Forecasting the Index of Financial Safety (IFS) of South Africa using neural networks

Matkovskyy, Roman

School of Economics,North-West University, South Africa

August 2012

Forecasting the Index of Financial Safety (IFS) of South Africa using neural networks

Roman Matkovskyy

PhD, Associate professor, post-doctoral research fellow (e-mail: roman.matkovskyy@gmail.com)

School of Economics

Private Bag X6001

North-West University

Potchefstroom Campus

Potchefstroom

South Africa

2520

2012

Forecasting the Index of Financial Safety (IFS) of South Africa using neural networks

Abstract

This paper investigates neural network tools, especially the nonlinear autoregressive model with exogenous input (NARX), to forecast the future conditions of the Index of Financial Safety (IFS) of South Africa. Based on the time series that was used to construct the IFS for South Africa (Matkovskyy, 2012), the NARX model was built to forecast the future values of this index and the results are benchmarked against that of Bayesian Vector-Autoregressive Models. The results show that the NARX model applied to *IFS* of South Africa and trained by the Levenberg-Marquardt algorithm may ensure a forecast of adequate quality with less computation expanses, compared to BVAR models with different priors.

JEL Classifications: G01, C38, C11, C32, C53, E50, G17

Keywords: Index of Financial Safety (IFS), neural networks, nonlinear dynamic network (NDN), nonlinear autoregressive model with exogenous input (NARX), forecast

1. INTRODUCTION

The recent financial crisis, which influenced a number of well-developed countries, again emphasised the role of financial system safety in maintaining macro-stability. The question of accuracy of predictions concerning future conditions of financial safety of a country may be of interest to investors and policy-makers alike.

For the purpose of forecasting financial perturbation, a variety of instruments based on time-series analysis may be applied, for example: Markov switching models (see, among others Abiad, 2003; Chen, 2005), state space methods and the mixed-measurement dynamic factor model (Koopman, Lucas & Schwaab, 2010), Bayesian vector autoregressive models (see Matkovskyy, 2012) etc. However, another class of models, namely neural networks, have also been found to be very effective at time series forecasting (Bacha & Meyer, 1992). One type of model, based on neural networks, namely nonlinear autoregressive exogenous (NARX) models, combines past values of the same series and present and past values of exogenous series. Additionally, this model includes an error (or residual) term. NARX models have broad applications, which include predictions, simulations and identifications. Some of its advantages include the approximation and simplicity of control design as well as simplicity and the fast convergence of identification.

The aim of this research is to use a constructed Index of Financial Safety of South Africa, estimated in Matkovskyy (2012), to predict the future conditions of financial safety in South Africa through the application of nonlinear autoregressive exogenous (NARX) modelling with different learning algorithms, as one of the most effective models of neural networks.

The methodological base of the current research is formed by means of the macro-prudential approach, system analyses, the basic principles of the theory of logical inference, the principal of parsimony, principal component analysis and neural networks, especially nonlinear autoregressive exogenous (NARX) models with different training algorithms (including the Levenberg-Marquardt algorithm and Bayesian regulation).

The structure of this paper is as follows. Section 2 briefly focuses on the theory and method underlying the construction of the Index of Financial Safety of a country. In Section 3, the methods of modelling and forecasting the IFS, based on the nonlinear autoregressive with exogenous inputs (NARX) model, are presented. In Section 4, the empirical results for the estimation of the IFS (as explained in Matkovskyy, 2012) and the forecasting of the IFS of South Africa are presented. The paper concludes in Section 5.

2. THE INDEX OF FINANCIAL SAFETY OF SOUTH AFRICA

For the purpose of this paper, the financial safety of a country is defined as a state in which the financial system, and all elements of this system, is shielded against real and potential internal and external threats. In other words, financial safety indicates a very small probability of the appearance of a crisis in a financial system (see Matkovskyy (2012) for a more detailed discussion).

The estimation of an index of financial safety of South Africa is based on the identification of key indicators. Such indicators should be able to capture the financial system's functions on macro-level that provide leading information on future performance and are relatively easy to estimate and use.

According to Matkovskyy (2012), the following indicators are used to estimate an index of financial safety for South Africa:

TABLE 1: The list of the indicators of financial safety estimation

Financial safety indicators	Character of financial safety indicators*
Money in circulation/M2*100	non-stimulant
Money in circulation/GDP*100	non-stimulant
M1/M2*100%	stimulant
M2/GDP*100	stimulant
M2/M0	stimulant
PPI / WPI	non-stimulant
Money market interest rates %	non-stimulant
GDP/M2	stimulant
M2/Money in circulation	non-stimulant
Monetary base/reserves	Stimulant
Coverage of import by reserves	Stimulant
Total domestic credit/GDP	non-stimulant
M2/ market capitalisation	non-stimulant
Changes of share price index % to a previous quarter	Stimulant
Real effective exchange rate	non-stimulant

*The difference between stimulants and non-stimulants lies in the nature of the influence, i.e. direct or indirect. The relationship between the integral estimation and indicator's stimulants is direct, and the relationship between and the indicator's non-stimulants is indirect

The process of aggregation of raw signals is based on the logic of the superposition principle (Oppenheim, Willsky & Nawab, 1997; Hsu & Hwei, 1995) applied to Economics. The selected indicators have different information "directions" that complicate the aggregation of the indicators. In order to perform the additive aggregation, it is therefore necessary to normalise the information contained in each indicator. The method of normalisation entails an equalisation of empiric (x_i) values with the optimum (x_{optim}) values, cordon (x_{cordon}) values, and extreme ($x_{extreme}$) values.

