Neural network-based prediction of the USD/GBP exchange rate: the utilisation of data compression techniques for input dimension reduction

Leonidas Anastasakis and Neil Mort

Dept. of Automatic Control & Systems Engineering
University of Sheffield
Mappin St., Sheffield S1 3JD, U.K.
Tel: ++44 114 222 5559
Fax: ++44 114 273 1729
Email: cop99la@sheffield.ac.uk & n.mort@sheffield.ac.uk

Abstract: Financial prediction is a research active area and neural networks have been proposed as one of the most promising methods for such prediction. In this paper we simulate an MLP network in order to perform one step ahead prediction of the USD/GBP exchange rate. Four different input vectors are tested and the best network architecture determined. In addition, an autoassociator MLP network has been applied to reduce input data dimension. It is shown that the generalisation performance of the network is improved when the reduced input vector is used.

Keywords: time series modelling, financial prediction, neural networks, autoassociator MLP network, data analysis, dimension reduction.

1. INTRODUCTION

Modelling real world systems is a difficult task as these systems are often characterised by high dimensionality, noise corruption and existence of significant non-linearities. Financial systems in general, belong to this category and numerous researchers have investigated the possibility of developing useful models. The model of a financial system like the stock market or the exchange rate can be used by the trader for prediction tasks in order to increase the profitability of the trades. The interactions between different countries economical systems and the consequences of the globalisation phenomenon in economies increase the dimensionality of exchange rate systems making the accurate prediction of an exchange rate difficult. Economists believe that the task of predicting exchange rates is not feasible. They claim that the evolution of any exchange rate follows the theory of efficient market hypothesis, which in its weak form asserts that the price of an asset reflects all of the information that can be obtained from past prices of the asset (Giles et al., 1997).

Agreement with this statement means that the best prediction of the exchange rate follows the random walk model and no profit can be taken by a trader. Giles et al. (1997) rejects this claim while Schwaerzel and Rosen (1997) point out that the random walk hypothesis is not always valid. Both researchers use neural networks to increase the accuracy of prediction with regard to other statistical models. Episcopos and Davis (1995) for example, tested the predicted ability of a random walk model, an EGARCH-M model and neural networks, and concluded that neural networks were superior. Economists using purely economic theory have developed financial models to predict exchange rates. Parikh (1991) describes the main theoretical formulations for modelling exchange rates known as (i) the purchasing power parity approach, (ii) the asset market approach and the (iii) portfolio approach. Alternatively, based on historical exchange rate data statisticians as well as economists have developed linear autoregressive models. The latter approach has two main drawbacks. First, it is not strictly rational to explicitly use past values of the exchange rate for the modelling task due to the globalisation phenomenon in economies. Secondly, the majority of financial systems are non-linear, so the application of linear models cannot accurately capture the relationships within data.

The introduction of techniques that can both reflect the non-linear behaviour of the data as well as using a variety of variables in the input vector can be proposed as a solution to these problems. Artificial neural networks is an approach that many researchers have adopted in order to predict either the exchange rate of two currencies or the fluctuations in a stock market index (Embrechts (1995) and Chen (1994)). Their non-linear structure enables the construction of models that efficiently describe real world systems. Several types of neural networks and associated learning algorithms exist; however, the multilayer perceptron (MLP) and backpropagation (BP) learning rule are the most popular choices in financial applications. In this paper a MLP network was applied to the prediction of the exchange rate between the US Dollar and British Pound (noted as USD/GBP throughout the paper). Overfitting problems and local minima in the error surface were avoided using a modified BP algorithm that included a momentum term and an adaptive learning rate. Four different input vectors were applied to MLP networks with various numbers of hidden units and the best network topology (i.e. number of hidden units) was identified for each input vector. The possibility of reducing the network's input dimension was examined and the performances of these input vectors were compared to the original unmodified cases.

