CBN Journal of Applied Statistics (JAS)

Volume 7 | Number 2 Article 6

12-2016

On Building Inference for the Statistical Neural Network with Application to Naira-Dollar Exchange Rate Efficiency: A Bootstrap Approach

Christopher Godwin Udomboso Dr
University of Ibadan, Ibadan, Nigeria

Francisco U. Saliu
University of Ibadan, Ibadan

On Building Inference for the Statistical Neural Network with Application to Naira-Dollar Exchange Rate Efficiency: A Bootstrap Approach

Christopher G. Udomboso[1] and Francisco U. Saliu[2]

In this study, we developed an inference procedure for the neural network using the bootstrap approach, and applied it to the market efficiency of the Nigerian exchange rate. Data used are exchange rate values from 2001 to 2015. We conducted a test on the market efficiency hypothesis, including test for relevance of individual and joint network inputs using method of partial derivative. The network architecture used is the multilayer perceptron. A valid statistical inference based on the estimated Statistical Neural Network was conducted using a well-known statistical resampling technique. Test of hypothesis that input or groups of inputs are relevant to a model was carried out at 1% and 5% levels of significance. Evaluation of model was carried out using the Durbin Watson and Ljung-Box tests (the test statistic are obtained are 1.98 and 0.5831 respectively). The tests showed that the residuals were independently and identically distributed and had no serial autocorrelation in the series. The exchange rates do appear to contain information that is exploitable for enhanced point prediction. Hence, the Implication of this is that there are possibilities of abnormal earning for the agents in this market.

Keywords: Nigerian exchange rate, bootstrap, inference, neural network, efficiency

JEL Classification: C450

1.0 Introduction

Neural networks are interesting because of their potential use in areas of forecasting and classification, and have been used for a wide variety of application where statistical methods are traditionally employed. For example, Prybutok et al. (2000) successfully applied Neural Network to systems such as ozone concentration prediction, and many others. According to Minns and Hall (1996), while Artificial Neural Networks (ANNs) can predict the output variable of a system based on a set of input variables, the inner workings of the network are not easily understood. This is because the ANN training and application usually involves the composition of nonlinear functions that can

[1] Department of Statistics, University of Ibadan, Ibadan, Nigeria, cg.udomboso@gmail.com
[2] Department of Statistics, University of Ibadan, Ibadan, Nigeria, francis_saliu@yahoo.com

be difficult to simplify and reduce to terms that can be understood physically. This also makes it difficult to analyse the network after it has been built and trained to determine the relative importance of the different input variables in predicting the output.

Over the years, mainstream economist (classical and neoclassical), have consistently maintained that, an unregulated market price is the best yardstick reflecting true scarcity or worth of a commodity. Belatedly, efficient market hypothesis is based on the notion that stock prices are informational efficient–reflecting all available information about the value of an asset in the financial market at every moment. This therefore implies that in an efficient market, stock prices are equal to the true worth of the stock, defined as discounted future cash flows (Idoloh and Braimah, 2015).

According to efficient market hypothesis, changes in stock prices are impossible to predict from available public information and the only thing that can move stock price is news that changes market's perception of a firm's asset value. Thus when good news about a firm's prospect becomes public, the value and stock price of the firm both appreciate and when the company prospect deteriorates both the value and stock price of the firm depreciates. This claim by the efficiency market hypothesis that neither technical analysis which is the study of past stock prices in an attempt to predict future stock price nor even fundamental analysis, which is the analysis of financial information such as company's earnings, asset price, etc., which is thought to be important indices for choosing undervalued shocks; have been confronted with mix reactions from researchers, academics and policy analyst (Malkiel, 2003).

Fama (1965) defined three forms of efficiencies and distinguished them by the information set considered. These include the weak, semi-strong, and strong efficiencies. In weak efficiency, the information set available is in the history of the series, that is, in its past values; while in semi-strong efficiency, the information set is made of all available public information (here the market instantly absorbs all the information as soon as it becomes publicly available), and in strong efficiency, the information set is made of all available public and private information, that is, also considering privileged information (here no agent could have abnormal earnings even if he has access to confidential data). Following the pioneering work by Fama (1965) on the US stock market, a number of studies have attempted to test the efficiency market hypothesis in different stock markets of the world (Vitali and Mollah, 2010; Aga and Kocaman, 2008; Cavusoglu, 2007).

