Artificial neural network application in modeling revenue returns from mobile payment services in Kenya

Kyalo Richard, Waititu Anthony, Wanjoya Anthony

Jomo Kenyatta University Department of Statistics and Actuarial Science, Nairobi, Kenya

Email address:
Kyalorichard11@gmail.com (K. Richard), agwaititu@gmail.com (W. Anthony), awanjoya@gmail.com (W. Anthony)

Abstract: Artificial Neural Networks has recently shown a great applicability in time-series analysis and forecasting thus correctly deducing the unseen part of the population even if the sample data contain noisy information. In this paper we used Neural Network to model revenue returns from mobile payment services using dataset extracted from Central Bank of Kenya website. The network with one or two hidden layers was tested with various combination of neurons, and results were compared in terms of forecasting error. It was observed that ANN if properly trained accurately forecast Revenue returns on mobile payments services in Kenya.

Keywords: Neural Network, Quasi Newton, Forecasting, Generalization

1. Introduction

Mobile payment is defined as a payment services that is operated under financial regulation and performed from a mobile device using Sms technology in place of cash. Kenya is one of a small number of innovative countries where financial services are offered by mobile network operators to people and there is considerable interest in the development of these services since it offers the prospect of providing services to people more conveniently and faster. Government benefits by taxing the revenue collected from mobile payments service and thus the aim of this research is to model the revenue of the mobile payments services in Kenya using Artificial Neural Network based on the actual annual data from central bank.

Nowadays a consumer can use mobile phone to pay for goods and services Mobile payment is being adopted all over the world in different ways and approximately the combined market for all types of mobile payments is expected to reach more than $600 Billion globally by 2014.

For decades, Auto-Regressive Integrated Moving Average (ARIMA) technique and Box and Jenkins has been widely used for time series forecasting. ARIMA is a univariate model and it is based on the assumption that the time series data being forecasted are linear and stationary. The Artificial Neural Networks has recently shown a great applicability in time-series analysis and forecasting thus correctly deduces the unseen part of the population even if the sample data contain noisy information.

Multivariate models can rely on countless information, where apart from the lagged time series being forecast, but also other exogenous variables are combined to act as predictors. In addition, ANN is more effective in analysis of non-stationary time series due to its unique non-parametric and adaptive properties.

A multilayer forward feedback neural network is used widely in building up forecast models, because it has a nonlinear mapping capability which enables bipolar neural network to describe complicated nonlinear relationship by studying input-output sample datasets and generalizing capability which also enables bipolar neural network to predict when a new data is introduced into the network.

ANN is therefore a strong data modeling tool that can represent complex input/output relationships. This study defines the application of neural networks in modeling revenue returns from mobile payment services.

2. Previous Research

Though mobile payment is a new entity in the capital market neural networks have widely been used in share market prediction, forex exchange and forecasting of the various share price predictions, as well as for time series modeling, for instance Akinwale et.al (2009) used error back

propagation algorithm and regression analysis to analyze and predict untranslated and translated Nigeria Stock Market Price. Translated NSMP prediction approach was more accurate than untranslated NSMP using either regression analysis or error back propagation algorithm.

Cao and Parry (2009) investigated the models for forecasting earnings per share using neural networks (comparing error backward propagation and genetic algorithms) in 283 firms in different industries. They used a neural network with 7 input variables. The results reveal that the genetic algorithm makes a more accurate forecast of earnings than the error backward propagation algorithm.

Fu (1998) in his study concluded that the Neural Network method is more precise than the regression model. Zhang et al. (2004) studied models of per share earnings forecasting of neural networks with four kinds of models in 283 firms: one-variable linear models, multi-variable linear models, one-variable neural network and a multi-variable neural network. This research shows that the use of neural network methods provides more accuracy in forecasting than linear forecasting models. G. P. Zhang, B. E. Patuwo and M. Y. Hu [59] presents an experimental evaluation of neural networks for nonlinear time-series forecasting. The effects of three main factors (input nodes, hidden nodes and sample size) are examined through a simulated computer experiment. Results show that neural networks are valuable tools for modelling and forecasting nonlinear time series while traditional linear methods are not as competent for this task.

El Shazly and El Shazly (1997) examined the one month forecasting performance of a NNR model compared with the forward rate of the British pound (GBP), German Mark (DEM), and Japanese Yen (JPY) against a common currency,. Evaluation was based on forecasting accuracy and in terms of correctly forecasting the direction of the exchange rate.

