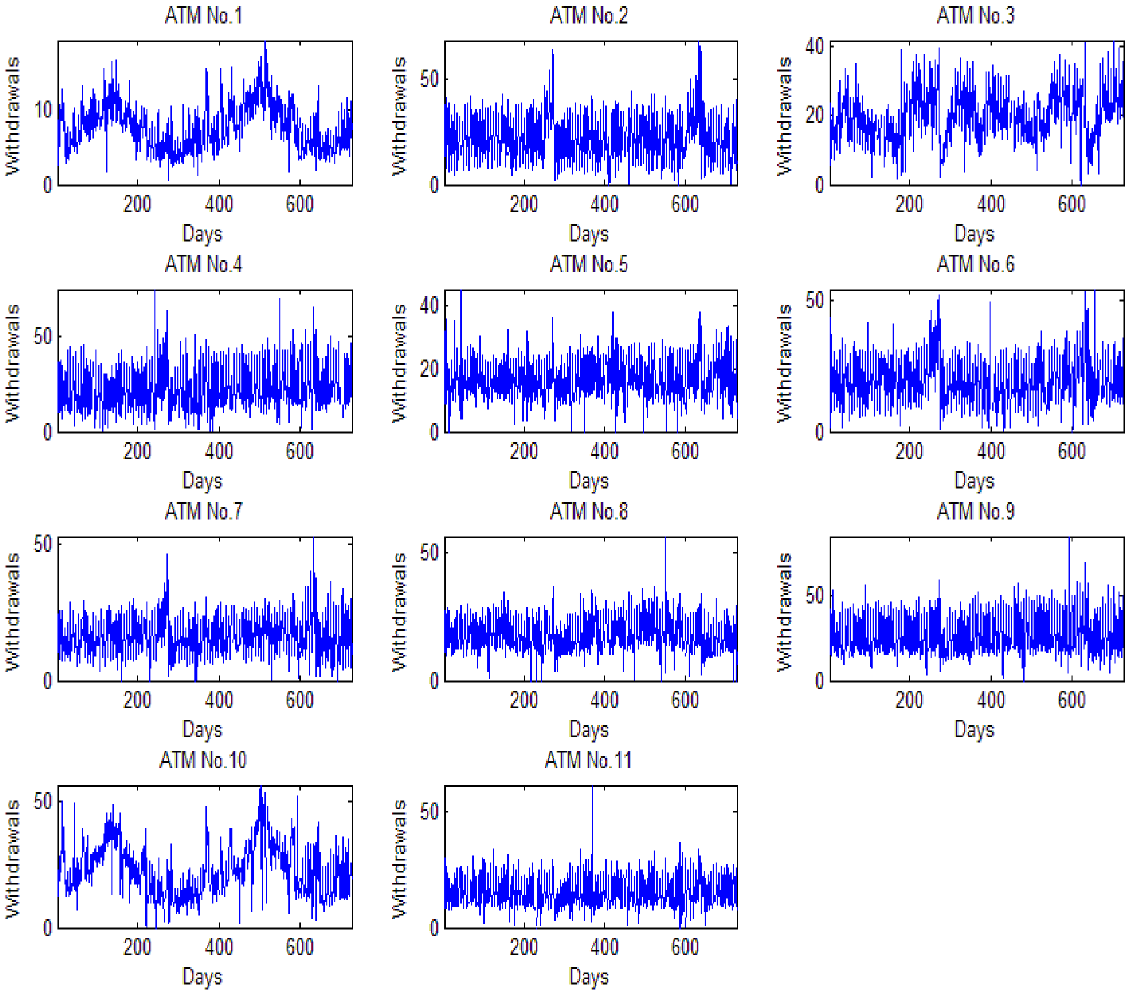

Figure 2: Autocorrelation and partial-autocorrelation function for the first ATM.