Several researchers have started to develop strategies based on statistical approach to model selection in neural networks. Kaashoek and Van Dijk (2002) proposed backward method, which starts from a simple model using an algorithm that reduces the number of parameters based on R^2 increment and the principal component analysis of network residuals until an optimal model is attained. Swanson and White (1997) applied a criterion of model selection; the Schwarz Information Criterion (SIC), on "bottom-up" procedure to increase number of unit nodes in hidden layer and select the input variables until finding the best Feed Forward Neural Network (FFNN) model. This procedure is also recognized as "constructive learning" and one of the most popular is "cascade correlation" and it can be seen as "forward" method in statistical modelling.

Inference in Artificial Neural Network (ANN) is a bit complicated since the estimators under the null hypothesis is itself complicated. Racine and White (2001) in their paper proposed a test for individual and joint irrelevance of network inputs. They used a Monte Carlo experiment to show that their test had reasonable level and power behaviour and then applied the method to examine if there are irregularities in foreign exchange rates. The objective of this study are therefore outlined as follows:

1. Model the Nigerian Dollar-Naira from 12/10/2001 to 4/20/2015 using the best Statistical Neural Network Model which shall be selected using the selection criterion; namely, Mean Squared Error, Coefficient of Determination, Akaike Information Criterion, Network Information Criterion and Schwarz Information Criterion.

2. Test the Market Efficiency Hypothesis of the best model obtained using the bootstrap experiment technique.

In other to carry out the objectives, the following researches were reviewed;
White (1989) provided statistics that can be used to test hypotheses of interest regarding optimal network weight, for example, that a given input or set of inputs is irrelevant in the sense that the optimal weights on connections from the inputs in questions to hidden units in the next layer are zero.
Dhoriva et al (2008) presented a statistical procedure based on hypothesis testing to build neural network model in multivariate time series case. Their

method involves strategies for specifying the number of hidden units and input variables in the model using inference in the R^2 increment. The result shows that the statistical inference can be applied successfully for modelling neural networks in multivariate time series analysis.

Udomboso (2014) in his doctoral thesis compared the levels of precision of different statistical neural networks using model criterion such as Adjusted Network Information Criterion and Network Information Criterion. The author compared the level of precision of the Heterogeneous Statistical Neural Network and Homogeneous Statistical Neural Network. The result showed that the Heterogeneous Statistical Neural Network performed better.

Howard et al (1998) used statistical inference to provide an objective way to derive learning algorithm, both for training and evaluating the performance of trained models. They proposed a solution to the over-fitting problem by model-selection methods using either conventional statistical approach or Bayesian approach. Subanar et al (2005) discussed and proposed a procedure for model selection in neural networks for time series forecasting. They focused on model selection strategies based on statistical concepts particularly on the inference of R^2 incremental, Principal Component Analysis (PCA) of the residual model and Schwarz Information Criterion (SIC). The results show that statistical inference of R^2 incremental combining by SIC criteria is an effective procedure for model selection in neural networks for time series forecasting.

Walde (2003) used the estimate from neural network modelling as a basis for formal statistical inference. The author tested the hypothesis that the dynamics of biomass distribution can be captured with the help of geo-registered and ortho-rectified colours images from the opposing hill slope. With the help of the bootstrap technique, the significance of the colour pattern for modelling phytomass was demonstrated.

Andreza and Alexandre (2006) used Artificial Neural Network to test the hypothesis of market efficiency for the Brazilian exchange rate from 1999 to 2004. The authors used the method of partial derivative suggested by Racine and white (2001) to choose the network architecture, estimate the model weight and then test the hypothesis. The result of the study showed that the

Brazilian exchange rare market is not efficient and therefore shows that there is a possibility of abnormal earning for the agents in the market.

2.0 Methodology

2.1 Data description

Data for this study are from the Nigerian exchange rate (purchase rate) in the period from 10th of December 2001 till 20th of April 2015, resulting from 3486 data point, obtained from the central bank website, Only Naira -Dollar exchange rate was considered.

