A Hybrid Forecasting Approach: Structural Decomposition, Generalized Regression Neural Networks and Theta Method

Marina Theodosiou
Imperial College
South Kensington, London
Email: m.theodosiou06@imperial.ac.uk

Murali Swamy
Imperial College
South Kensington, London
Email: murali.swamy@imperial.ac.uk

***Abstract*— A new hybrid forecasting methodology is proposed which leverages statistical and Artificial Intelligence (AI) techniques to perform multi-step ahead forecasting. This methodology is based on decomposing the time series into its structural components, predicting of each component individually and then reassembling the extrapolations to obtain forecast for a time series. The STL decomposition procedure is implemented to obtain the seasonal, trend and irregular components of a time series whilst Generalized Regression Neural Networks (GRNN) is used to perform out-of-sample extrapolations based on dynamic calibration of the training process for each component individually. The univariate Theta model is used to reinforce the directional stability and reliability of predictions. The proposed methodology is applied on 111 time series from the NN3 competition to obtain 18 out-of-sample predictions.**

I. Introduction

Since the pioneering work of Persons (1919) on the decomposition of time series, a large number of decomposition procedures were developed that perform additive or multiplicative disaggregation of the data into salient components such as trend, seasonality and error (Dagum (1988), Cleveland *et al.* (1981a, 1981b) and Cleveland *et al.* (1990)). These were primarily employed to facilitate time series analysis and understanding of the business cycle (Burns & Mitchell (1946)). However, after the imperative work of Box Jenkins (1970), on modelling seasonally adjusted data, decomposition techniques were also seen as a useful tool in forecasting. Some of the recent literature in this field includes the works of Landram, Abdullat & Shah (2004), Hansen & Nelson (1997) and Hansen & Nelson (2003). The latter implemented the Census X-11 (Shiskin *et al.*(1967)) procedure for additive decomposition of the time series into trend-cycle, seasonality and error, and used a time-delay neural network to obtain forecasts for the seasonality and trend components which were then combined into a single forecast through a backpropagation algorithm.

In this paper we propose a hybrid forecasting methodology that leverages three techniques: decomposition, artificial neural network and statistical model, to perform multi-step ahead prediction of reasonable accuracy. The data used in the analysis of the proposed methodology are the 111 time series from the NN3 neural network forecasting competition. Each time series is decomposed using STL decomposition procedure to obtain seasonal, trend and residual components. These individual components are predicted multi-step ahead using generalized regression neural network (GRNN). Theta forecasting model is used to reinforce long term behaviour for the component predictions. Finally, the predicted components are combined to produce the final multi-step ahead forecast.

The paper unfolds as follows. In section two a brief overview of Artificial Neural Networks (ANNs) and GRNNs is given followed by a description of the methodology developed in section three. Descriptions of results is given in section four, and conclusion is made in section five.

II. An Overview of Artificial Neural Networks

Neural Networks (hereafter NN) were developed through the realization that the human brain, being a complex, nonlinear and parallel computer, functions in a radically diverse manner to a computer processor. Since the early 1950s, scientists have studied extensively the structure and capabilities of the human brain in processing information, and have tried to encompass as many of those characteristics into a NN. This novel approach to information processing has gained a lot of popularity amongst various scientific disciplines such as engineering, computer science, biochemistry and physics.

With their successful implementation in other sectors, NNs are now emerging in finance, economics and the business industry. Their characteristic properties make them an attractive modelling methodology for addressing financial and economic problems. These include the *nonlinearity* of their structure, their built-in capability to adapt to new information, as well as the *universality* of their design, being the same in all the domains that involves their application (Haykin (1994)). In addition, no explicit assumptions need to be made on the functional relationship between the desired output and the independent variables of the model. Therefore NNs provide important advantages over the current statistical techniques used in modelling and predicting financial and economic data.

Despite this growing interest in the use of NNs as forecasting tools, statisticians and econometricians have been reluctant in trusting their predictive capabilities over well-established statistical techniques. This is largely the result of mixed conclusions from various studies comparing the performance of NNs against traditional statistical models (Timo Teräsvirtaa, Dick van Dijk, Marcelo C. Medeiros

(2005), Heravi, S., Osborn, D. R., & Birchenhall, C. R. (2004)). Zhang, Patuwo et al. (1998) provides a synthesis of published research in the area and draws upon some general conclusions on the limitations and advantages of NNs over traditional model-based methods.