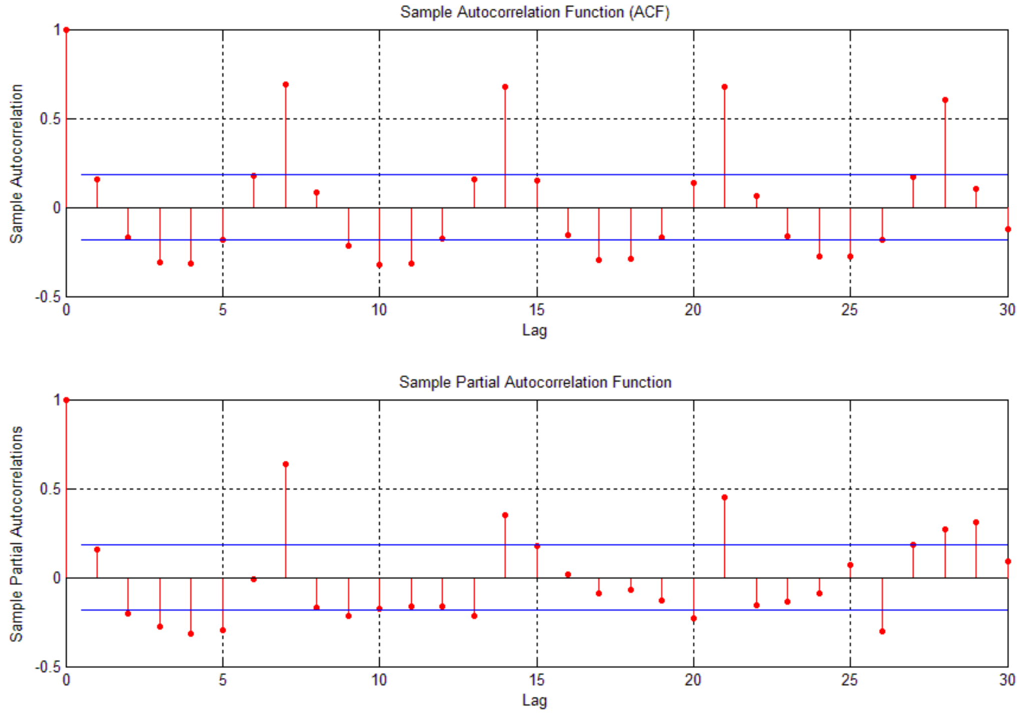

Figure 3: Average cash demand for the first ATM.

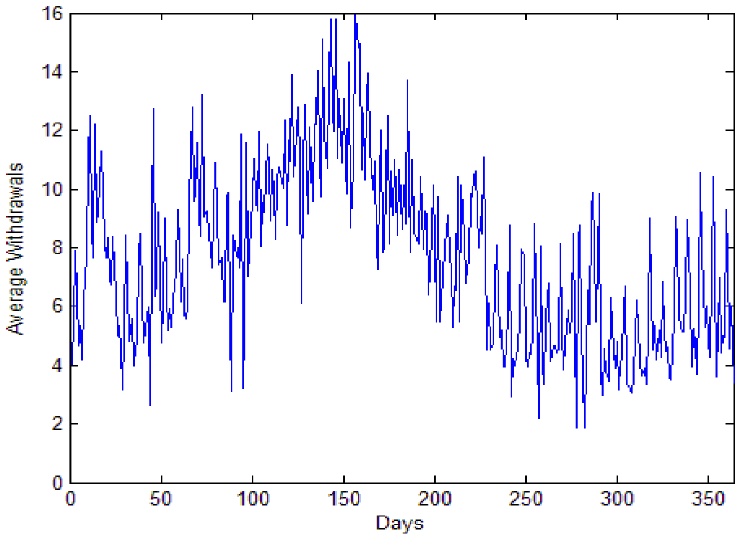

Figure 4: The Wavelet decomposition, the details d_1-d_7, the approximation a_7 and the synthesized signal s