2.2 Statistical Neural Network

The statistical neural network proposed by Anders (1996) is given as

$$y_t = f(X_t, w) + e_t \tag{1}$$

Where y_t is the dependent variable, $X_t = (x_0 \equiv 1, x_{t1}, ..., x_{tI})$ is a vector of independent variables, $w = (\alpha, \beta, \gamma)$ is vector network weight: $'\alpha'$ is the weight of the input unit, $'\beta'$ is the weight of the hidden unit, and $'\gamma'$ is the weight of the output unit, and e_t is the stochastic term that is normally distributed (that is, $e_t \sim N(0, \sigma^2 I_n)$). The assumption of the statistical neural network is the same as the usual regression models. Basically, $f(X, w)$ (Anders, 1996) is the artificial neural network function expressed as;

$$f(X_t, w) = \alpha X_t + \sum_{h=1}^{H} \beta_h g\left(\sum_{i=0}^{I} \gamma_{hi} x_i\right) \tag{2}$$

Where $g(.)$ is the transfer function, $g : X \to Y$.

For the purpose of this work, the Tangent sigmoid (TANSIG) transfer function will be used. By definition

$$\tan sig = f(n) = \frac{2}{1 + e^{-2n}} - 1$$

2.3 Significance of input

With the help of the optimised network, we want to test whether or not a given set of input are significant. Racine and White (2001) investigated the statistic and its distribution for the null hypothesis that some (or all) input variables are not significant.

From (1)

$$Y_t = f(x_t, w) + \varepsilon = g(x) + \varepsilon$$

The optimised model using the least squared method gives

$$Y_t^* = f(x_t, w^*) + \varepsilon \tag{3}$$

We are interested in whether certain elements of x_t have any effect on Y_t. Thus, we shall test the hypothesis that:

$$H_0 = \frac{\partial f(x, w^*)}{\partial x_i} = 0 \text{ for all } i \in I_0 \tag{4}$$

Equation (4) can be interpreted as testing the hypothesis that the inputs specified by I_0 do no enter the approximation of the target itself Y_t provided using weight w^* (i.e. $f(x_t, w^*)$).

Consider the quantity;

$$m^* = \sum_{i \in I_0} \int f_i(x, w^*)^2 \, d\mu(x) \tag{5}$$

Where $f_i(\cdot, w)$ denotes $\partial f(\cdot, w)/\partial x_i$. From (4), we have that $m^* = 0$

Two variables prevent us from calculating m^* directly. The first is that w^* is unknown; this can be estimated using the least square estimation. The second is μ, the distribution of X_t, is unknown; however, μ is consistently estimated by the empirical distribution $\hat{\mu}_n$ which place a mass n^{-1} at every realization observed in a training sample of size n. a computable statistic is therefore given as;

$$\hat{m}_n = n^{-1} \sum_{t=1}^{n} \sum_{i \in I_0} f_i(X_t, \hat{w}_n)^2 \tag{6}$$

Racine and white (2001) stated in their paper that under suitable conditions, the statistic is distributed thus;

$$n\hat{m}_n \sim N_2(0, C^*; M^*) \tag{7}$$

Where $N_2(0, C^*; M^*)$ is the mixture of independent chi-squares($\chi^2 s$) defined by White (1994) with C^* to be a finite positive semi-definite covariance matrix and $M^* = E(\nabla^2 m(X_t, w^*))/2$. ∇ is the gradient operator with respect to the weight.

The distribution (7) is not found in tables, but we can approximate it using the bootstrap technique.

Algorithm for testing the significance of the input is as follows;

i. Obtain the best Statistical Neural Network Model using the model criteria.
ii. Obtain the best predicted value with neural network; $\hat{y}$.
iii. Obtain the set of residuals; $\hat{e} = y - \hat{y}$.
iv. Randomly sample the vector, $\hat{e}$ with replacement and obtain the first set of shocks for the first bootstrap experiment, $\hat{e}^{b1}$.
v. With this set of first randomly sampled shocks from the base residuals, $\hat{e}^{b1}$, we generate a new dependent variable for the first bootstrap experiment, $y^{b1} = \hat{y} + \hat{e}^{b1}$ and use the new data set (y^{b1}, x) to re-estimate the neural network and obtain the partial derivatives and other statistics of interest.
vi. We then repeat this procedure for 100 or 500 or 1000 times, obtaining $\hat{e}^{bi}$ and y^{bi}.