In order to transform the raw data, which are possibly strongly correlated between themselves, into new, uncorrelated components' factors, the data are further normalised using factor analysis (especially principal component methodology). To ensure the transformation of the data into the new set with values from '0' to '1', a varimax rotation is applied (Kaiser, 1958).

The following time series are used for the construction of an Index of Financial Safety for

South Africa (source of data: International Financial Statistics database; 1992Q1-2011Q1): M0, M1, M2, M3, money in circulation, GDP, total reserves (minus gold), exchange rate (ZARs per USD), real effective exchange rate, imports, money market interest rate, share prices, and market capitalisation. Data of total domestic credit was obtained from the Reserve Bank of South Africa. These raw data series are treated as input for the IFS of South Africa's estimation and prediction. To calculate the integral index of financial safety (IFS) of a country (in our case, of South Africa), one should sum the multiplications of weight coefficients and normalised values of indicators. The detailed process of the estimation of the Index of Financial Safety of South Africa can be found in Matkovskyy (2012). The graphical representation of it is provided in Figure 1.

3. NONLINEAR AUTOREGRESSIVE WITH EXOGENOUS INPUTS (NARX) MODEL

Throughout this study, the nonlinear autoregressive with exogenous inputs (NARX) model is used to model and forecast. This type of model was chosen because a properly designed NARX is equal to an observable general recurrent neural network, which, in turn, can emulate any nonlinear, dynamic state space model (Haykin, 1999).

3.1. The structure of NARX

The NARX is a recurrent dynamic network, with feedback connections enclosing several layers (incl. hidden layers) of the network. This model is characterised by non-linear relations between past inputs, past outputs and the predicted process output and can be delineated by a high order differential equation. The defining equation for the NARX model is as follows (Leontaritis & Billings, 1985; Siegelmann *et al.*, 1997; Haykin, 1999; Lin *et al.*, 2000; Menezes Jose *et al.*, 2006 *etc.*):

$$y_t = f\left(y_{t-1}, y_{t-2}, \dots, y_{t-n_y}, u_{t-1}, u_{t-2}, \dots, u_{t-n_u}\right) + e_t\ , \qquad (1)$$

where the next value of the dependent process y_t is regressed on previous values of the output signal (y) and previous values of an independent input signal (u); n_u and n_y are the corresponding maximum lags for input and output; and e_t explains the uncertainties and possible noise. The function f is a nonlinear function. The system input vector with a known dimension $n = n_y + n_u$ is:

$$\bar{X} = [y_{t-1}\ y_{t-2} \dots\ y_{t-n_y}\ u_{t-1}\ u_{t-2} \dots\ u_{t-n_u}]'. \qquad (2)$$

The function f is unknown. Therefore, it is approximated by the regression model of the form:

$$y_t = \sum_{i=0}^{n_u} a_i . u_{t-i} + \sum_{j=0}^{n_y} b_j . y_{t-j} + \sum_{i=0}^{n_u} \sum_{j=1}^{n_u} a_{i,j} . u_{t-i} . u_{t-j} + \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} b_{i,j} . y_{t-i} . y_{t-j} + \sum_{i=0}^{n_u} \sum_{j=1}^{n_y} c_{i,j} . u_{t-i} . y_{t-j} + e_t \ , \tag{3}$$

where a_i and $a_{i,j}$ are the linear and nonlinear coefficients of the originating exogenous term, respectively; b_j and $b_{i,j}$ are the linear and nonlinear coefficients of the autoregressive term, respectively; $c_{i,j}$ are the coefficients of the nonlinear terms.

The implementation of the approximation of the function *f* also allows for a vector ARX model to be estimated (i.e. the input and output can be multidimensional). Therefore, equation (3) may be re-written in the matrix form:

$$\begin{bmatrix} y_t \\ y_{t+1} \\ \vdots \\ y_{t+n_y} \end{bmatrix} = a.u' + b.y' + A.[U]' + B.[Y]' + C.[X]', \tag{4}$$

where

$$a = [a_0 \ a_1 \ \dots \ a_{n_u}]' \ , \tag{5}$$

$$b = [b_1 \ b_2 \ \dots \ b_{n_y}]', \tag{6}$$

$$A = [a_{0,0} \ a_{0,1} \ \dots \ a_{0,n_u} a_{1,1} \ \dots \ a_{n_u,n_u}]', \tag{7}$$

$$B = [b_{1,1} \ b_{1,2} \ \dots \ b_{1,n_y} b_{2,2} \ \dots \ b_{n_y,n_y}]', \tag{8}$$

$$C = [c_{0,1} \ c_{0,2} \ \cdots \ c_{0,n_y} c_{1,1} \ \cdots \ c_{n_u,n_y}]', \tag{9}$$

$$u = [u_{t-1} \ u_{t-2} \ \dots \ u_{n_y}], \tag{10}$$

$$y = [y_{t-1} \ y_{t-2} \ \cdots \ y_{n_y}], \tag{11}$$

$$U = [u_t . u_t u_t . u_{t-1} . u_{t-2} \dots u_t . u_{t-n_u} . u_{t-1} . u_{t-1} \dots u_{t-n_u} u_{t-n_u}], \tag{12}$$

$$Y = [u_t . y_{t-1} u . y_{t-2} \dots u_t . y_{t-n_y} u_{t-1} . y_{t-1} \ \cdots \ u_{t-n_u} . u_{t-n_y}], \tag{13}$$

Alternatively, equation (4) may be re-written as:

$$y_t = [u' y' U' Y' X'] \begin{bmatrix} a \\ b \\ A \\ B \\ C \end{bmatrix}. \tag{14}$$