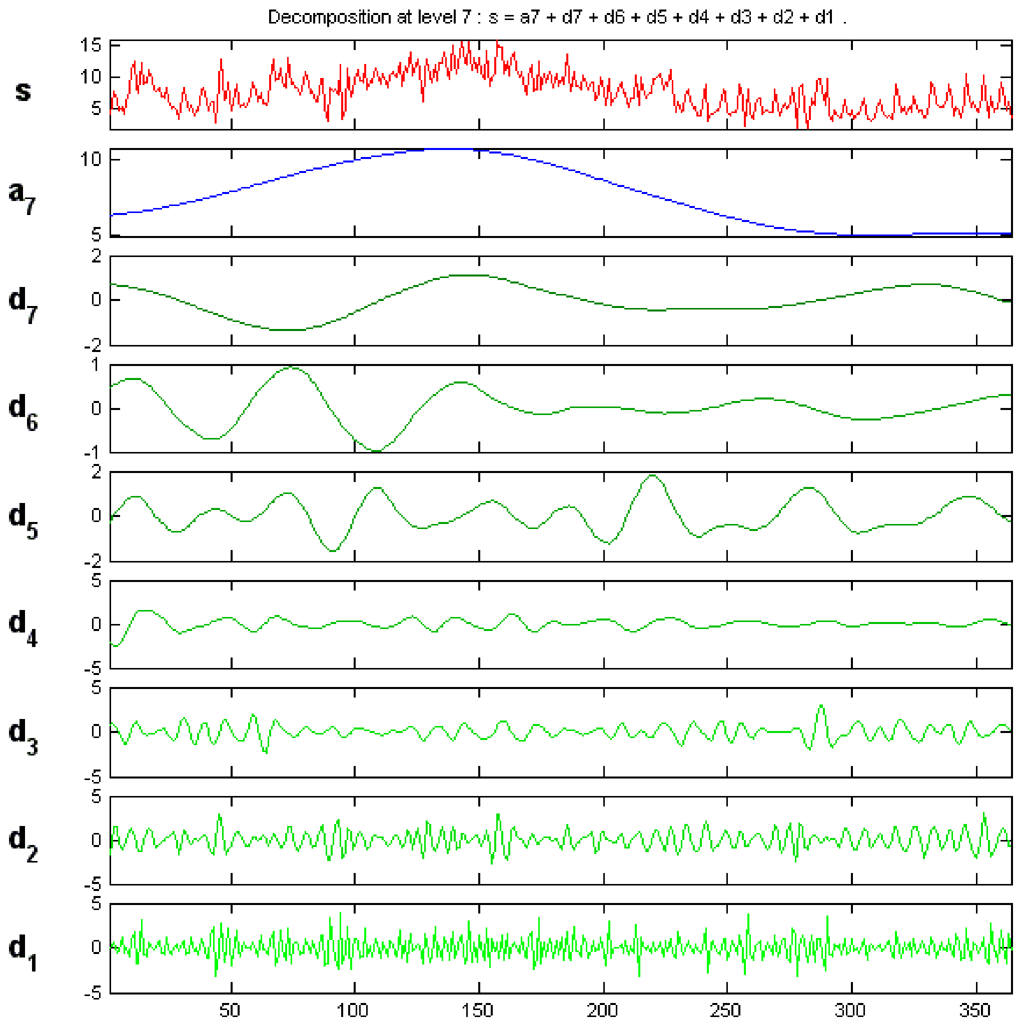

Figure 5: The structure of a wavelet neural network.