Many experts in the field expressed the view that the realization of the potential of NNs involves more art than science (Zhang, Patuwo et al. (1998), Chatfield (1993)). This statement stems from the fact that the application of NNs involves a large number of degrees of freedom. Consequently, a great deal of experimentation and investigation, through trial-and-error, is required in order to establish parameter settings that provide the best architectural structure of the network. These degrees of freedom include, but are not exclusive to, the type of *activation function*, initial *synaptic weights*, connectivity of the *neurons*, *learning process* of the network and the number of *inputs* and *hidden neurons*. These decisions significantly affect the accuracy and performance of the network as well as its generalization capabilities. Therefore the application of NNs in forecasting involves a great level of modelling complexity, making it a time consuming process. Despite a considerable number of articles written with the scope of developing a stepwise selection of the parameters involved, no conclusive decision has been made on the best way to address this problem (Michel Nelson, Tim Hill & Marcus O'Connor (1999), Anders, U. and Korn, O. (1999),Teräsvirta, T. and Lin, C.-F. J. (1993), Crone, F. Sven (2004), Zhang, G. P. & Qi, M. (2005)).

In the current research Generalized Regression Neural Networks are used for prediction. This class of neural networks was proposed by Specht (1991) and presents important advantages over standard neural network architectures such as the feed-forward back-propagation algorithm (hereafter FFBP). The need for extensive designing and experimentation with free parameters is significantly reduced through the use of GRNNs. Apart from the small number of free parameters involved in their designing, GRNNs are also characterized by fast learning and convergence to the optimal regression surface. Unlike the standard FFBP neural network which needs a large number of iterations, and hence large computational time to converge to the desired functional form, GRNN adopts a one-pass learning algorithm and therefore reduces this limitation significantly. In addition, GRNN applications do not face the frequently encountered local minima problem of the FFBP applications and do not generate forecasts that are physically implausible (Cigizoglou (2005)).

Like any neural network, GRNN is used to form any arbitrary mapping between input and output variables through the process of training. It can approximate any arbitrary function between input and output vectors, through kernel regression, by creating a relationship between the given input variables and the expectation of the output variables. The latter is defined by:

$$E[y] = \frac{\int_{-\infty}^{+\infty} y f(X,y)dy}{\int_{-\infty}^{+\infty} f(X,y)dy} \quad (1)$$

where y is the output variable estimated by GRNN, X is the input variable to GRNN and $f(X, y)$ is the joint probability density function of X and y learned by GRNN through the training process. During the training process, GRNN forms the joint density function, $f(X, y)$, by observing each input and output pair (X, y). If a new input is given, GRNN estimates the most likely output value $E[y|X]$.

Particular to the GRNN is the use of the smoothing factor (or spread), σ, which alters the degree of generalization of the network. High smoothing factors increase the network ability to generalize, they may also degrade the error of prediction. Smoothing factors approaching 1 will straighten the path of the prediction line. Conversely, low smoothing factors degrade the network ability to generalize and may even prevent it from doing any prediction at all. Smoothing factors approaching 0 essentially create a dot-to-dot map.

Despite the attractive features of GRNNs, these have not been extensively applied to forecasting problems. Whilst there are numerous applications of the FFBP neural networks to predict financial and economic time series ((Trippi & Turban (1993), Azoff (1994), Refenes (1995), Gately (1996)), these are very limited in the case of GRNN. In the current research GRNNs are used to forecast the trend and residual components from each time series.

III. Research Methodology

We propose a forecasting methodology which is based on the estimated inherent structural properties of the observed time series. The structural evolution of a time series is studied through its decomposed seasonal, trend and residual components. Each of these components is modelled individually and appropriate forecasting methods are employed to perform multi-step ahead prediction on each of the obtained sub-series. The forecasting methods applied to each component must be capable of predicting multi-step ahead, i.e. these methods must model short-term properties as well as long-term trend of a time series.