Carney and Cunningham (1996) used four datasets over the period 1979 to 1995 to examine the single-step and multi-step prediction of the weekly GBP/USD, daily GBP/USD, weekly DEM/SEK (Swedish Krona) and daily GBP/DEM exchange rates. The neural network models were benchmarked by a naïve forecast and the evaluation was based on forecasting accuracy. The conclusion was that neural network models are useful techniques that can make sense of complex data that defies traditional analysis.

Mahdavi and Behmanesh (2005) investigated the forecasting of the stock prices of investment firms using artificial neural networks. The results of their investigation showed that if a neural network is trained properly, it will be able to identify the relation between the variables and will be effective in forecasting stock prices of investment firms with the least error(0.044).

3. Artificial Neural Network

Artificial Neural Network is an artificial representation of the human brain that tries to simulate its learning process. It is a collection of interconnected processing elements referred to as nodes or neurons, with each connection in a neural network having a weight associated with it. The nodes make up the input layer, one or more hidden layers in between and the output layer.

Based on the direction of flow of information. Neural network models can be classified as either feed forward or recurrent. In a feed forward network the information flows from the input layer, through the hidden layers to the output layer. In other hand, recurrent neural networks have a bidirectional flow of information, where the information initially propagates in a feed forward manner and then back from the later processing stages to the earlier stages.

Neural networks have been used in stock market prediction

And other financial predictive model across the world this is a clear implication of a good performance in modeling monthly revenue returns from mobile payment services with

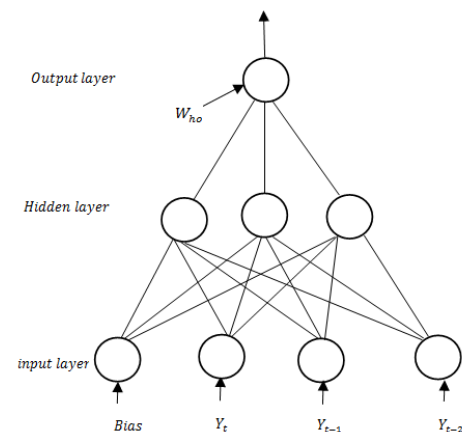The architecture of the neural network is shown I Fig 1 below.

Fig 1. *Multi-layered feed forward neural network*

Fig 1 above show a three-layer multi layered feed forward network with d input nodes, H hidden nodes, and one output node, and activation function φ (x) . The input at hidden layer nodes are connected by weights W_{hj} h ϵ (1,…, H) and jϵ (0,…,d) and W_{h0} is the bias for the ith hidden node. The hidden and output layers are connected by weights α_h for h ϵ (1,…,H)

The hidden and output layers are connected by weights α_h for h ϵ (1,…,H)

The input $V_h(x)$ to the hth hidden node is the value

$$V_h(x) = W_{h0} + \sum_{j=1}^{4} W_{hj}x_j \qquad (1)$$

The output $\emptyset_h(x)$ of the hth hidden node is the value

$$\emptyset_h(x) = \varphi(V_h(x)) \qquad (2)$$

The net input to the output node is the value

$$Z(x) = \alpha_0 + \sum_{1}^{4} \alpha_0 \emptyset_n(x) \qquad (3)$$

The output of network is

$$Z(x) = \psi\big(Z(x)\big) \qquad (4)$$

The weights are adjusted through supervised training, for the feed-forward ANN, a sigmoid function is selected as the activation functions to generate relationship between input and output nodes.

$$\psi(x) = \frac{1}{1+exp(-a(x-b))} \quad (5)$$

Where 'a' is the learning rate and 'b' is the bias.

Feed forward neural network is designed to learn a nonlinear mapping between the input and output data. Generalization is used to predict the possible outcome for a particular task. This process is split into two phases known as the training phase and the testing phase.

The functional relationship estimated by the ANN for an extrapolative or time series problem, the inputs are typically the past observations of the series and the output is the future value. The function mapping performed by the ANN is as shown:

$$Y_{t+1} = F(y_t, y_{t-1}, y_{t-2}, \dots y_{t-n}) \quad (6)$$

Where y_t is the observation at time t. Thus the ANN is the same as the nonlinear autoregressive model along the series.