2.4 Evaluation of Network estimation

2.4.1 Serial Independence: Durbin Watson and Ljung-Box tests

The Durbin Watson (DW) autocorrelation test is used to test for serial independence and constancy in the variance.

$$DW = \frac{\sum_{t=2}^{T} [\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}]^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2} = 1.98$$

Since DW is approximately equals to 2, we conclude that there is no autocorrelation in the series.

The Ljung-Box Q-Statistics is further used to examine the joint significance of the residual autocorrelation. That is the testing for autocorrelation at multiple lags. The hypothesis to be tested is stated thus;

H_0 : The data conforms to a white noise process.

H_1 : The data are do not conform to a white noise process.

$$Q = T(T+2)\sum_{m=1}^{M}\frac{\hat{\rho}_m^2}{T-m} = 0.5831$$

Where, T is the number of observation and m is the lag of the series.

Since the calculated value is less that $\chi_{0.05,3}$ we do not reject the null hypothesis, and therefore conclude at 5% level of significance that the data conforms to a white noise process.

2.4.2 Valid statistical inference in ANNs

Statistical inference will now be performed on this net to test the validity of the market efficiency hypothesis in its weak version.

We shall test the hypothesis that;

$H_0 : \dfrac{\partial f(x, w^*)}{\partial x_i} = 0$; (the network inputs are not relevant)

$H_1 : \dfrac{\partial f(x, w^*)}{\partial x_i} \neq 0$ (the-network-inputs are relevant)

2.5. Network Architecture

In this work, the architecture choice followed by Racine and White (2001) will be adopted.

Number of input: The number of input, that is, the number of independent variables is the number of lags of the exchange rate. Since our interest is to test the market efficiency hypothesis (in its weak form), we tested for the significance of the lags. For the purpose of this work, only lags 1 to 10 (that is inputs 1 to 10) is considered and the best model chosen by the Swartz Information Criterion (SIC), Akaike Information Criterion (AIC).

Number of Hidden Layers: Only one hidden layer is used for the purpose of the study. It has been theoretically shown that one hidden layer is enough to approximate any non-linear function (Hornik et al, 1989).

Number of Neurons in the hidden layer: We used 1 to 10 hidden units. The configuration with the least Swartz Information Criterion (SIC), Akaike Information Criterion (AIC), Mean Squared Error (MSE), Network Information Criterion (NIC) is chosen.

Number of output neurons: Since our interest is on the forecast of exchange rate, we required only one output neuron.

The notation q-p-o means that the network under consideration has "q" lags or input(s), "p" neurons in the hidden layer and "o" outputs.

3.0 Results and Discussions

The chosen network configuration based on the selection criteria for each input configuration is shown below. Table 1 shows the best model configuration for each inputs 1 to 10 and the corresponding criterions.

Table 1: Results of the inputs showing the chosen configurations

Input(s)	1	2	3	4	5	6	7	8	9	10
Chosen configuration	1-8-1	2-6-1	3-9-1	4-10-1	5-10-1	6-2-1	7-6-1	8-7-1	9-7-1	10-10-1
MSE	6.161	6.177	6.093	8.333	7.086	9.864	7.901	8.962	8.639	8.102
SIC	6.186	6.215	6.114	8.419	7.173	10.007	8.015	9.11	8.819	8.287
AIC	6.164	6.183	6.1	8.345	7.098	9.884	7.916	8.982	8.664	8.128
NIC	6.186	11.765	6.295	8.336	7.004	10.531	7.897	11.33	8.615	8.07

Table 1 shows the results of the ten inputs with the best model configuration.

The networks were trained using the Matlab 6.5 software and was considered trained after 500 epoch. We trained the networks using 1 to 10 input(s) (number of "independent variable", i.e. the number of lags of the exchange rate). For each hidden neuron 1 to 10 of each input(s), the model criterions namely; Schwarz Information Criterion (SIC), Akaike Information Criterion (AIC), Network Information Criterion (NIC) and Mean Squared Error (MSE) were calculated and the "most appropriate" network was selected. The table above displays the selection criterion values of the "most appropriate" network of each input(s). The network configuration notation p-q-r implies p inputs, q hidden neuron and r output. It can be seen above that the ANN with 3 inputs has the least MSE, SIC, and AIC (i.e. the network configuration 3-9-

1) and is therefore chosen. Statistical inference will be performed on this network.