2. INPUT VECTOR

The data consists of the daily prices of the USD/GBP exchange rate for the period 01/01/91 – 31/12/93 creating a series of 784 observations, which were collected from the web financial data provider http://www.accu-rate.ca that contains historical data for many exchange rate pairs. The data were divided into the training set used by the network to learn the underlying relationship between input-output and the testing set of the latest observations that helped the evaluation of the network's prediction performance.

The results of previous work and a basic knowledge in financial systems were considered for the construction of the input vector. The first input vector consisted of past values of the USD/GBP series itself. A period up to four days lag including the same day's price of USD/GBP was used to predict next day's price. The extent of past values that were included in the input vector was selected subjectively based on the behaviour of the stock market system. For example, a small crisis in a stock market that occurs on a Monday can affect the fluctuations of the index throughout the week. A similar type of behaviour is assumed for the case of the exchange rate market. The subjective selection was justified by the results of the correlation analysis, as will be described later in the paper.

According to Azoff (1994) it is worth considering the inclusion of technical indicators and the fundamental series in the input vector. Indicators are a popular approach in analysing future prices in the technical analysis methodology. Azoff (1994) describes one of the most popular indicators, the Random Walk Index (RWI), which is given by the form:

$RWI = \frac{\Delta p_t(N\tau)}{\overline{\Delta p_t}\sqrt{N}}$ (1), where $\Delta p_t(N\tau)$ is the price difference over the time interval $N\tau$, τ is the sample interval (i.e. τ=1 in our case), $\overline{\Delta p_t}$ is the average first differences of the series. A set of three RWIs computed for one, two and three look back periods as well as the four past values of USD/GBP were included in the second input vector.

The reason for choosing fundamentals series was to capture the influence of the most important factors in the system. The choice of these factors depends on the globalisation of the financial systems and the interactions between them. In this case the exchange rates between US Dollar and Japanese Yen, and between British Pound and German Mark (noted as USD/JPY and GBP/DEM respectively throughout the rest of the paper) were considered. Germany, which is one of the leading and most industrial powerful European countries, and Japan, as the major economical – industrial rival against USA in the world, have a significant effect on the global exchange rate market. Specifically the differences between the same day's and last day's price of USD/JPY and GBP/DEM as well as the four past values of USD/GBP were included in the third input vector. Finally, in the fourth input vector the complete set of variables was included in order to predict the next day's price for the USD/GBP exchange rate.

3. PRE-PROCESSING STAGE

The pre-processing stage, which is a necessary step prior to network training, can be divided into two main parts: First, data analysis is employed for the examination of the data quality. Then specific actions are taken in order to ensure a suitable input vector to train the network. A neural network's superiority over economical and autoregressive models lies in its ability to eliminate any subjectiveness in the input vector. A wide variety of input variables can be considered but it must be remembered that network's generalisation performance is sensitive to them. The inclusion of variables in the input vector that have no underlying relationship with the output will cause problems in the training procedure. Data analysis is therefore an important task in modelling and prediction since it explores the data and provides useful information on its structure, etc.

The first step in data analysis is the graphical plotting of the series. This step helps in the identification of potential outliers and provides a visual inspection of the series stationarity. Any potential outliers should be removed in order to avoid problems with network training and generalisation performance.