To simplify we denote:

$$\bar{Y} = y_t, \tag{15}$$

$$\bar{U} = [u'y'U'Y'X'], \tag{16}$$

and

$$\bar{C} = [a\ b\ A\ B\ C]'. \tag{17}$$

Which means that equation (14) is now simplified as:

$$\bar{Y} = \bar{U}.\bar{C} \tag{18}$$

while the solution to the problem can be presented by:

$$\bar{C} = \bar{U}\backslash\bar{Y}. \tag{19}$$

3.2. Identification of the number of neurons in the hidden layer

The identification of the number of neurons in the hidden layer has no commonly accepted definition. The best number of hidden units depends, in a complex way, on: (i) the numbers of input and output units; (ii) the number of training cases; (iii) the amount of noise in the targets; (iv) the complexity of the function or classification to be learned; (v) the architecture of the type of hidden unit activation function; (vi) the training algorithm; and (vii) regularisation.

There are a number of theories for the optimal neural-network size determination. For example:

- the final prediction error, *FPE* (Akaike, 1970) or generalized final prediction error, *GPE* (Moody, 1992);
- the Network Information Criterion, *NIC* (Amari, 1995; Ripley, 1995 *etc*.), which is a generalisation of the Akaike Information Criterion;
- the Vapnik-Chervonenkis, *VC,* dimension (Cohn & Tesauro, 1992; Bartlett, 1993; Maass, 1995 *etc*.).

The *NIC* depends on a single, well-defined minimum and can be unreliable in the case of several local minima (Ripley, 1995). The evaluation of *NIC* or *GPE* is expensive for large networks, while the computation of *VC* for practical networks is also difficult (for an in-depth discussion on the problem of number of neurons, please refer to Lawrence *et al.,* 1996; Elisseeff & Paugam-Moisy, 1997).

On the other hand, there are algorithms like the Cascade Correlation (Fahlman & Lebiere, 1991), which starts with a minimal network, and then adds hidden nodes during the training.

A good result may also be obtained from experiments with multiple architectures, which select the one that performs best on held-out data. This approach will be used in our case to forecast the IFS for South Africa.

3.3. Performance estimation

The forecasting performance of the system model is mainly compared using a mean square error normalised with respect to the variance of the target signal. Mean Squared Error (MSE) is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2, \tag{20}$$

where t_i – are target outputs, a_i – are network outputs. The use of MSE will allow the comparison of the empirical results from this study with the results obtained by Matkovskyy (2012).

3.4. Algorithm for training

Training is performed using the Levenberg-Marquardt algorithm and Bayesian Regularization. The Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) is an iterative technique that locates the minimum of a multivariate function, which is expressed through the sum of squares of non-linear real-valued functions. The advantage of this algorithm is that it is guaranteed to converge.

The basis of the Levenberg-Marquardt algorithm is the linear approximation to *f*. When the performance function is in the form of a sum of squares, the Hessian matrix is approximated in the following way:

$$H = J'J, \tag{21}$$

and the gradient is calculated as:

$$g = J'\epsilon, \tag{22}$$

where *J* is the Jacobian matrix with the first derivatives of the network errors with respect to the weight and biases; ϵ is a vector of network errors. Next, the algorithm applies this approximation to the Hessian matrix:

$$x_{k+1} = x_k - [J'J + \mu I]^{-1}J'\epsilon. \tag{23}$$

If the scalar $\mu = 0$ in equation (23), the algorithm becomes Newton's method. If not, this method becomes a gradient descent with a small step size. A detailed analysis of the Levenberg-Marquardt algorithm is beyond the scope of this research (for a more comprehensive discussion please refer, among others, to Nielsen, 1999; Nocedal & Wright, 1999; Kelley, 1999).

Bayesian Regularization takes place within the Levenberg-Marquardt algorithm described above (see MacKay (1992); Foresee & Hagan (1997) for more detailed discussions of Bayesian Regularization). It minimises a linear combination of squared errors and weights. Therefore, this algorithm also modifies the linear combination. Its advantage is that it causes the resulting network, at the end of training, to have good generalisation qualities. In the algorithm, Bayesian Regularization back-propagation is used to calculate the Jacobian jX of performance, taking into account the weight and bias variables (X). Therefore, each variable is adjusted based on the Levenberg-Marquardt algorithm,

$$jj = jX * jX, \tag{24}$$

$$je = jX * E, \tag{25}$$

$$dX = -(jj + I * \mu)\backslash je, \tag{26}$$

where E is all errors; and I is the identity matrix.

4. EMPIRICAL RESULTS

The IFS of South Africa is considered as an output series. All calculations were made in MatLab. The general scheme of the NARX model of the IFS of South Africa is provided in Figure 2.

Both the Levenberg-Marquardt algorithm and Bayesian Regularization for training were applied to the network. The application randomly divides input vectors and target vectors into three sets, as follows: 75% are used for training; 15% are used to validate that the network is generalising and to stop training before overfitting; the last 15% are used as a completely independent test of network generalisation.

Table 2: The results of the NARX model training for the IFS of South Africa prediction (Levenberg-Marquardt algorithm)

	Target value	*MSE*	*R*
Training	59	4.02946e-17	1.000000
Validation	13	9.91773e-3	9.31977e-1

Testing	13	4.92878e-3	9.59343e-1

Hidden neurons = 48, delay = 4

As we can see from the Table 2, the small values of *MSE* and the high values of *R* give us reason to consider the obtained NARX model as adequate. If one compares the forecasting results to the ones obtained in the research of Matkovskyy (2012), using a BVAR (2) model with Minnesota prior, the MSE of training (4.02946e-17) is much lower than that of the BVAR(2) model (0,0000009). Compared to the BVAR(2), the same indicator for validation and testing is higher: 0,00991773 and 0,00492878, respectively.