The main steps of the proposed methodology are:

1) Decompose a time series using STL decomposition technique into trend, seasonal and residual components (sub-series).
2) Using seasonality component, obtain the value for the forecast horizon looking back at the same point in the seasonal cycle.
3) Apply GRNN to the trend component and obtain estimate for the forecast horizon. Use Theta method to obtain another estimate for the same forecast horizon. The average of these two estimates is the prediction for the trend component.
4) Apply GRNN to the residual component to obtain estimate for the forecast horizon.
5) The estimates from the seasonal, trend and residual components are linearly combined to obtain prediction for the forecast horizon.

The above steps are repeated for each various forecast horizons. For the NN3 competition the predictions are made from 1-step ahead to 18-steps ahead.

A. The Decomposition Procedure

For the decomposition of the time series into its constituent components, the well-established STL decomposition procedure (Clevelant *et al.* (1990)) is employed. It is a filtering procedure which decomposes the data through a sequence of applications of time series smoothing operations using locally-weighted regression (LOESS). The parameters for the STL procedure are obtained from the eigenvalue and frequency response analysis of a given time series. Several elegant design features of STL make it an attractive choice for time series decomposition. Among these is its ability to handle any amount of variation in the trend and seasonal components, and most importantly the robustness of the returned trend and seasonal components which are not susceptible to distortion by transient and aberrant behaviour in the time series. In addition, it is computationally tractable and can be easily implemented in statistical software packages[1].

Hence, for every time series x, STL returns, m, s and e such that:

$$x = m + s + e \tag{2}$$

where m is the trend component in the time series, s is the seasonality component and e is the residual component.

B. Extrapolating the Seasonality Component

The STL decomposition procedure, unlike other procedures such as SABL, imparts almost perfect seasonality which remains static across a time series with the range of values for each period varying depending on the level of seasonality inherent in the observed series. One can therefore obtain estimates of future seasonality by looking back at the same point in the seasonal cycle.

The data considered in the current research is specified as time series of monthly interval. Given this information, it was assumed that all time series contain annual seasonality and is incorporated in the STL decomposition procedure. Equation (3) was therefore used to obtain out-of-sample forecasts of the seasonality component for multi-steps ahead, assuming that the seasonality component will not change in the near future. The following relation therefore applies:

$$\hat{s}_i = s_{T-12+i} \tag{3}$$

where $\hat{s}_i$ denotes the forecast of the seasonality component at time point i and T denotes the total length of the time series.

C. Calibration of GRNN

The training (and testing) process of GRNN is designed to achieve high accuracy levels across the two sub-series, trend and residual. For training (calibration) the GRNNs, a training set representing the neural network inputs and outputs (targets) are constructed as described. The decomposed sub-series (trend & residual) is broken down into sequential non-overlapping windows of length equal to the forecast horizon starting from the last observed value in the sub-series. For each of this forecast window, a corresponding input (look-back) window is framed just prior to the forecast window, and its length is determined dynamically. The time series values in the input windows and the corresponding forecast windows, excluding the last window set, constitute the input and output pairs for the training of the GRNN. The last input window and forecast window is used for testing the predictability of the network.

These training and testing pairs are then scaled by the mean of the observed time series. This step is necessary in neural network calibration to ensure stable performance of the network in estimating the nonlinear relationship between these pairs and to obtain consistent and reliable predictability of forecasts. The same scaling process is also applied for out-of sample prediction.

Three important factors that can influence considerably the performance of the GRNN are the choice of the input window size, forecasting horizon and smoothing parameter. In this study, the size of the forecasting window was chosen to be the same as the specified forecast horizon (one for single or 18 for multi-step ahead forecasts), however the size of the input window was obtained dynamically.

For a given time series and for a particular forecast horizon, the GRNN was trained and tested using a number of different input window sizes starting with the smallest size (one) and increasing to the largest possible size accounting for the fact that the last forecast window is reserved for testing. The optimal input window size for each time series was then chosen based on the corresponding minimum mean absolute error (MAE) of this iterative procedure. The optimal input window is tuned based only on the last forecasting window. This is based on our belief, which is in line with general consensus, that future values of a time series depend more on the recent features of the data, than on features that lag far behind in time. So if the network is able to predict the last few observations in the time series with reasonably good accuracy, it is likely to give the same level of prediction accuracy for the out-of sample forecast window. The chosen optimal input window is then used for out-of-sample prediction.