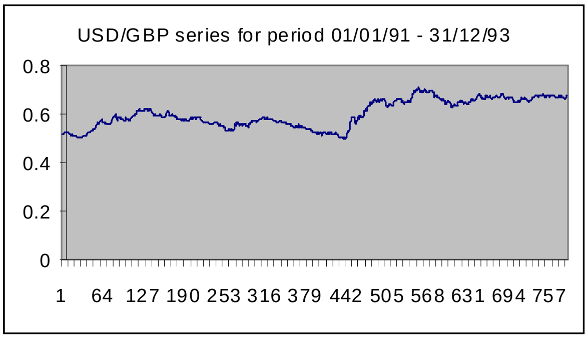

Figure 1a

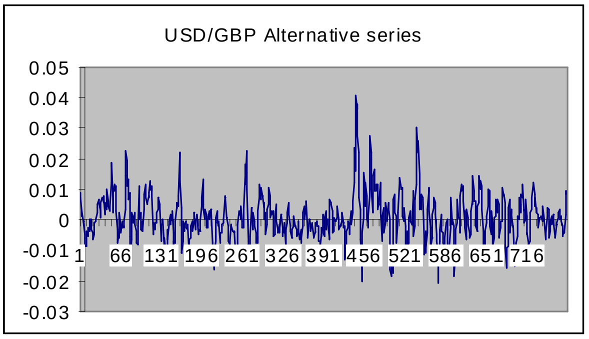

Figure 1b

It is important to explicitly model stationary series, otherwise the prediction may be biased. Figure 1a illustrates that while no significant outliers exist, the USD/GBP series is clearly not stationary. The introduction of either the first difference series or the log difference series is a popular method in financial modelling for overcoming nonstationarity. Robinson (1998) suggests that for regression tasks an alternative difference series given by the equation (2) could be applied.

$r_t = p_t - \frac{1}{5}\sum_{i=1}^{5} p_{t-i}$ (2), where p_t is the raw value for USD/GBP exchange rate.

This modification has a smoothing effect on the primary series due to the averaging term (figure 1b). The stationarity of the series was further examined by applying the Dickey-Fuller test (Said et al., 1984). According to these results there was no indication of unit roots within the data. Additionally, the Dickey-Fuller value was significantly below its critical value confirming nonstationarity of the data.

Other more advanced methods like correlation analysis are also included in data analysis. Using the basic equations for the calculations of autocorrelation and cross-correlation coefficient at lag k, only linear relationships within the data can be extracted. However, most real world data exhibits non-linear behaviour, so the basic equations are not useful in such cases. Billings and Zhu (1994) suggest a solution to this problem by introducing non-linear forms for both autocorrelation and cross-correlation coefficients. In equation (3) both linear and non-linear equations are presented:

$$\Phi_{xx}(k) = \frac{\frac{1}{N}\sum_{t=1}^{N-k}(x_t - \overline{\mu_\chi})(x_{\tau+\kappa} - \overline{\mu_x})}{\sigma^2} \qquad \Phi_{x^2x^2}(k) = \frac{\sum_{t=1}^{N-k}(x_t^2 - \overline{\mu_{x^2}})(x_{t+k}^2 - \overline{\mu_{x^2}})}{\sum_{t=1}^{N-k}(x_t^2 - \overline{\mu_{x^2}})^2}$$

$$\Phi_{xy}(k) = \frac{\sum_{t=1}^{N-k}(x_t - \overline{\mu_x})(y_{t+k} - \overline{\mu_y})}{\sqrt{\left[\sum_{t=1}^{N-k}(x_t - \overline{\mu_x})^2\right]\left[\sum_{t=1}^{N-k}(y_{t+k} - \overline{\mu_y})^2\right]}} \qquad \Phi_{\chi^2y^2}(k) = \frac{\sum_{t=1}^{N-k}(x_t^2 - \overline{\mu_{x^2}})(y_{t=k}^2 - \overline{\mu_{y^2}})}{\sqrt{\left[\sum_{t=1}^{N-k}(x_t^2 - \overline{\mu_{x^2}})^2\right]\left[\sum_{t=1}^{N-k}(y_{t+k}^2 - \overline{\mu_{y^2}})^2\right]}} \quad (3)$$

where, k= 0,1,2, …K, $\overline{\mu}$ is the mean of time series, s^2 its variance and $\overline{\mu_{x^2}} = \frac{1}{N}\sum_{i=1}^{N} x_{t_i}^2$.