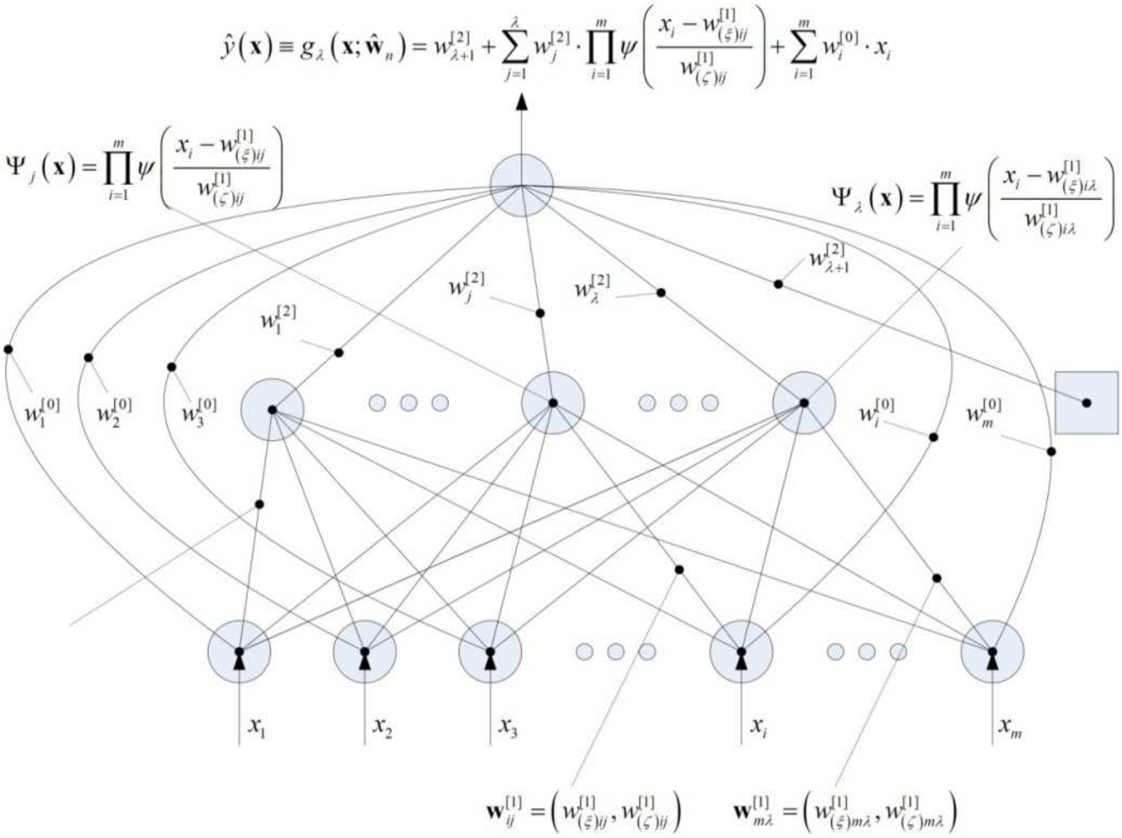

Figure 6: Out of sample forecasts using wavelet networks. Real data (blue line) and forecasts (red line)