The best validation performance is provided in Figure 3. It shows that the process of the network's performance improved during training. This performance is measured in terms of mean square error and it is shown in log scale. It is evident that the MSE is decreasing rapidly while the network is trained. In this case, the results are reasonable, because of the following:

- the final mean-square error is very small (less than 10^{-15});

- the test set error and the validation set error have similar characteristics.

In Figure 4, the training data indicate a good fit. The validation and test results also show *R* values greater than 0.9. Although it is not a perfect fit, the reason may be sought in the length of the time series – in this case, we do not have long time series.

Figure 5 shows how the error sizes are distributed. Typically, when most errors are near zero, it suggests a better trained model. In this case, we also have errors near zero.

The correlation between input and error is provided in Figure 6. This figure illustrates how the errors are correlated with the input sequence. The perfect prediction model means that all the correlations should be zero. In our case, all of the correlations are within the confidence bounds around zero.

The function of autocorrelations of errors is used to validate the network performance. Autocorrelation describes how the prediction errors are related in time. For the perfect model, there should be only one nonzero value of the autocorrelation at zero lag (this is the mean square error). This means that there is no correlation in prediction errors with each other and there is white noise. In our case, the correlations, except the one at zero lag, are within the 95% confidence limits around zero. Based on the various diagnostics described up to now, we may conclude that the model is adequate.

Figure 8 confirms that the responses, obtained from the NARX prediction model for the IFS of South Africa, are adequate, since the errors are quite small.

The predictions, obtained based on the Bayesian Regularization method of network training, do not improve the ones obtained based on ML network training (see Figures 9a and 9b).

Although the errors are not correlated with the input sequence, all the correlations are not within the 95% confidence limit around zero. The response of output element (the IFS of SA) in the NARX model, obtained using Bayesian Regularization, is also worse compared to the model trained by the Levenberg-Marquardt algorithm. This is illustrated in Figure 10.

5 CONCLUDING REMARKS

In this study, the nonlinear autoregressive with exogenous input (NARX) model with different network training methods was applied to the Index of Financial Safety (*IFS*) of South Africa to obtain high quality forecasts. The Levenberg-Marquardt algorithm and Bayesian Regularization for training of the network were used. The first conclusion of the paper is that although neural models may frequently suffer from a certain degree of inaccuracy, the results showed that the NARX model applied to the *IFS* of South Africa may ensure forecasts of adequate quality, while using less of computational expanses compared to Bayesian Vector-Autoregressive Models with different priors. The NARX model, with n_u=4 and n_y= 4 and the number of hidden neurons=48, was chosen since it resulted in the best performance, according to *MSE*. Therefore, the NARX models have the potential to capture the dynamic of nonlinear systems.

The second conclusion is that the NARX models are mainly dependent on the applied architecture and training method. Within the context of the architecture, the behaviour of NARX models mostly depends on the numbers of neurons in hidden layers. Too many hidden neurons in network cause over-fitting that, in turn, leads to poor predictions.

Future applications of the NARX model to *IFS* of South Africa may be based on a harmonically limited N-D Fourier transformation of the right-hand side of the NARX equation that entails the possibility of adaptive characteristics of NARX appearing. This new advantage of the NARX model will ensure the updating of the system of IFS monitoring with quarterly incoming data online that may be of interest to policy-makers and investors.

APPENDIX

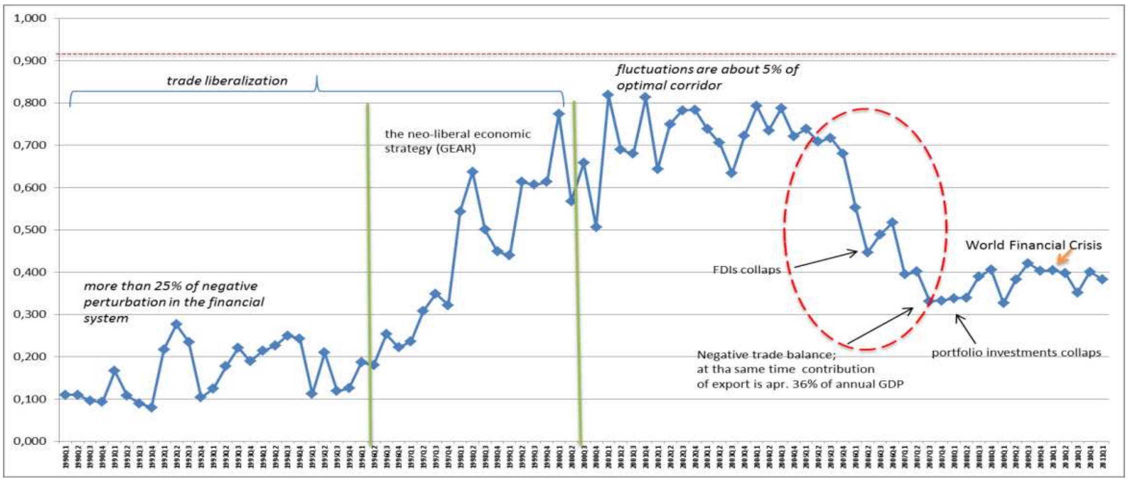

FIGURE 1: Dynamics of the estimated Index of Financial Safety (IFS) of South Africa, 1990Q1-2011Q1, (Matkovskyy, 2012)