3.1 Significance of Network Input

Table 2: Statistical inference on bootstrap estimate

Mean	1761.506
Standard error	11091.36
C_α at 1%	16519
C_α at 5%	5736.5
Statistic	22100.55

In this section, hypothesis test about the significance of the network input is carried out using the bootstrap technique (hypothesis statement is shown in Section 2.4.2). A 100 bootstrap samples are generated then for each bootstrap experiment, the estimated partial derivatives and other statistic of interest is computed. The set of estimated partial derivatives are then ordered from lowest to highest values and the probability distribution of these statistics are then obtained. The table above gives the test statistic and the significance value.

As shown above, the null hypothesis that the network input variables are not relevant is rejected at both 1% and 5% levels of significance.

4.0 Concluding Remarks

This study sets out to test the efficiency of the Nigerian exchange rate. The methodology of Racine and White (2001) was applied, and statistical inference was conducted on the SNN models used. The study tested the validity of the efficient market hypothesis on the Naira-Dollar exchange rate market in the period of 2001 to 2015.

In economics, the efficient market hypothesis states that "*prices already reflect all the available information, therefore it is impossible to outperform market prices*". This hypothesis could be correct if linear methodologies are used. However, as pointed out by Granger and Terasvirta (1993), most economic variables possesses non-linear relationships, as such, a linear model may not be able to fully explain it. The use of Non-linear model like the

Statistical Neural Network, which is known for its great capability of modelling non-linear relationships, has shown to be rather useful and suitable. Different SNN models were estimated at different lags (ranging from lag 1 to lag 10) and various hidden neurons (from 1 hidden neuron to 10) and the best model at each lag was selected using the selection criteria such as the coefficient of variation, Mean Squared error (MSE), Akaike Information Criterion (AIC), Network Information Criterion (NIC) and The Schwartz Information Criterion (SIC). The five least (increasing) ordered values of the AIC, SIC and MSE were observed to in the Network configurations; 3-9-1, 1-8-1, 2-6-1, 5-10-1 and 7-6-1respectively. For the NIC, the five least (increasing) ordered values were found to be the Network configurations; 1-8-1, 3-9-1, 5-10-1, 7-6-1 and 10-10-1.From the best models at each input (1 to 10), the overall best model was selected to have the Network configuration 3-9-1 since it has the least Mean Square Error, Akaike Information criterion and Schwartz Information Criterion.

The evaluation of model was carried out using the Durbin Watson and Ljung-Box tests. The tests showed that the residuals were independently and identically distributed and had no serial autocorrelation in the series.

The bootstrap technique introduced by Efron and Tibshirani (1993) was employed and 100 bootstrap experiments were obtained. The bootstrap statistic was computed from each bootstrap and the critical regions at 1% and 5% level of significance for the rejection of the null hypothesis were calculated. The null hypothesis that a given set of inputs has no effect was rejected and the alternative was accepted.

The hypothesis of no effect inputs was rejected at 1% and 5% level of significance (that is the past lags of the Exchange rate are not relevant). Therefore the study do not support the hypothesis of market efficiency for the Naira-Dollar exchange rate series for the period in question.
The main findings of this study can be summarised as follows:

The Statistical Neural Network, a non-linear model, was used to successfully model the Naira-Dollar exchange rates for the period in question. The selected model had the least Mean Square Error, Akaike Information Criterion and Schwartz Information Criterion.

The model selected was used to test the validity of the market efficiency hypothesis for the Naira-Dollar exchange rate in its weak form. The results do not support the market efficiency hypothesis as it shows that the inputs (that is the lags) are relevant. The implication of this is that the exchange rates do appear to contain some information that can be exploited for enhanced prediction. Therefore, there are possibilities of abnormal earning for the agents in this market.

In other to prevent arbitrary earning of agents partaking in the Nigerian exchange rate markets, policies should be implemented to cut the extra earnings in the market.