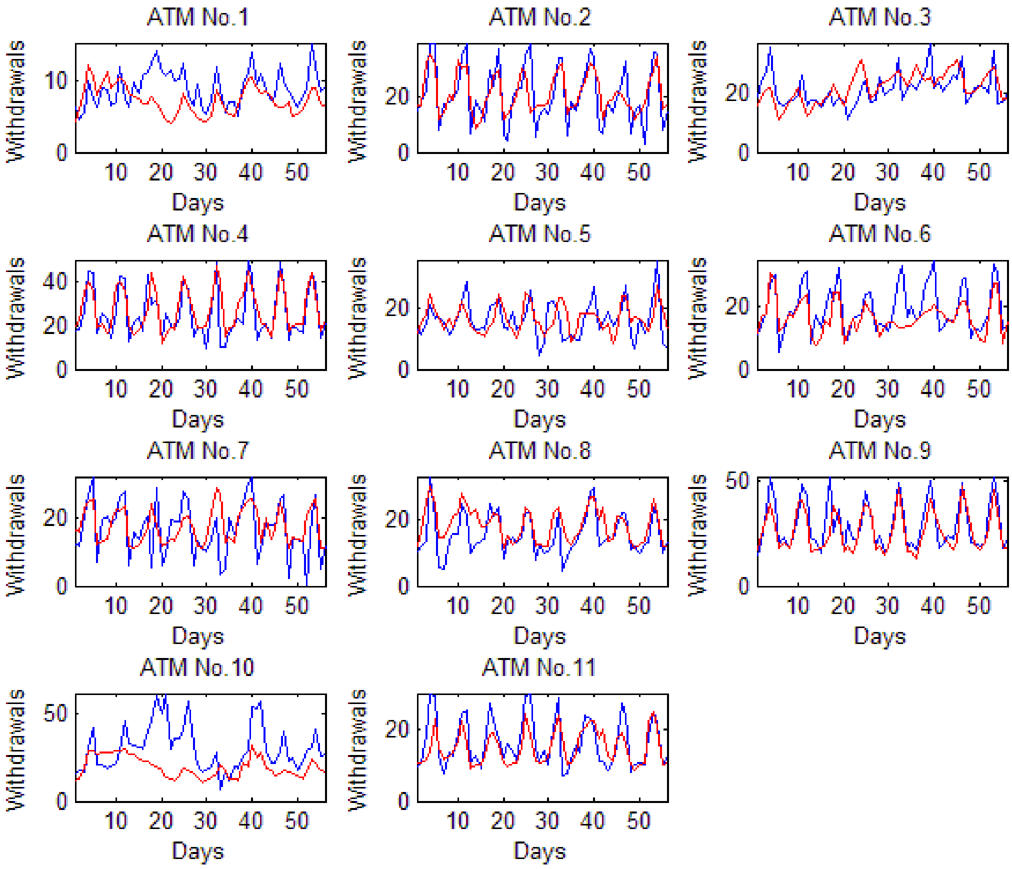

Figure 7: Out of sample forecasts using linear models. Real data (blue line) and forecasts (red line)

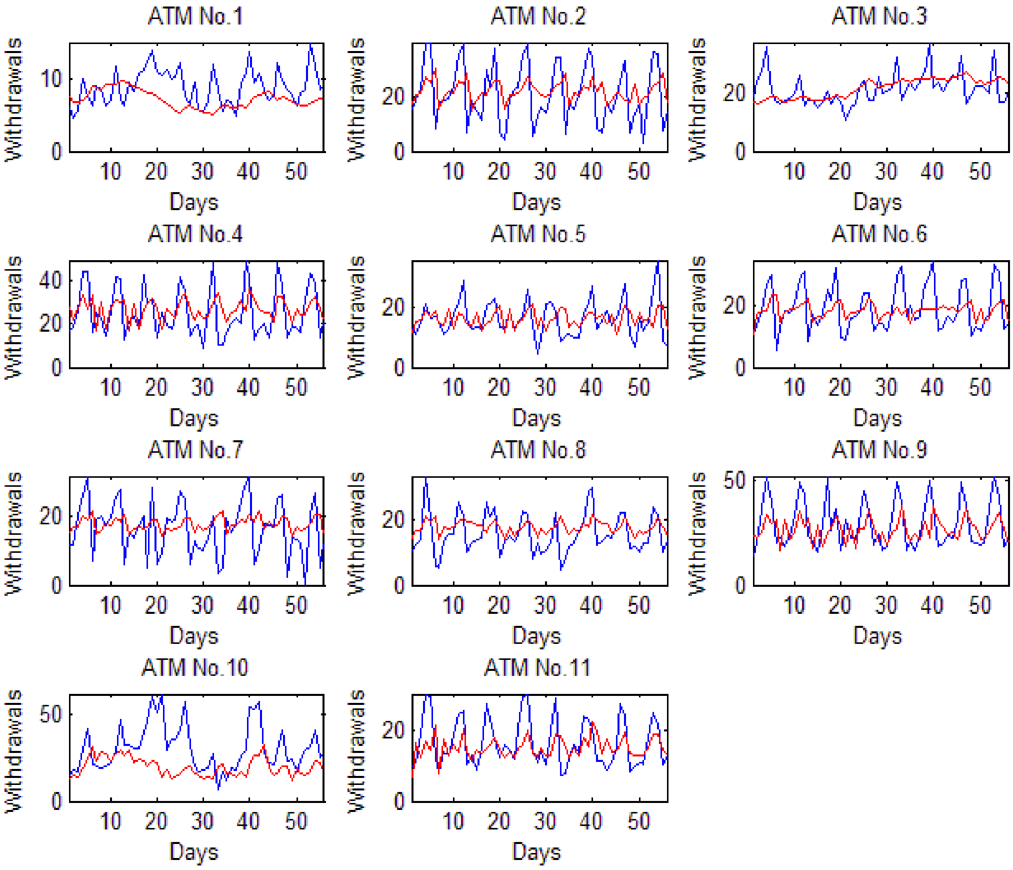

Table 1. Descriptive Statistics.