Both linear and non-linear autocorrelation and cross-correlation coefficients between USD/GBP and USD/JPY – GBP/DEM series respectively were calculated over a period of zero to ten lags and the results are presented in figure 2. In the case of the autocorrelation function for USD/GBP series the existence of significant correlation at lags 1, 2, 3, 4, justifies the selection of four past values of USD/GBP data in the input vector. There are also significant non-linear autocorrelations at lags 5 and 6 but it was found that the inclusion of these days in the input did not significantly improve the performance of the network.

Good generalisation requires the training of the network with input variables that are relevant to the output. In this case the existence of significant cross-correlation between the USD/GBP series and the other two fundamental series indicates the relevance between input and output vectors.

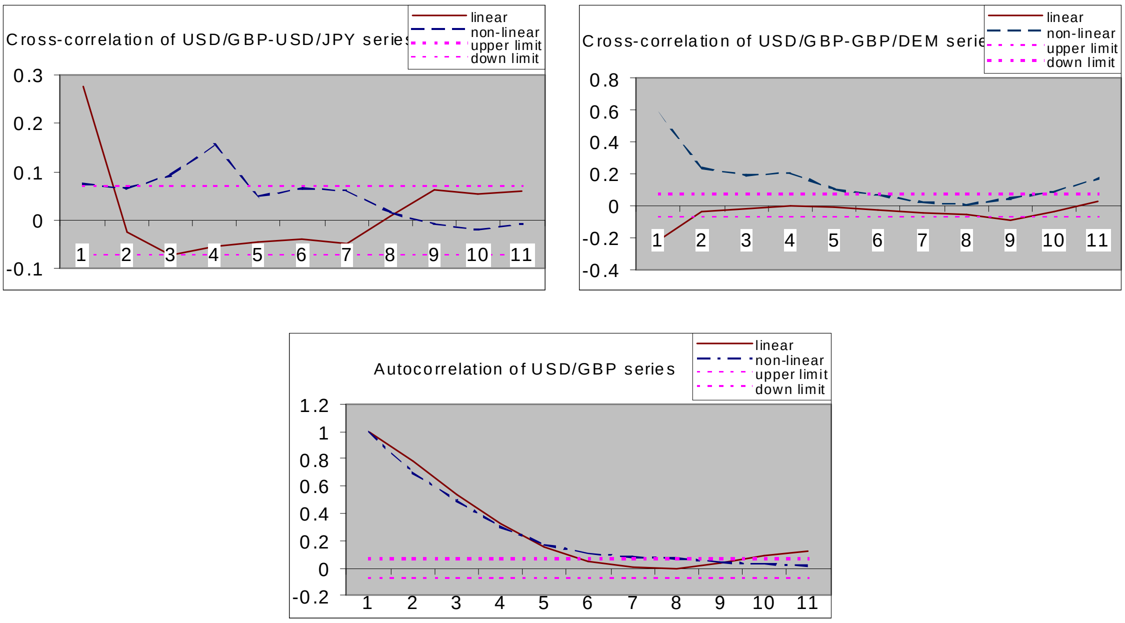

Figure 2

Normalisation of the input vector is an essential task during the pre-processing stage. 'The normalisation is important especially when different types of variables are recorded since their values may differ by either an order of magnitude or their relative size may not reflect their importance to the output variable. Normalising the data ensures that all have the same order of magnitude' (Robinson, 1998). Azoff (1994) describes several types of normalisation procedure such as along, across, mixed and external normalisation. In this paper we performed along channel normalisation of the whole input vector as well as the output vector in the range [-1 1], given by the form:

$$p_i^j = \frac{2*(p_i^j - \min_j)}{\max_j - \min_j} - 1$$ (4), where j is the column and i the row of the vector.

