Application of support vector machines to the forecast of direction of exchange rate in Nigeria

[1]Alabi Nurudeen Olawale, [2]Are Stephen Olusegun
[1]Department of Mathematics and Statistics, Federal Polytechnic Ilaro
[2]Department of Mathematics and Statistics, Federal Polytechnic Ilaro
[1]e-mail: nurudeen.alabi@federalpoly.edu.ng, stephen.are@federalpoly.edu.ng

Abstract

In this paper, we have considered mainly four of the most common computationally efficient linear and non-linear classification kernels (i.e. support vector classifier or SVC, polynomial basis kernel, radial basis kernel and sigmoid kernel) called the support vector machines (SVMs) to model the direction of exchange rate using crude oil price and inflation rate in Nigeria between 2004 and 2017. This study aims at testing the SVMs accuracies and forecast strength on the direction of exchange rate in Nigeria data. We based our study on 167 observations collected from the database of the Central Bank of Nigeria (from 2004 to 2017) which were divided into two parts; training dataset (2004 to 2007) and test dataset (2008:2017). One-against-one approach and 10-fold cross-validation resampling technique were used to select amongst several cost parameters C and γ values for the linear and non-linear basis kernels. Specifically, cross-validation selected $C = 0.5, 10^5, 0.1, 10^3$ for the SVC, polynomial basis kernel of degree = 2, sigmoid kernel and radial basis kernel respectively. For the radial basis kernel, $\gamma = 0.001$ was optimal amongst 153 positive γ values implemented in the cross-validation procedure. A total of 155, 148, 158 and 150 support vectors were generated for the SVC, polynomial basis kernel of degree = 2, sigmoid kernel and radial basis kernel respectively. All kernels except the radial basis kernel performed slightly better on the test dataset. However, the confusion matrices showed that only the polynomial basis kernel of degree = 2 classified one of the three observations in the "unchanged" class correctly on the test data set. Packages from various ***R*** *libraries were deployed throughout the paper.*
Keywords*: Support vector machine, direction of exchange rate, crude oil price, inflation rate, one-against-one approach, 10-fold cross-validation.*

1.0 Introduction

Forecasting the direction of exchange rate has been a subject of interest in both developed and developing nations. Many researchers have done a great deal of research on forecasting exchange rate using different approaches and statistical analysis models. For instance, [1] established an ARDL (4,4,0) model using bounds testing procedure. Their empirical analysis showed that these macroeconomic variables have highly significant level relationship with exchange rate irrespective of the underlying properties of the time series. They emphasized that the conditional level relationship model and the associated conditional unrestricted equilibrium correction model (ECM) in the long- and short-run relate crude oil prices negatively and inflation rate positively with exchange rate. [2] worked on Exchange Rate Forecasting with Neural Networks. The paper presents the prediction of foreign exchange rate using artificial neural networks. Since neural networks can generalize from past experience, they represent a significant advancement over traditional trading system, which require a knowledge expert to define trading rules to represent market dynamics. It is practically impossible to expect that one expert can devise trading rules that account for, and accurately reflect, volatile and rapidly changing market conditions. With several neural networks, a trader may use the predictive information alone or with other available tools to fit the trading style, risk propensity, and capitalization. Numerous factors affect the foreign exchange market, as they were described in the paper. The neural network helped minimize these factors by simply giving an estimated exchange rate for the future day (given its previous knowledge gained from extensive training). Because the field of financial forecasting is too large, the scope in this paper is narrowed to the foreign exchange market, specifically the value of the Japanese Yen against the United States Dollar, two of the most important currencies in the foreign exchange market.