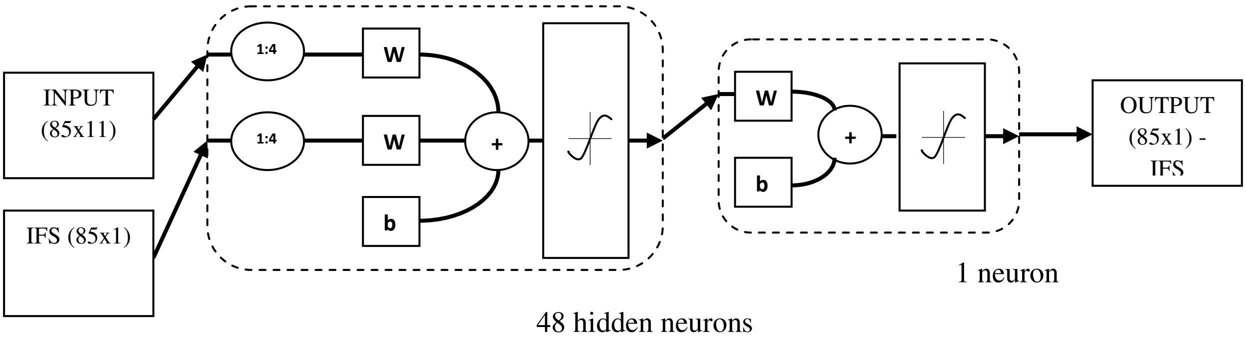

FIGURE 2: The general scheme of the Series-Parallel Architecture of the NARX model of the IFS of South Africa