4. INPUT VECTOR REDUCTION BY USING AUTOASSOCIATOR MLP NETWORK

The speed of convergence of neural networks is effected by either the number of hidden units or the input vector's dimension. The reduction of the input dimension not only increases convergence speed, but also eliminates the possibility of duplicated information within the variables of input vector that would effect network performance. For these reasons it is essential to reduce the dimension of the input vector. Azoff (1994) describes three techniques for dimension reduction, the data compression method also known as the autoassociator MLP network, the principal component analysis (PCA) and the Wavelet transform.

Steurer and Siegler (1998), while attempting to forecast the German stock index DAX with neural networks, applied PCA and concluded that the achieved reduction in the number of inputs was disappointing in relation to the decrease in network's performance. In this paper we attempt to achieve dimension reduction by applying the data compression technique known as the autoassociator MLP network. The autoassociator MLP network is a three-layer network in which the number of input nodes is equal to the number of output nodes.

The output vector is identical to the input vector so that the MLP network is trained to reproduce the input vector. Using a number of hidden nodes less than the number of input units we seek to compress the input vector without losing any information within it. It is obvious that there will be a small error in the output since it is virtually impossible to reproduce the input vector exactly. When the error reaches an acceptable level the values of the hidden layer weights are frozen and the hidden layer output is applied as input to the normal neural network, thus achieving a reduction of the input vector. The inability of the MLP to reduce the error below an acceptable level implies that dimension reduction is not feasible.

5. TRAINING THE NETWORK

In this paper the MATLAB Neural Network Toolbox (Ver 2.0) was employed for the simulation of a MLP neural network. Three different neural networks were trained with respect to the number of hidden units. In each case a three-layer MLP network was constructed with the number of input units equal to the elements of the input vector. The network used a single hidden layer with hyperbolic tangent activation function and 10, 6, 4 hidden units respectively and a single output neuron for one step ahead forecasting of the USD/GBP exchange rate. The piece-wise linear function was applied as the activation function in the output layer.

The four input vectors were introduced to each network topology (i.e. number of hidden units) and their performance was evaluated by the root mean squared error:

$RMSE = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (\overline{y_i} - y_i)^2}$, where N_t is the number of testing data and $\overline{y_i}$ the predicted value.

The training procedure terminated if either the sum-squared error in the output was below an acceptable level or the number of epochs (i.e. number of iterations of the whole training set) was equal to 10000. Preliminary experiments revealed that every network that was trained for more than 10000 epochs did not significantly improve its generalisation performance. In some cases the performance even decreased, suggesting the appearance of overtraining problems. In order to tackle the overtraining problem as well as eliminate the possibility of being trapped in local minima, a set of ten runs were performed using each network topology with different initialisation of the weights. The mean value of the RMSE was calculated and the topology with the smallest mean was classified as the best network topology of each particular input vector. 'Averaging of neural networks forecasts is based on the idea of combination of predictions', (Steurer et al., 1998). By taking the average of the ten networks we stabilise the results of the network as well as avoiding the issue of becoming trapped in a local minimum. So far little research has been published about combining models but Steurer and Siegler (1998) presents a number of references relevant to this particular subject.

Reduction of input dimension was also attempted by using the autoassociator MLP network for each one of the four input vectors. For every input vector a number of autoassociator MLP networks with different numbers of hidden units were trained to reconstruct the output. On each occasion the best network topology was determined according to the value of the RMSE error over the whole set of data. Each network was trained for 25000 epochs and a value of 0.1 for the mean of RMSE of the output neurons was considered as the threshold value for the selection of the best topology. In table 1 the number of hidden units and the corresponding mean value of RMSEs error that was selected to reduce each input vectors are presented.

Network's input	Number of hidden units	Mean of RMSEs
1^{st} input vector	2	0.069
2^{nd} input vector	3	0.049
3^{rd} input vector	3	0.067
4^{th} input vector	4	0.056

Table 1

The hidden layer output was used to train an MLP network with six hidden neurons in order to compare the network's performance of the original input vector with that of the reduced version. Similar to the procedure that was followed for the original input vector, ten different runs were performed and the mean of their RMSE was computed.