The choice of the smoothing parameter is another important issue in the implementation of GRNNs. In the current research, this parameter is selected in a way similar to that of the procedure applied for selecting the optimal input window size. After the selection of the input window size, the GRNN is trained and tested with a range of spreads between 0.2 and 2. This range was decided based on our experimentation to ensure proper predictive properties across all time series considered in this study. As mentioned in the previous section, a very small smoothing parameter can prevent the network from generalizing or even predicting, whilst a large smoothing parameter can degrade the results considerably. Hence, the optimal spread for each time series and forecast horizon was the spread corresponding to the minimum testing error of the iterative testing process and was tuned based only on the last forecast window. This was then used for out-of sample prediction.

[1]The statistical software used in the current research is the **R Language** and is free to download from: `www.r-project.org`.

The process of training and selecting the optimal training parameters (input window & spread) is done independently for the trend and residual components of the time series. This can be justified based on the fact that these two components differ fundamentally in their structure. By doing so, forecasts of better accuracy can be achieved by modelling the component nonlinearity to the best possible extent.

D. Short Length Time Series

In order to predict 18 steps into the future with a reasonable accuracy, an appropriate sized window must be considered for GRNN training, thereby introducing inherent restriction on the minimum length of the time series for the implementation of the iterative procedure described above, for the selection of the training parameters. Consequently, for short time series (of length 51 or less), the proposed forecasting approach was modified. For such series the input window size and smoothing parameter are give reasonable values obtained through our experimentation.

E. Reinforcement of Trend with Theta Method

Since NN3 competition requires predicting forecasts of longer forecast horizon (18 steps ahead), the prediction methodology should posses the ability of modelling the long-term behaviour of a time series. From our analysis we found that for the time series in this study the GRNN is good at predicting short-term nonlinear structures. However, it does not fair well in estimating long-term trends of a time series. A similar observation was also made by Zhang & Qi (2005), whereby they noted "the inability of the feedforward neural network model to model a trend seems to be at odds with its universal approximation theory". So in our methodology we reinforced our estimate of GRNN with that of Theta statistical method (Assimakopoulos & Nikolopoulos (2000)) for the trend component to reinforce the directional reliability of our results.

The Theta method is univariate forecasting model based on the modification of the local curvatures of a time series through the *Theta* coefficient. This modification retains the mean and slope of the original data but not their curvature. The value of the Theta parameter is attributed to the long-term behaviour of the time series or the augmentation of the short-term behaviour. The smaller (larger) the value of the Theta coefficient, the larger the degree of deflation (inflation). In the extreme case where $\theta = 0$, the time series is transformed into a linear regression line. This model has been selected for best performance in the M3 forecasting competition (Makridakis & Hibon (2000)) and is defined as follows:

Let x_t be a time series with n observations. The equation of the a **Theta-Line** is defined as:

$$y_t = \alpha + \beta(t-1) + \theta x_t \tag{4}$$

where

$$\alpha = (1-\theta)\bar{x} - \beta(n-1)/2 \tag{5}$$

$$\beta = c_1 \sum_{i=1}^{n} i x_i - c_2 \sum_{i=1}^{n} x_i \tag{6}$$

$$c_1 = \frac{12}{n(n+1)(n-1)} \tag{7}$$

$$c_2 = \frac{6}{n(n-1)} \tag{8}$$

Theta method is, as proved in previous forecasting competition, has better modelling capability of long run behaviour of a time series. The GRNN estimate of the trend component is averaged with that of the estimate provided by Theta method to obtain the prediction for different forecast horizons. This adjusts the point forecast to be inline with the long term trend of the time series and hence produce better accuracy.

F. Offsetting of GRNN Prediction

It has been observed from experimentation on the in-sample data that the predictions given by the GRNN for the trend component lie at a vertical offset to the actual values, and thus magnifying significantly the error of the estimation. This seems to be the case even for time series with a relatively regular structure and small variability within the directional component.

In this study, in order to alleviate this problem, the predicted series was shifted by an offset distance, i.e. the vertical distance between the last observation of the actual time series and the first observation of the predicted series. By negating this offset, the predicted series is brought back to lie within the same range as the actual values thereby reducing the error significantly. We expect this behaviour to continue even on to the out-of sample predictions and therefore apply the offset adjustment for out-of sample extrapolation.