3.1. Multilayer Perceptron

MLP network is also the type of ANN most commonly used in time series forecasting. The univariate time series modeling is normally carried out through this ANN by using a determined number of the time series lagged terms as inputs and forecasts as outputs. The number of input nodes determines the number of prior time points to be used in each forecast, while the number of output nodes determines the forecast.

3.2. Stopping Rule

Sum of Squared Errors (SSE) is used as a stopping rule to determine when training should stop. In this methods the weights are adjusted in such a way that the SSE between the targets y and the goal of output Z is minimized. The SSE is defined as:

$$Error, E = \frac{1}{n} \sum_{k=0}^{n}(t_k - o_k))^2 \quad (7)$$

Where t_k is the target output and o_k is the actual output of the neuron. This error is used to decide when to stop the ANN training process. To examine how close of the forecast to the actual value. The activation function is defined as:-

$$S^2(x_i;\theta) = \sum(Y_i - Z(x_i;\theta))^2 \quad (8)$$

$$= \sum(Y_i - Z(x_i; W, \alpha))^2 \quad (9)$$

$$= \sum(Y_i - \psi(\alpha_0 + \sum \alpha_n \psi(W_{h0} + W_{hj}x_j))^2 \quad (10)$$

4. Methodology

In this paper we analyze the Monthly revenue returns from mobile payment services between March 2007 and June 2013. The data sets are obtained from the Central Bank of Kenya website

4.1. Selection of Input Variables

The distinct feature of input variables of neural networks is that the input variables should not be much correlated, since the correlated input variables may worsen the prediction performance by interacting with each other and generating a biased effect. In order to let the input variables correlated to the output variable be used in the model, we should choose the input variable with less lags such as $y_{t-1}, y_{t-2} y_{t-3} \dots, y_{t-n}$ where, y_t *is* the value at time t and since Neural network is a nonparametric method choosing the number of input variables is very vital since with less of lags between each input variable may result to over fitting whereas increase in the lag between input variable may result to under-fitting. To handle this issue of under fitting or over fitting, we select few number of lags and test the performance of the neural network on a trial and error basis to get the optimum lag combination by checking goodness of fit for each model.

In neural networks applications the input data is usually normalized into the range of [0, 1] according to the activation function of the neurons, and then the neural networks trained and tested using the Quasi newton algorithm.

4.2. Number of Hidden Layers

The hidden layer(s) provide the network with its ability to generalize. In practice as suggested by (Baum and Haussler, 1989), neural networks with one and occasionally two hidden layers are widely used and have performed very well. Increasing the number of hidden layers also increases computation time and the danger of over fitting which leads to poor out-of-sample forecasting performance. In the case of neural networks, the number of weights, which is adamantly interrelated to the number of hidden layers and neurons, and the size of the training set, determine the likelihood of over fitting. It was applied neural networks structure with one and two hidden layers to a comparison in our study we used one hidden layer

4.3. Training and Testing of Data

The logarithmic rates were divided into training set from March 2007 to October 2011 and testing sets from November 2011 to June 2013. The training set was used to optimize the weights and the bias of the network, while testing set was used to determine the generalization ability of the network. Once the network topology and activation functions have been specified, the network needs to make adjustments to the synaptic weights to learn the function, the transformed data was scaled using a linear function to an interval of 0 and 1. The quasi-newton method known as BFGS (Broyden, Fletcher, Goldfarb and Shanno). Is used in training data to train our model and validated the model with testing data. The data was spit such that 70% of the data was used as training set while 30% was used for validation and testing of the model to check on goodness of fit.

4.4. Model Selection Criterion

The Akaike Information criterion(AIC) and Bayesian Information criterion(BIC) are more commonly and sufficiently used criterion for choosing the best model that fit the data, adding more parameters to the model increases the likelihood and by this the model is over fitted thus, BIC and AIC resolve this problem by introducing a penalty term for the number of parameters in the model, the penalty term is larger in BIC than in AIC, thus we selected the best model by observing the smallest AIC value from all the validated model.