	Mean	St.Dev.	Max.	Min.	Skewness	Kurtosis
X1	7.743	3.177	19.48	0.652	0.545	2.913
X2	22.426	11.090	68.03	0	0.621	3.603
X3	19.320	7.237	41.60	0	0.228	3.021
X4	23.311	12.173	75.45	0.113	0.716	3.332
X5	16.150	6.562	44.92	0	0.560	3.501
X6	19.509	9.190	53.89	0	0.602	3.485
X7	16.783	7.353	52.97	0	0.573	5.204
X8	17.507	6.841	56.29	0	0.514	4.221
X9	26.554	11.977	85.20	0	0.767	3.294
X10	22.797	10.424	56.21	0	0.459	2.692
X11	15.178	6.818	61.89	0	0.918	5.525

Table 2. Unit root tests.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
P-Value	0. 3250	0.0881	0. 6475	0.4202	0.2018	0.1179	0.4678	0.2111	0.2092	0.318	0.4149
Unit Root	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 3. Coefficients of the linear trend in each ATM.

	a	P-Value	*b*	P-Value
X1	7.74298	0.0000	-	-
X2	22.4259	0.0000	-	-
X3	17.0259	0.0000	0.006434	0.0000
X4	20.2877	0.0000	0.008468	0.0001
X5	16.1493	0.0000	-	-
X6	19.5089	0.0000	-	-
X7	15.7236	0.0000	0.002935	0.0249
X8	17.5069	0.0000	-	-
X9	26.5541	0.0000	-	-
X10	23.9871	0.0000	-0.003302	0.0873
X11	15.1783	0.0000	-	-

The linear trend: $a+bt$

Table 4. Final observations for each ATM.

	Missing Values	Outliers	Zeroes	Final Obs.
X1	22	10	1	331
X2	18	8	2	336
X3	10	7	0	347
X4	20	6	0	338
X5	10	8	6	340
X6	16	11	1	336
X7	14	5	4	341
X8	9	9	7	339
X9	13	4	1	346
X10	15	5	2	342
X11	15	7	2	340

Table 5. Network topology for each ATM.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
Hidden Units	9	13	2	13	9	13	14	13	14	1	11

Table 6: Out of sample results using wavenet networks.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
MdAE[1]	4.403	2.047	3.824	3.782	3.507	3.147	4.176	2.723	4.207	8.478	2.577
MAE[2]	5.620	2.628	4.198	5.090	4.407	4.329	5.220	3.806	5.180	12.809	3.640
MaxAE[3]	24.425	7.370	13.987	29.160	14.319	19.390	23.887	14.425	27.694	48.325	16.662
RMSE[4]	7.186	3.187	5.364	7.263	5.532	6.016	7.142	5.071	6.924	16.608	4.983
NMSE[5]	0.448	1.684	0.982	0.400	0.827	0.621	0.783	0.657	0.372	1.608	0.529
MSE[6]	51.646	10.159	28.779	52.758	30.607	36.195	51.011	25.721	47.939	275.824	24.782
MAPE[7]	49.787	28.459	20.407	27.154	34.562	22.402	577.620	34.973	17.163	40.035	22.782
SMAPE[8]	33.122	32.425	18.935	21.225	28.286	24.041	36.87%	24.746	18.306	47.658	22.289
POCID[9]	81.82%	60.00%	78.12%	87.27%	65.45%	70.91%	76.36%	89.09%	90.91%	54.54%	83.64%
IPOCID[10]	61.82%	69.09%	67.28%	76.36%	60.00%	65.45%	72.73%	69.09%	83.64%	67.28%	70.91%