One can argue that if the last value of the time series happens to lie on an extreme value compared to the mean of the time series then this offsetting will increase the prediction error instead of decreasing. Since this offset is applied only to the trend component where we average the estimate of GRNN (with the offset negated) with the estimate of Theta method we think that the effect of these extreme value will reduce.

G. Prediction of Residual Component

To our knowledge, none of the existing statistical and Artificial Intelligence Methodologies deals with the extrapolation of the residual component obtained through a decomposition procedure. The irregular component is the residual variability left in the data after the removal of the main components (usually trend-cyclical and seasonal component), of a time series. Although academics agree that important information can still be found in the irregular component, and therefore its exclusion from the prediction process can result in negative ramifications for forecast accuracy, nevertheless, existing statistical methodologies are unable to predict and model the erratic behaviour of its sub-series and therefore in all applications the irregular component is excluded or is assumed to be a white noise process with zero mean (Newbold (1991)). In the current paper, an attempt is made to include the irregular component into the prediction process rather than discard it completely.

In the case of statistical techniques however, the exclusion of the error component is offset by enhanced prediction results obtained from de-noised data. It is well known that the elimination of noise from the data can improve significantly the forecasting performance of statistical techniques and thus is usually advised for the data to be pre-processed and filtered before extrapolation.

This is not the case however with Artificial Intelligence techniques. Even though, it is supported in the literature that pre-processing of the data such as deseasonlising and detrending, can improve predictions significantly when implementing neural network techniques (Nelson, Hill, Remus & O'Connor (1999)), as shown by Zhang & Qi (2005), predicting a series with high level of noise such as the residual series after the elimination of the trend and seasonal component, is possible with artificial intelligence techniques and can reduce the estimation error significantly. The reason behind this is the fact that neural networks, and particularly GRNN, cannot overshoot but are always likely to return a prediction within a reasonable range from the real observation. As noted by Cigizoglou (2005), GRNN "do not generate forecasts which are physically implausible". For these reasons, it was decided that much benefit can be obtained from the inclusion of the residual component in the forecasting process rather than its exclusion.

IV. Results & Discussion

Our proposed hybrid methodology is applied to the complete set of 111 times series of NN3 competition. All the time series examined were of monthly frequency with above zero positive values. The structural characteristics varied greatly across these 111 time series. The shorter time series (from NN1-NN50) with 51 or 50 observations (in the case of NN22 & NN31) were mostly dominated by noise with insignificant seasonal and trend components. Most of the longer time series were dominated by a defined seasonal structure, and some time series (NN59, NN102, NN103) are perfect seasonal with almost no noise. There were other time series with both trend and seasonal behaviour, and in few cases some outliers were also be observed (e.g. NN108, NN110).

The time series for the analysis are not subjected to any data preprocessing and are decomposed using STL in the first stage. The scaling of the training data with the mean of the time series was helpful in limiting the bounds of the predictions. Based on this we say that our methodology is robust to outliers.

Preliminary testing of our methodology is performed by forecasting multi-step ahead the last 18 observations of each time series. The predictions obtained from this testing process seem to be inline with the observations. Following four figures display the actual series and corresponding predicted series for the last 18 observations. From the small mean absolute error of these predictions for all the time series we conclude that our methodology is capable of performing well in both for short-term and long-term. Finally, we predicted 18 steps ahead out-of-sample forecasts for the NN3 competition.

V. Conclusion

We propose a new hybrid forecasting methodology that makes use of structural decomposition technique along with neural network and statistical method to perform multi-step ahead forecasting to a reasonable extent. The time series is decomposed into seasonal, trend and residual components. These components are then predicted separately into the future. The forecasts thus obtained from these components are combined to obtain the final prediction. We used Generalized Regression Neural Networks for predicting the trend and residual components. The prediction from the trend is reinforced with the prediction from the Theta methods for long term directional stability. We tested our methodology on the 18 in-sample observations and found the mean absolute errors of acceptable range. For our experimentation we claim that this methodology is robust to outliers and capable of capturing both short-term and long-term behaviour of a time series.