4.5. Performance Measure

To measure how well a neural network performs, an objective function must be defined. This is demarcated as forecasting error which is the difference between actual and predicted values. A usual way used to evaluate predictive power of the model includes:-

i.   Mean Squared Error(MSE)

$$\frac{\sum(y_t - \widehat{y_t})^2}{N} \tag{11}$$

ii.   Mean Absolute Error(MAE)

$$\frac{\sum|y_t - \widehat{y_t}|}{N} \tag{12}$$

iii. Mean Absolute Percentage Error(MAPE)

$$\sqrt{\frac{\sum(y_t - \widehat{y_t})^2}{N}} \tag{13}$$

iv.   Root Mean Squared Error(RMSE)

$$\frac{1}{N}\left\{\frac{\sum|y_t - \widehat{y_t}|}{y_t}\right\}100 \tag{14}$$

The prediction errors are calculated using all the above performance measures

5. Results

The ANN model (developed based on the training data) with 1 hidden node was found to show the least error, when compared with the testing data, thereby resulting in maximum capture of the actual trend observed with respect to monthly revenue collection as shown in Fig2.

Data set between March 2007 and October 2011 was used for training purpose and then predict the revenue from November 2011 to June 2013.

Error is calculated using the following formula

$$RelaiveAbsoluteError = \frac{|Actual - Predicted|}{Actual}x100$$

Table 2 shows the percentage relative error in prediction for the specified period. From table 1 the model consisting of lag 1 and lag 2 respectively was identified as the optimal model using AIC selection criterion with AIC(-140.8966), BIC(-134.2051) and one hidden node. Therefore the final

model included two input nodes, one hidden node and one output node. To validate the ANN architecture the in sample forecasting was used where Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error(RMSE) and Mean Absolute Percentage Error (MAPE) test where used for forecasting accuracy measures. Using this specified technique to test for the performance measure of the fitted model, table 3 represent the output of the various selected techniques the results confirms that model with lag1 and 2 performed better than any other model in predicting monthly revenue.

The Relative Mean squared Error (RMSE) of neural networks along with Mean Absolute Error (MAE) is very much low. This consistency proves their accurate prediction power and absolutely true for the mobile payment revenue monthly data.

***Table 1.** ANN model AIC, BIC values*

LAGS	AIC	BIC
1,2,3,4,5	-135.7117	-120.733
1,2,3,4	-136.3776	-124.3947
1,2,3	-138.4102	-129.4230
1,3,4	-137.3849	-128.3977
1,2,3,5	-136.8190	-124.8361
1,2,4	-138.2407	-129.2535
1,2*	-140.8966	-134.2051
4,5	-125.6273	-119.6359
2,5	-134.0179	-128.0264
2,3,5	-132.3777	-123.3905
1,5	-140.2525	-134.2611

*indicates the 'best' ANN model prediction

***Table 2.** Relative Error values*

	Date	Actual (Billions)	Predicted (Billions)	Error
1	November 2011	112.332	111.1390	1.062009
2	December 2011	118.080	115.4632	2.216123
3	January 2012	114.060	114.5957	-0.469643
4	February 2012	116.691	115.1108	1.354199
5	March 2012	126.093	121.2801	3.737643
6	April 2012	117.360	118.6759	-1.121242
7	May 2012	128.403	122.9049	4.281906
8	June 2012	124.020	123.2972	0.582802
9	July 2012	129.280	125.2487	3.118247
10	August 2012	131.380	127.8802	2.663890
11	September 2012	130.690	128.0265	2.038040
12	October 2012	137.680	131.9020	4.196717
13	November 2012	138.990	134.5020	3.229019
14	December 2012	150.160	141.1923	5.972108
15	January 2013	142.653	139.7824	2.012276
16	February 2013	141.126	137.0157	2.912507
17	March 2013	134.446	132.7265	1.278932
18	April 2013	142.609	135.7218	4.829450
19	May 2013	158.770	146.9280	7.458606
20	June 2013	152.500	147.5791	3.226846

Table 3.below summarizes the performance measure all the ANN model developed

***Table 3.** performance measure for the selected models*

Lags	MAE	MSE	RMSE	MAPE
1,2,3,4,5	7.0202	77.602	8.8092	5.1459
1,2,3,4	6.7848	76.8481	8.7626	4.9354
1,2,3	3.7673	21.8320	4.6725	2.7171
1,3,4	6.4166	70.3445	8.3872	4.7011
1,2,3,5	6.7199	73.3335	8.5635	4.9209
1,2,4	6.7835	76.4936	8.7460	4.9496
1,2*	3.9977	23.7633	4.8748	2.8881
4,5	6.2191	65.1414	8.0710	4.4573
2,5	7.85672	95.1691	9.7555	5.6828
2,3,5	5.2717	49.5858	7.0417	3.8516
1,5	9.2252	118.7600	10.8977	6.7356