References

Aga, M., Kocaman, B. (2008).Efficient Market Hypothesis and Emerging Capital Markets. Empirical Evidence from Istanbul Stock Exchange, International Research Journal of Finance and Economics, Issue 13.

Andreza, P., Alexandre, S. (2006).Inference in Artificial Neural Network: Testing Market Efficiency for the Brazilian Exchange Rate. https://www.researchgate.net/publication/265527761_Inference_in_Artificial_Neural_Networks_Testing_Market_Efficiency_for_the_Brazilian_Exchange_Rate

Anders, U. (1996).Statistical Model Building for Neural Networks.AFIR Colloqium.Nunberg, Germany. http://www.actuaries.org/AFIR/Colloquia/Nuernberg/Anders.pdf

Cavusoglu, T. (2007).Testing the Efficient Market Hypothesis for the Athens Stock Exchange. 4th AFE Samos International Conference on Advances in Applied Financial Economics, 12-14 July 2007.

Dhoriva, U. W., Subanar, S. G.,Zanzawi, S. (2008) Statistical inference for modelling Neural Network in Multivariate Time Series. Journal ILMU DASAR, Vol. 9 No. 1, January 2008: page 15-21

Efron, B., Tibshirani, R. J. (1993). An introduction of the bootstrap. Chapman and hall/CRC: New York, ISBN 0-412-04231-2

Fama, E. F. (1965). The Behavior of Stock Market Prices, Journal of Business Vol. 38, pp 34-105.

Granger, C.W.J.; Teräsvirta, T. Modelling nonlinear economic relationships. Oxford University Press, 1993.

Hornik, K., Stinchcombe, S.M., & White, H. Multilayer feedforward networks are universal approximators. Neural Networks, 2, p.359-366, 1989.

Howard, H. Y., Noboru, M., Amari, S. (1998). A statistical Inference; Learning in Artificial Neural Network. Trend in cognitive Sciences – Vol. 2, No 1, 1 January 1998.

Idolor, E. J., Braimah, A. (2015). A test of Fama's Efficient Capital Market's II position in the Nigerian bourse, Indian Journal of Economics & Business, Vol. 14, No. 3, (2015) : page 375

Kaashoek, J.F., Van Dijk, H.K. (2002) Neural Network Pruning Applied to Real Exchange Rate Analysis, Journal of Forecasting, 21, pp. 559–577.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and its critics, the journal of Economic Perspectives 17(1), 59-82.

Minns, A. W.and Hall, M. J. (1996). Artificial Neural Networks as Rainfall-runoff models. Hydrological Science Journal.Vol. 41, No 3.

Prybutok, V. R., Yi, J., Mitchel, D. (2000).Comparison of Neural Network Models with ARIMA and Regression Models for prediction of Houston's Daily Maximum Ozone Concentrations.European Journal of Operational Research.Vol. 122.

Racine, J. White, H. (2001). Statistical inference, the bootstrap, and neural network modelling with application to foreign exchange rates.IEEE Transactions on Neural Networks, v. 12, n. 4, p.1-19, 2001.

Subanar, Guritno, S., Dan Hartati, S. (2005). Neural Network, PemodelanStatistik dan Peramalan Data Finansial.LaporanPenelitian HPTP Tahun I, UGM, Yogyakarta.

Swanson, N. R., H. White (1997), A model-selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks, Review of Economic and Statistics, 79, 540–550

Udomboso, C. G. (2014). On the level of precision of an heterogeneous statistical neural network model. Ph.D Thesis, University of Ibadan.

Vitali, F., Mollah, S. (2010). Stock Market Efficiency in Africa, Evidence from Random Walk Hypothesis, Mid-West Finance Association Annual Meeting, West Chicago River North, March 2-5.

Walde, J. F. (2003). Valid Statistical Inference based on Feedforward Artificial Neural Network Models. Proceeding of the seventh Young Statisticians Meeting.

White, H. (1994). Estimation, inference, and specification analysis. New York: Cambridge University Press, ISBN 0-521-25280-6.

White, H. (1989). “Learning in artificial neural networks: A statistical perspective”, Neural Computation Volume 1 Issue 4, 425-464, ISSN: 0899-7667. Dec. 1989,doi:10.1162/neco.1989.1.4.425.
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6796856&isnumber=6795285