6. DISCUSSION

For all four input vectors that were tested the best network topology includes six hidden units. Table 2 suggests that every time new elements were added in the first input vector, better performance was achieved.

These results justify the suitability of the proposed input series as well as their relevance to the network's output. The best network according to the mean of RMSE is obtained for the fourth input vector where all the proposed input variables are included. Its performance over the test data can be seen in figure 3.

Network's input (i.e. hidden units: 6)	RMSE
1^{st} input vector (4 past USD/GBP)	0.1585544
2^{nd} input vector (4 past USD/GBP + RWIs)	0.154401
3^{rd} input vector (4 past USD/GBP + Rates)	0.1512797
4^{th} input vector (4 past USD/GBP + RWIs + Rates)	0.1468761

Table 2

Averaging the output of ten different networks creates smoothing of the outputs and at the same time defines an output that was never produced by any network. For these reasons we chose as the prediction of USD/GBP exchange rate the network's output with the closest value of RMSE to the mean of RMSEs. In that way we gain the advantages offered by the multiple runs offer and define an output that was given by a trained neural network.
The generalisation is improved by training a network with the same number of hidden units as before (i.e. six hidden neurons) but with the reduced input vector. Moreover the convergence speed of network is improved, since the number of free parameters of the network is reduced.

Network's input	RMSE train data (original)	RMSE train data (reduced)	Improvement (%)	RMSE test data (original)	RMSE test data (reduced)	Improvement (%)
1st input	0.1800965	0.1661311	7.754	0.1585544	0.1355729	14.494
2nd input	0.1738541	0.1522578	12.422	0.154401	0.1358631	12.006
3rd input	0.174732	0.1617476	7.433	0.1512797	0.1365206	9.756
4th input	0.1712605	0.1483922	13.353	0.1468761	0.1388306	5.477

Table 3

In table 3 the percentage of improvement over the training set and test set for the reduced input vector against the original input vector is presented. It is clear that in all cases the reduction of the input dimension produced better results. For example, in the case of the reduced first input vector improvement about 7.7% in training data and about 14.5% on testing set of data was achieved (figure 4). Based on those results we conclude that the reduction of the input dimension through the autoassociator MLP network was successful.