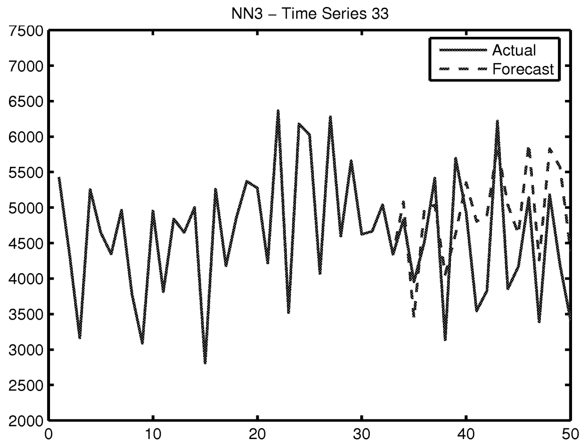

Fig. 1. NN3 Time Series 33

References

[1] Anders, U. and Korn, O. (1999). *Model Selection in Neural Networks.* Neural Networks, **12**, 309323.
[2] Assimakopoulos, V. & Nikolopoulos, K. (2000), *The Theta Model: A Decomposition Approach to Forecasting*. International Journal of Forecasting, **16**, 521530.
[3] Azoff, E.M. (1994), *Neural Network Time Series Forecasting of Financial Markets*. John Wiley and Sons, Chichester.
[4] Burns, A.F & Mitchell, W.C. (1946), *Measuring Business Cycles*, National Bureau of Economic Research, New York.
[5] Chatfield, C. (1993). *Neural Networks - Forecasting Breakthrough or Passing Fad.* International Journal of Forecasting **9**(1): 1-3.

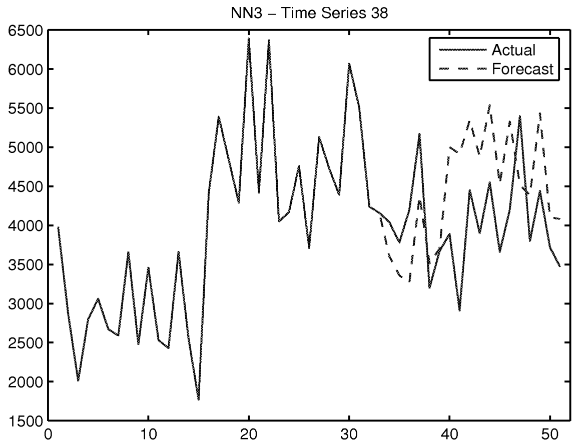

Fig. 2. NN3 Time Series 38

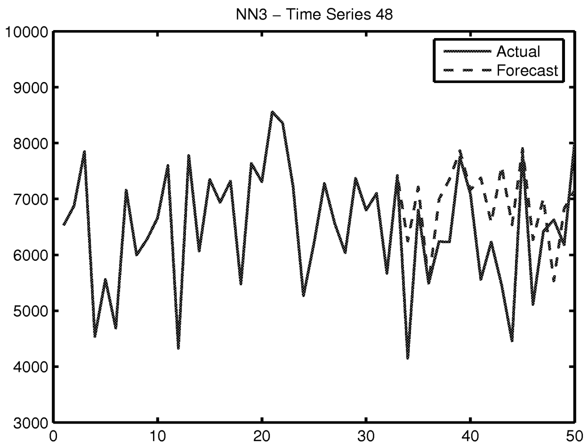

Fig. 3. NN3 Time Series 48

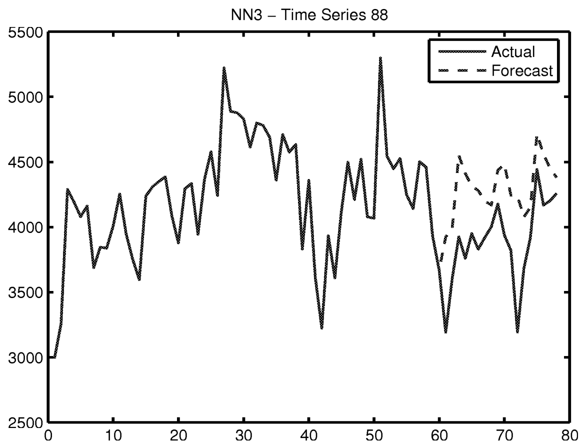

Fig. 4. NN3 Time Series 88

[6] Cleveland, W. S., Devlin, S. J. & Terpenning, I. J. (1981), *The SABL statistical and graphical methods*. Computing Information Library, Bell Labs, 600 Mountain Ave., Murray Hill, N.J. 07974, USA.