[3] research on the Performance of Deterministic and Stochastic Trends Model in Forecasting the Behavior of the Canadian dollar and the Japanese Yen against the US Dollar, showed that complex

forecasting exchange rate models do not outperform ARIMA models, the same forecasting models applied to forecast the behavior of the Canadian dollar and the Japanese Yen against the US dollar produced varying forecast performance. [4] research on A New Approach to Forecasting Exchange Rates, built on purchasing power parity theory; this paper proposed a new way approach to forecasting rates using his Big Mac data from The Economist magazine. Their approach is attractive in three aspects. Firstly, it uses easily-available Big-Mac data as input. These prices avoid several serious problems associated with broad price indexes, such as Consumer Price Index (CPI), that are used in conventional PPP studies. Secondly, this approach provides real-time exchange rate forecasts at any forecast horizon. Such real-time forecast can be made on a day-to-day basis if required, so that the forecasts are based on the most up-to-date information set. These high-frequency forecasts could be particularly appealing to decision makers who want up-to-date forecasts of exchange rates. Finally, their forecasts were obtained through Monte Carlo simulation; estimation uncertainty was made explicit in their framework which provided the entire distribution of exchange rates, not just a single point estimate. A comparison of their forecasts with the random walk model shows that although the random walk is superior for very short horizons, their approach tends to dominate over the medium to longer term.

[5] research on Exchange Rate Predictability tested various models such as Single Equation Linear Models, Error Correction Model (ECM), Non-Linear Model, Time-Varying Parameter (TVP) Model, Multivariate Models and Panel Models. Her researched revealed that predictability of exchange rate is dependent on the choice of predictor, forecast horizon, sample period, model and forecast evaluation method. It also showed that predictability is more apparent when one or more of the following hold: the predictors are Taylor rule or net foreign assets, the model is linear, and a small number of parameters are estimated. [6] used Markov-switching vector error correction model (MSVECM) to test the Trade-weighted nominal and real effective exchange with three different country indices (broad, main, OITP), WTI crude oil price (in USD/barrel), US CPI, three-month treasury bill rate in the US from January 1974 – November 2011 (monthly data). The result showed that Effective depreciation of the dollar triggers an increase in oil Prices (in nominal terms). Increase in real oil prices is associated with a real appreciation of the dollar (stems from price effects). [7] used Markov-switching vector error correction model (MSVECM) WTI nominal oil price expressed in USD, CPI and exchange rates of 12 oil exporting and importing countries against the US dollar In oil exporting countries: Brazil, Canada, Mexico, Norway, Russia; And Oil importing: Euro Area, India, Japan, South Africa, South Korea, Sweden, and the UK from 1974 – 2011 (monthly data) showed that most important causality runs from exchange rates to oil prices, with a depreciation of the dollar triggering an increase in oil prices. Nonlinearities are an important issue when analyzing oil prices [8] used Time Domain Model to forecast the exchange rate between Nigerian naira and US dollar using Box Jenkins fundamental approach for the period, January 1994 to December 2011. Their sample forecast for period of 12-month term revealed that the naira will continue to depreciate on the US dollar for the period forecasted.

[9] working paper series of the European Central Bank on Exchange Rate Forecasting with

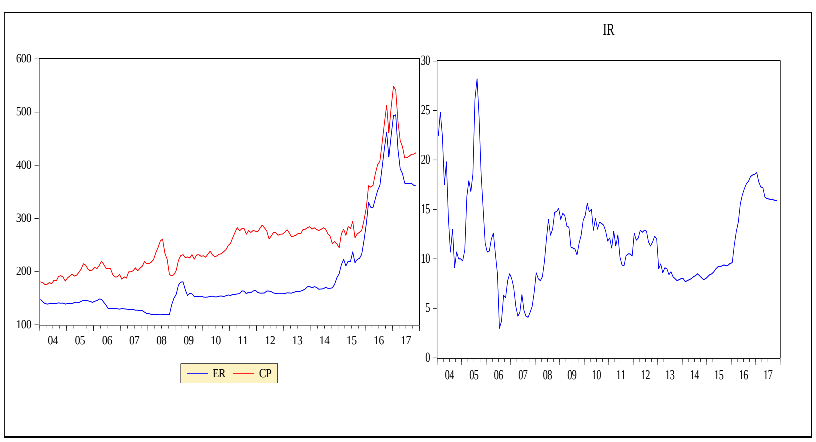


Figure 1. Time Series Plots on Exchange rate, crude oil price and inflation rate in Nigeria (2004 to 2017)