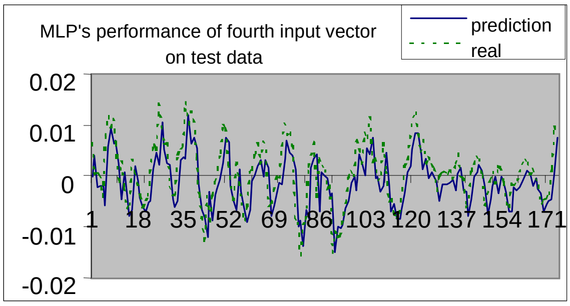

Figure 3

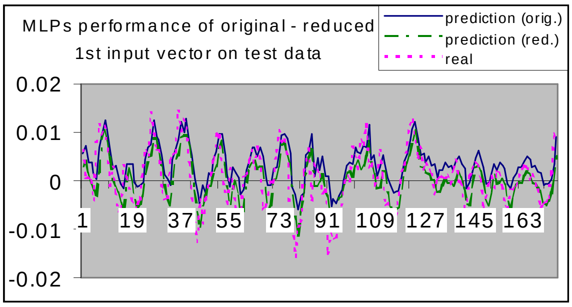

Figure 4

7. CONCLUSIONS

In this paper we tested four different input vectors in order to predict one step ahead the USD/GBP exchange rate. A number of different network topologies were trained to choose the best for each input. Moreover the autoassociator network was applied to reduce input vector with satisfactory results. Using a fixed number of epochs for the termination of the training procedure, even though considered acceptable (Chen, 1994 and Jasic et al., 1995) can lead to overtraining problems. Despite the precautions that were taken in our case, like the modified BP algorithm and the averaging of neural network forecasts for tackling these problems, other well-known methods should be used. Unfortunately the MATLAB Neural Network Toolbox 2.0 that was applied to simulate the MLP network does not offer the possibility of using stopping techniques like the early stopping and cross-validation methods. An attempt to compare our results with those from a network that uses such a stopping rule will be worth considering. Moreover, the adaptation of other modelling methods based on the principal of self-organisation like the Group Method of Data Handling (GMDH) which reduce the extent of human's involvement in the construction of input vector even more, would be useful (Scott et al., 1976).

8. ACKNOWLEDGEMENTS

Mr. Anastasakis would like to thank the Greek State Fund (I.K.Y.) for sponsoring his postgraduate studies in the University of Sheffield.

9. REFERENCES

Azoff, E.M. (1994) “Neural Network Time Series Forecasting of Financial Markets”, Chichester, John Wiley & Sons - (Wiley finance editions).

Billings, S.A. and Zhu, Q.M. (1994) “Nonlinear model validation using correlation tests”, *International Journal of Control*, 60, 6, 1107-1120.

Chen, C.H. (1994) “Neural Networks for Financial market prediction”, *IEEE International Conference on Neural Networks*, 7, 1199-1202.

Embrechts, M.J. (1995) “Forecasting Foreign Exchange Rates with Artificial Neural Networks”,

Episcopos, A. and Davis, J. (1995) “Predicting Returns on Canadian Exchange Rates with Artificial Neural Networks and EGARCH-M Models”, in Refenes, A-P.N., Abu-Mostafa, Y., Moody, J. and Weigend, A. (eds.) *Neural Networks Financial Engineering. Proceedings of the third International Conference on Neural Networks in the Capital Markets, Singapore*, London, World Scientific, 135-145.

Giles, C.L., Lawrence, S., Tsoi, A.C. (1997) “Rule Inference for Financial Prediction Using Recurrent Neural Networks”, *Proceedings of the IEEE/IAFE Computational Intelligence for Financial Engineering*, New York, IEEE, 253-259.

Jasic, T. and Poh, H.L. (1995) “Financial Time Series Prediction Using Neural Networks: A Case Study for the TOPIX Data”, *Proceedings of the Sixth Australian Conference on Neural Networks*, Sydney, 65-68.

Parikh, A. (1991) “Modelling the Behaviour of the Dollar/Pound Exchange Rate”, *Discussion Paper (University of East Anglia. Economics Research Centre)*, no. 9115, Norwich, School of Economics and Social Studies, University of East Anglia.

Robinson, C. (1998) “Multi-Objective Optimisation of polynomial Models for Time Series Prediction Using Genetic Algorithms and Neural Networks”, Dept. of Automatic Control & Systems Engineering, University of Sheffield, Sheffield.

Said, E.S. and Dickey, D.A. (1984) “Testing for unit roots in autoregressive-moving average models of unknown order”, *Biometrika*, 71, 3, 599-607.

Schwaerzel, R. and Rosen, B. (1997) “Improving the Accuracy of Financial Time Series Prediction Using Ensemble Networks and High Order Statistics”, *IEEE International Conference on Neural Networks. Proceedings*, 4, New York, IEEE, 2045-2050.

Scott, D.E. and Hutchinson, C.E. (1976) “The GMDH algorithm-a technique for economic modelling”, *Modeling and Simulation*, 7, 729-733.

Steurer, E. and Siegler, W. (1998) “Forecasting of the German Stock Index DAX with Neural Networks: Using Daily Data for Experiments with Input Variable Reduction and a Modified Error Function”, in Xu, L., Chan, L.W., King, I. and Fu, A. (eds.) *Intelligent Data Engineering and Learning . Perspectives on Financial Engineering and Data Mining. 1st International Symposium*, Hong Kong, Springer, 113-120.