[7] Cleveland, W. S., Devlin, S. J. & Terpenning, I. J. (1981), *The details of the SABL transformation, decomposition and calendar methods.* Computing Information Library, Bell Labs, 600 Mountain Ave., Murray Hill, N.J. 07974, USA.

[8] Cleveland, R.B., Cleveland, W.S, McRae, J.E & Terpenning, I. (1990) *STL: A Seasonal-Trend Decomposition*

[9] Cigizoglou, H.K. (2005), Generalized Regression Neural Network in Monthly Flow Forecasting*, Civil Engineering and Environmental Systems,* **22**, 71-84. Procedure Based on Loess, Journal of Office Statistics, **6**, 3-73.

[10] Crone, F. Sven. *Stepwise Selection of Artificial Neural Networks Models for Time Series Prediction*. Journal of Intelligent Systems, **14**, 2-3.

[11] Dagum, E.D. (1988), *The X-11-ARIMA/88 Seasonal Adjustment Method Foundations and Users Manual*, Canada.

[12] Gately, E. (1996), *Neural Networks for Financial Forecasting*. John Wiley, New York.

[13] Hansen, J.V & Nelson, R.D. (1997), *Neural Networks and Traditional Time Series Methods: A Synergistic Combination*, IEEE Transactions on Neural Networks, **8**, 863-873.

[14] Hansen, J.V & Nelson, R.D. (2003), *Forecasting and Recombining Time-Series Components by Using Neural Networks*, Journal of Operational Research Society, **54**, 307-317.

[15] Haykin, S.S. (1994) *Neural Networks: A Comprehensive Foundation*. Prentice Hall.

[16] Heravi, S., Osborn, D. R., & Birchenhall, C. R. (2004). *Linear Versus Neural Network Forecasts for European Industrial Production Series*. International Journal of Forecasting, **20**, 435 446.

[17] Landram, F., Abdullat, A. & Shah, V. (2004), *Using Seasonal and Cyclical Components in Least Squares Forecasting Models*, Southwestern Economic Review, **31**, 189-196.

[18] Makridakis, S. & Hibon, M. (2000), *The M3-Competition: Results, Conclusions and Implications*, International Journal of Forecasting **16**, 451 476.

[19] Nelson, M., Hill, T. Remus, W. & O'Connor, M. (1999). *Time Series Forecasting Using Neural Networks: Should the Data be Deseasonalized First?* Journal of Forecasting, **18**, 359-367.

[20] Newbold, P. (1991), *Review of Quantitative Finance and Accounting*, **1**, 259-279.

[21] Persons, W.M. (1919), *Indices of Business Conditions*, Review of Economics and Statistics, **1**, 5-107.

[22] Refenes, A.N. (1995), *Neural Networks in the Capital Markets*. John Wiley, Chichester.

[23] Specht, D. F. (1991), *A General Regression Neural Network*, IEEE Trans. Neural Networks, **2**(6), 568576.

[24] Shiskin, J., Young, A.H. & Musgrave, J.C. (1967), *The X-11 Variant of the Census II Seasonal Adjustment Program*, Technical Report #15, US Department of Commerce, Bureau of Economic Analysis, Washington DC.

[25] Teräsvirta, T. and Lin, C.-F. J. (1993). *Determining the number of hidden units in a single hidden-layer neural network model.* Technical report. Bank of Norwey.

[26] Timo Teräsvirtaa, Dick van Dijk, Marcelo C. Medeiros (2005). *Linear Models, Smooth Transition Autoregressions, and Neural Networks for Forecasting Macroeconomic Time Series: A Re-examination.* International Journal of Forecasting **21**, 755 774

[27] Trippi, R.R. & Turban, E. (1993), *Neural Networks in Finance and Investment: Using Artificial Intelligence to Improve Real-world Performance*. Probus, Chicago.

[28] Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). *Forecasting with Artificial Neural Networks: The State of the Art*. International Journal of Forecasting, **14**, 3562

[29] Zhang, G. P. & Qi, M. (2005). *Neural Network Forecasting for Seasonal and Trend Time Series.* European Journal Of Operational Research, **160**, 501-514.