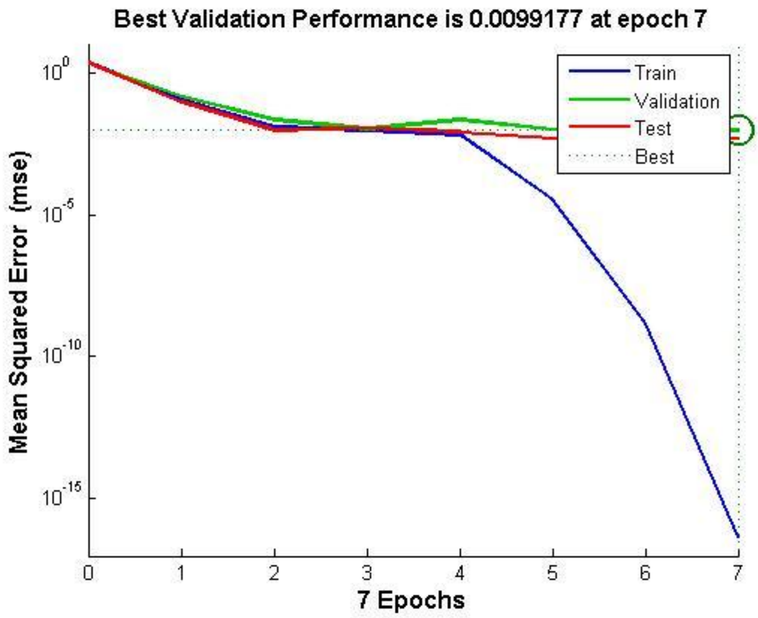

FIGURE 3: The process of the network's performance

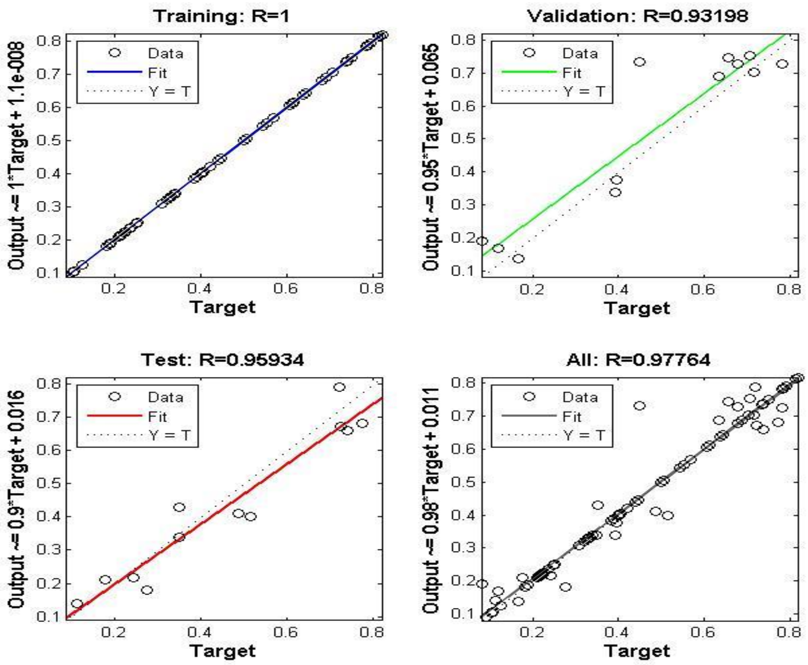

FIGURE 4: Regression plot for the NARX to predict IFS for South Africa

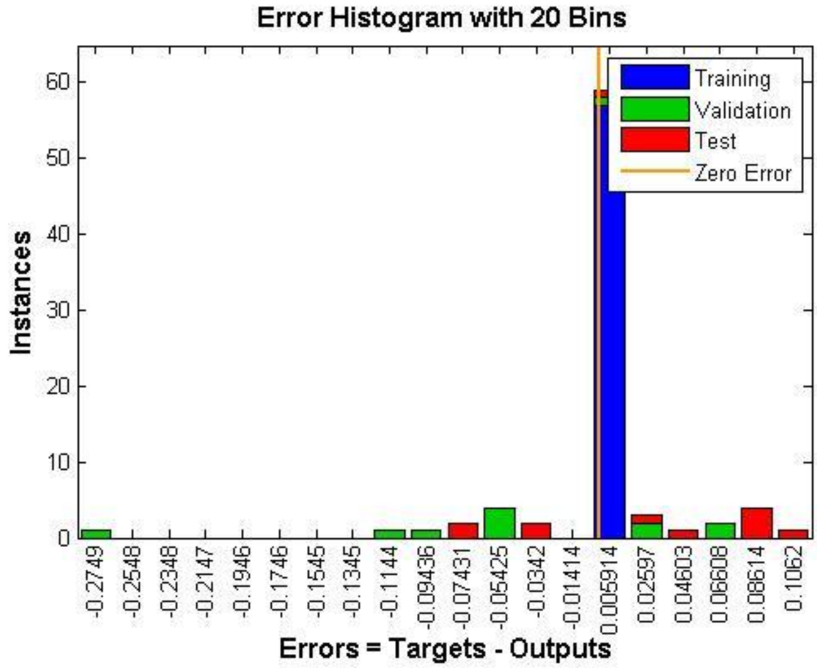

FIGURE 5: Error histogram of the NARX prediction model for IFS of South Africa

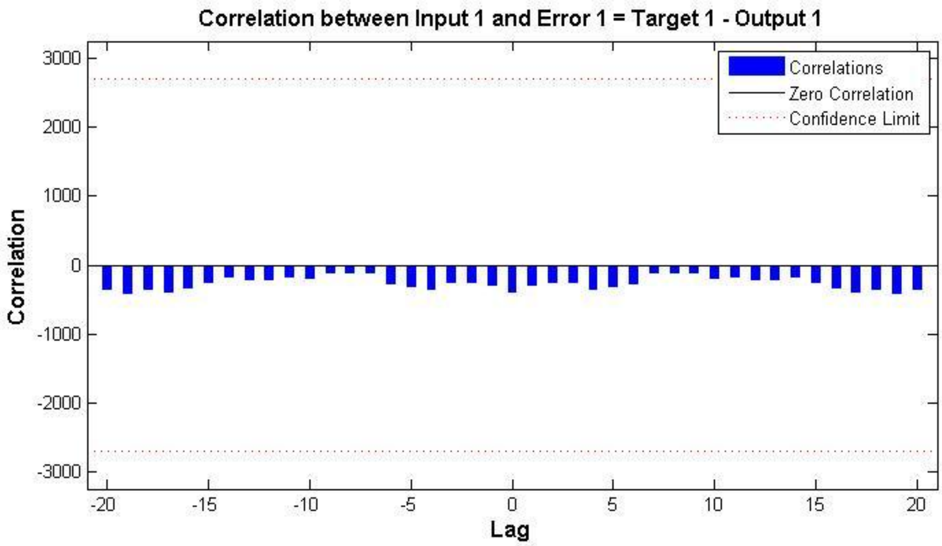

FIGURE 6: The input-output cross-correlation function

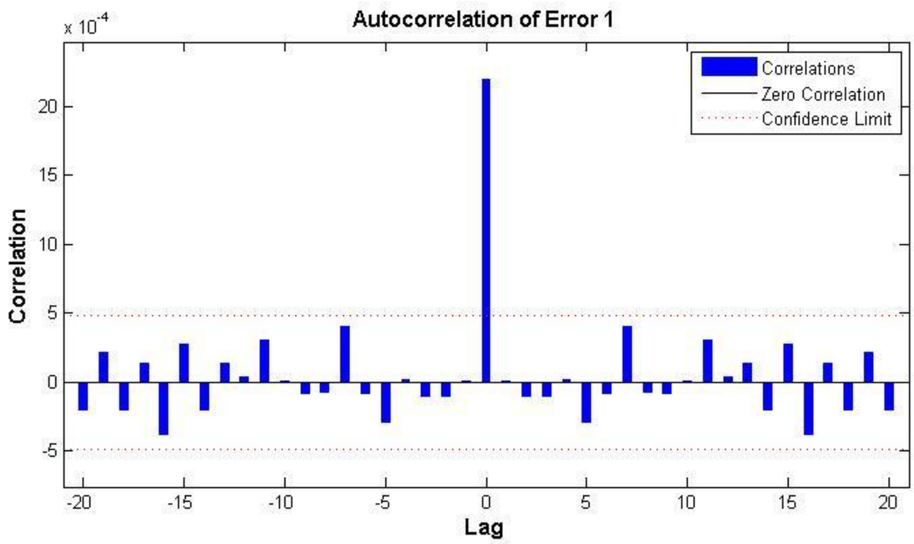

FIGURE 7: Autocorrelation of errors of NARX prediction model for IFS of South Africa