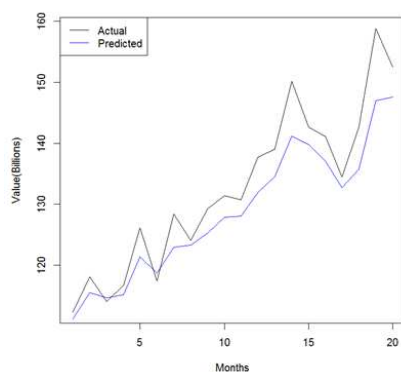*indicates the best ANN model for out-of-sample prediction

***Fig 2.** Actual versus predicted values*

Fig 2. Shows a plot of actual values for the monthly revenue between November 2011 to June 2013 in black and predicted values in blue using the trained neural network model. As observed from the plot the accuracy of the model may not be perfect as per the data set since mobile payment is a new entity in financial market and we still lack sufficient data to facilitate sufficient training but with time the model may improve with the growth of the dataset.

6. Conclusion

The studies reveal a high potential of ANN in predicting the monthly revenue collection from mobile payment services in Kenya and thus a good measure by government to implement a tool for giving accurate prediction of revenue collection for future planning

Neural network application in predicting revenue returns on mobile payment services creates a research gap for further investigation whether the model can be extended to Random forest which incorporates machine learning algorithm measures like boosting and bagging of decision trees have added predictive capability over these approach and I addition the tree base algorithm lacks sensitivity to noise and not subject to over fitting which are the challenges

experienced in mobile payment dataset and thus using RF will guarantee accuracy in forecasting. The policy implication of this study is that ANN can be used to model revenue returns from mobile payments services, which is certainly useful for various financial players such as government and policy makers of the country. This will catalyze the desired objective of economic growth in the country economy significantly.

References

[1]   Zhang, G., Patuwo, B. E. and Hu, M. Y. (1997) El-Shazly, M. R. and El-Shazly, H. E. (1997), 'Comparing the Forecasting Performance of Neural Networks and Forward Exchange Rates', Journal of Multinational Financial Management, 7, 345-356.

[2]   Mwita, P., Franke, J., Odhiambo, R. and Waititu, A. (2005). On conditional quantiles: Direct Kernel Estimator and its Consistency. African Journal of Science and Technology, Vol. 6(2), 67-76.

[3]   Tang , Almeida and Fishwick, Simulation, Time series forecasting using neural networks vs. Box-Jenkins methodology ,November 1991, pp303- 310.

[4]   Medeiros M, Terasvirta, T, Rech, G. (2006) "Building Neural Network Models for Time Series: A Statistical Approach." Journal of Forecasting. 25(1) pp. 49-75.

[5]   Zhang, G., Patuwo, B.E., Hu, M.Y. (1998), Forecasting with artificial neural networks: The state of the art International journal of forecasting, 14:35 -62.

[6]   Mahdavi Gh, Behmanesh MR (2005). The Forecasting of Stock Price of Investment Firms by Using Artificial Neural Networks. J. Econom. Res. 19(4): 211-233.

[7]   Zhong Luo, Liu Li-sheng. The application of Neural Network in Lifetime Prediction of Concrete. Journal of Wuhan University of Technology. 2002,17:79-81.

[8]   Teräsvirta T, Lin, C-FJ. 1993. Determining the number of hidden units in a single hidden-layer neural network model. Research Report 1993/7, Bank of Norway.

[9]   J. Yao, Y. Li and C. L. Tan, "Option price forecasting using neural networks," OMEGA: Int. Journal of Management Science, vol. 28, pp 455-466, 2000.

[10] HILL, T., OCONNOR, M. & REMUS, W. (1996) neural network models for time series forecasts. Management Science, 42, 1082-1092.

[11] Pacelli1, V., Bevilaqua, V., Azzollini, M. (2011), An Artificial Neural Network Model to Forecast Exchanges rates, Journal for Intelligent Learning Systems and Applications, 3:57 - 69.

[12] Kuan, C.M., Liu, T. (1995),   forecasting exchange rates using feedforward and recurrent neural networks, Journal of Applied Econometrics, 10, (4): 347 – 364.