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

Table 7: Out of sample results using a linear model.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
MdAE[1]	2.269	8.360	2.373	7.554	4.036	4.982	5.342	4.177	7.832	9.196	3.758
MAE[2]	2.516	8.284	3.876	8.823	4.805	5.744	6.605	4.954	8.476	12.710	4.849
MaxAE[3]	8.607	22.055	17.472	24.404	15.064	17.998	17.583	16.362	27.189	48.099	15.186
RMSE[4]	3.221	10.515	5.346	10.687	5.939	7.446	8.372	6.183	10.766	16.677	6.299
NMSE[5]	1.720	0.960	0.976	0.867	0.953	0.952	1.077	0.976	0.901	1.622	0.845
MSE[6]	10.376	110.57	28.612	114.2	35.274	55.447	70.101	38.228	115.910	278.140	39.687
MAPE[7]	26.160	79.077	18.223	45.265	35.421	35.821	490.700	44.943	29.659	38.900	30.649
SMAPE[8]	29.532	42.809	17.488	34.953	30.110	29.384	44.303	31.688	29.309	46.241	29.002
POCID[9]	63.63%	74.55%	74.54%	81.82%	69.09%	70.91%	61.82%	80.00%	70.91%	52.72%	78.18%
IPOCID[10]	34.54%	41.82%	54.54%	54.54%	70.91%	47.27%	50.91%	49.09%	42.27%	50.91%	65.46%

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

References

1. Brown, R.G., and P.C.Y. Hwang. *Introduction to random signals and applied Kalman filtering.* New York: J. Wiley and Sons, 1992.
2. Daubechies, I. *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics, 1992.
3. Donoho, D. L., and I. M. Johnstone. "Minimax Estimation Via Wavelet Shrinkage." *Annals of Statistics*, 1998: 879-921.
4. Donoho, D. L., and I. M. Johnstone. "Ideal Spatial Adaption by Wavelet Shrinkage." *Biometrika*, 1994: 425-455.
5. Efron, B., and R. J. Tibshirani. *An Introduction to the Bootstrap.* USA: Chapman & Hall, 1993.
6. Mallat, S. G. *A Wavelet Tour of Signal Processing.* San Diego: Academic Press, 1999.
7. Oussar, Y., and G. Dreyfus. "Initialization by Selection for Wavelet Network Training." *Neurocomputing*, 2000: 131-143.
8. Oussar, Y., I. Rivals, L. Presonnaz, and G. Dreyfus. "Trainning Wavelet Networks for Nonlinear Dynamic Input Output Modelling." *Neurocomputing*, 1998: 173-188.
9. Simutis, R., D. Dilijonas, L. Bastina, J. Friman, and P. Drobinov. "Optimization Of Cash Management For ATM Network." *Information Technology And Control*, 2007: 117-121.
10. Snellman, H., and M. Virén. "ATM networks and cash usage." *Bank of Finland Research Discussion Papers*, 2006: 1-33.
11. Zapranis, A., and A. Alexandridis. "Modelling Temperature Time Dependent Speed of Mean Reversion in the Context of Weather Derivetive Pricing." *Applied Mathematical Finance*, 2008: 355 - 386.
12. Zapranis, A., and A. Alexandridis.. "Weather Derivatives Pricing: Modelling the Seasonal Residuals Variance of an Ornstein-Uhlenbeck Temperature Process With Neural Networks." *Neurocomputing*, 2007.
13. Zapranis, A., and A. P. Refenes. *Principles of Neural Model Indentification, Selection and Adequacy: With Applications to Financial Econometrics.* Springer-Verlag, 1999.
14. Zhang, Q. "Using Wavelet Network in Nonparametric Estimation." *IEEE Trans. Neural Networks*, 1997: 227-236.
15. Zhang, Q., and A. Benveniste. "Wavelet Networks." *IEEE Trans. Neural Networks*, 1992: 889-898.