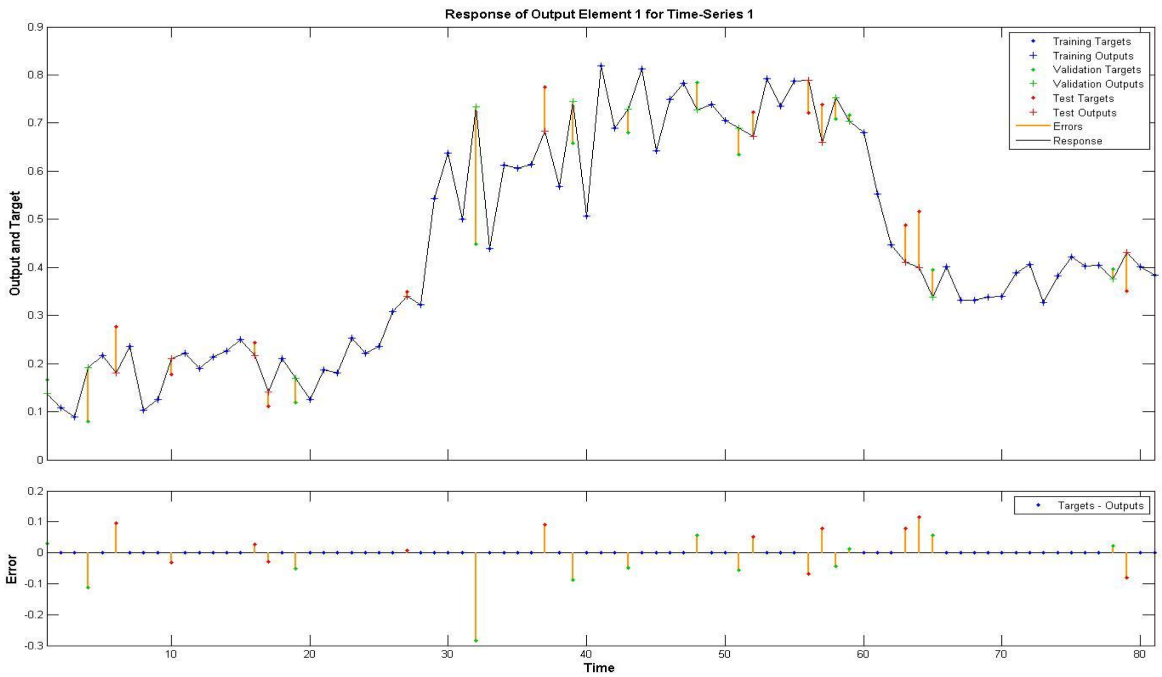

FIGURE 8: Response of NARX prediction model for IFS of South Africa (trained by the Levenberg-Marquardt algorithm)

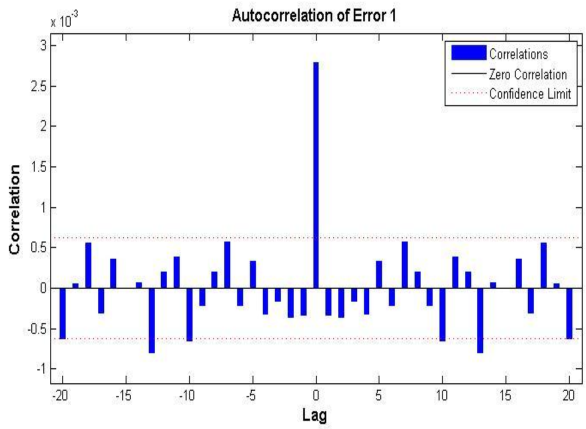

FIGURE 9a: Autocorrelation of errors of the SA IFS NARX prediction model trained by Bayesian Regulation

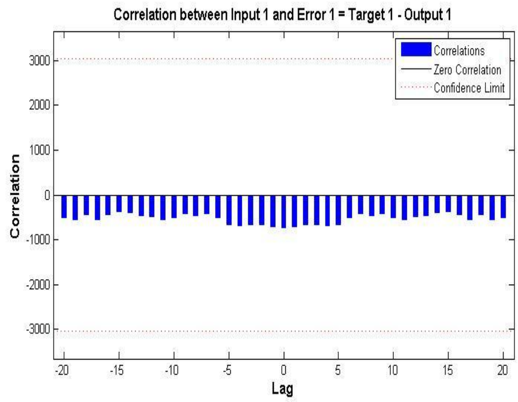

FIGURE 9b: The input-output cross-correlation function for the SA IFS NARX prediction model trained by Bayesian Regulation

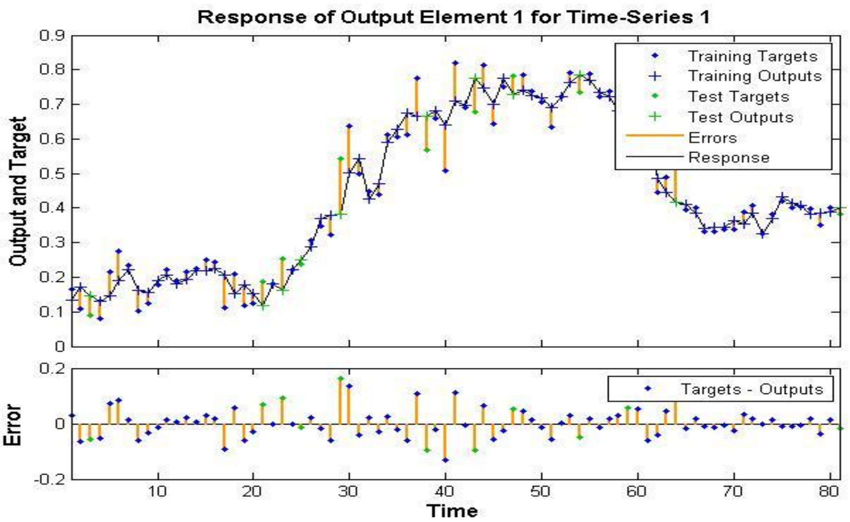

FIGURE 10: Response of NARX prediction model for IFS of South Africa (trained by the Bayesian Regulation method)

REFERENCES

ABIAD, A., G. (2003). Early Warning Systems: A Survey and a Regime-Switching Approach. *IMF Working Paper No. 03/32.*

AKAIKE, H. (1970). Statistical predictor identification, *Annals of the Institute for Statistical Mathematics* **22**, 203-217.

AMARI, S. (1995). Learning and statistical inference, in M. A. Arbib, ed., '*The Handbook of Brain Theory and Neural Networks*', MIT Press, Cambridge, Massachusetts, 522-526.

BACHA, H. & MEYER, W. (1992). *Neural Network architecture for Load Forecasting, Proceedings of International Joint Conference on Neural Networks*, IEEE Publisher, 442-447.

BARTLETT, P. (1993). Vapnik-Chervonenkis dimension bounds for two-and three-layer networks, *Neural Computation* **5**(3), 371-373.

COHN, D. and TESAURO, G. (1992). How tight are the Vapnik-Chervonenkis bounds?, *Neural Computation* **4**(2), 249-269.

ELISSEEFF, A., and PAUGAM-MOISY, H. (1997). Size of multilayer networks for exact learning: analytic approach". *Advances in Neural Information Processing Systems* **9**, Cambridge, MA: The MIT Press, 162-168.

FAHLMAN, S. E. and LEBIERE, C. (1991). The Cascade-Correlation Learning Architecture, Available at http://www.cs.iastate.edu/~honavar/fahlman.pdf [Accessed 07 February 2012].

FORESEE, F.D. and HAGAN, M.T. (1997). Gauss-Newton approximation to Bayesian regularization, *Proceedings of the 1997 International Joint Conference on Neural Networks*, pp. 1930-5.

FRASER, A. M., SWINNEY, H. L. (1986). Independent coordinates for strange attractors from mutual information, *Physical Review,* **A 33**, 1134-40.

GIRGUIS, S., AHMED, K. M., El MAKKY, N. M., & HAFEZ, A. M. (2006). *Mining the Future: Predicting Itemsets' Support of Association Rules Mining*. Retrieved January 28, 2007, Available at: http://www.cs.pitt.edu/~shenoda/files/miningthefuture-fdm06.pdf [Accessed 07 February 2012].

HAYKIN, S. (1999). *Neural Networks, A Comprehensive Foundation*, 2nd Edition, Prentice Hall.

HSU, HWEI P. (1995). *Signals and Systems*. McGraw-Hill. p.18.

KAISER, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**(3): 187-200.

KELLEY, C.T. (1999). *Iterative Methods for Optimization*. SIAM Press, Philadelphia.

KOOPMAN, S. J., LUCAS, A. and SCHWAAB, B. (2010). Macro, frailty, and contagion effects in defaults: Lessons from the 2008 credit crisis. *Tinbergen Institute Discussion Paper* **2010-004/2**, 1–40.

LAWRENCE, S., GILES, C.L., and TSOI, A.C. (1996). What size neural network gives optimal generalization? Convergence properties of backpropagation. *Technical Report UMIACS-TR-96-22 and CS-TR-3617, Institute for Advanced Computer Studies, University of Maryland*, College Park.

LEONTARITIS, I. L., BILLINGS, S.A. (1985). *Input-output parametric model for nonlinear systems – Part II: stochastic nonlinear systems,* Int.J.Control, **41**(2): 329-344.

LEVENBERG, K. (1944). A Method for the Solution of Certain Non-linear Problems in Least Squares. *Quarterly of Applied Mathematics*, **2**(2): 164-168.

LIN, TSUNGNAN, HORNE, BILL G., TINO, PETER & LEE, G., C. (2000). Learning long term dependencies in NARX recurrent neural networks, *Recurrent neural networks design and applications*, CRC Press, 133-146.

MAASS, W. (1995). Vapnik-Chervonenkis dimension of neural networks, in M. A. Arbib, ed., *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge, Massachusetts, 522-526.

MACKAY, D.J.C. (1992). Bayesian Interpolation, *Neural Computation*, **4**(3): 415-447.

MARQUARDT, D.W. (1963). An Algorithm for the Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal of Applied Mathematics*, **11**(2): 431-441.

MATKOVSKYY, R. (2012). The Index of the Financial Safety (IFS) of South Africa and Bayesian estimates for IFS Vector-Autoregressive Model. Forthcoming.

MENEZES, Jose M.P. & BARRETO, G. A. (2006). A new look at nonlinear time series prediction with NARX recurrent neural network, *Proceedings of Ninth Brazilian Symposium on Neural Networks*, pp.160-165, 23-27 Oct. 2006.

MOODY, J. (1992). The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems, in J. Moody, S. J. Hanson and R. P. Lippmann, eds, *Advances in Neural Information Processing Systems*, **Vol. 4**, Morgan Kaufmann, San Mateo, CA, 847-854.

NIELSEN, H.B. (1999). Damping Parameter in Marquardt's Method. *Technical Report IMM-REP-1999-05*, Technical University of Denmark, 1999. Available at http://www.imm.dtu.dk/˜hbn.

NOCEDAL, J. and WRIGHT, S.J. (1999). *Numerical Optimization*. Springer, New York.

RIPLEY, B. (1995). Statistical ideas for selecting network architectures, *Invited Presentation, Neural Information Processing Systems* **8**.

SIEGELMANN, HAVA T., HORNE G. & LEE G.C. (1997). Computational capabilities of recurrent NARX neural network" IEEE Transactions on Systems, *Man and Cybernetics, Part B*, **27**(2): 208-215.

OPPENHEIM, A. V., WILLSKY, A. S., & NAWAB, S. H. (1997). *Signals & Systems*. Prentice Hall.