Dynamic Stochastic General Equilibrium (DSGE) models, ran a real exchange rate forecasting “horse race”, which highlights that two principles hold. First, forecast should not replicate the high volatility of exchange rates observed in sample. Second, models should exploit the mean reversion of the real exchange rate over long horizons. Abiding by these principles, an open-economy DSGE model performs well in real exchange rate forecasting. However, it fails to forecast nominal exchange rates better than the random walk. They found that the root cause is its inability to predict domestic and foreign inflation. This short coming leads us towards simpler ways to outperform the random walk. [10] used Autoregressive Integrated Moving Average (ARIMA) model to analyze exchange rate of Nigeria naira to the US dollar. His research work covers the period (1982 - 2011), through Box-Jenkin methodology on AR (1): order one generated model was preferred as it was proved through the diagnostic rate of Naira – Dollars based om its potentials for better prediction and computational requirement. [11] used Autoregressive Distributed Lag (ARDL) model bound testing approach, Wavelet Coherence (WTC), wavelength based signal detection and frequency domain approach to determine the relationship between oil prices, real exchange rate, GDP, government spending, terms of trade, productivity differential in Russia from 1993Q1 – 2009Q4 (Quarterly data). Their result showed that Oil price causes sharp real EXR in lower frequencies. The link between oil price and real exchange rate seems conditioning upon GDP, government expenditures, terms of trade and productivity differential. [12] used Vector Error Correction Model (VECM), Impulse Response Function (IRF) and other models to determine the impact of interest rate, inflation and money supply in exchange rate volatility in Pakistan. Their results showed that a short run as well as long run relationship exist between inflation and exchange rate volatility. They concluded that high money supply and increase in interest rate, raises inflation rate which leads to exchange rate volatility. [13] working paper series of the European Central Bank on Exchange Rate Prediction Redux: new models, new data, new currencies, looked at previous assessment of nominal exchange rate determination. [14] augmented the usual suspected with productivity based models and “behavioral equilibrium exchange rate” models and assessed performance of horizon of up to 5 years. In their paper, they further expand the set of models to include Taylor rule fundamentals, yield curve factors, and incorporate shadow rates and risk and liquidity factors. The performance of these models was compared against the random walk benchmark. The models were estimated in Error Correction and first difference specification. They examined model performance at various forecast horizons (1 quarter, 4 quarters, 20 quarters) using different metrics (Mean Square Error, direction of change), as well as the “consistency” test of [15]. No model consistently outperforms a random walk, by a mean square error measure, although purchasing power parity does fairly well. Moreover, along a direction-of-change dimension, certain structural models do outperform a random walk with statistical significance, while one finds that these forecasts are co-integrated with the actual values of exchange rate, in most cases, the elasticity of the forecasts with respect to the actual values is different from unity. They concluded overall, that model/specification/currency combination that work well in one period will not necessarily work well in another. [16] research on the Prediction of Chaotic Exchange Rate Time Series Applying Dynamic Component Predicting Model. In order to forecast chaotic variable of exchange rate, the paper integrated RBF neural network model, Lyapunov exponent model into a dynamic component model, weights of which could be adjusted by series themselves. Empirical research on 5 exchange rates showed both error indexes and direction statistics of component model could obtain better result than individuals, especially for JPY/USD and SEK/USD. Moreover, a compare on performance was taken between component model and random walk model. Both D-M and H-M test refused null hypnosis, showed the component model could obtain obvious advantages than RW model as expected.

In light of the above, this study seeks to predict and forecast the direction of exchange rate from crude oil price and inflation rate in Nigeria using Support Vector Machine (SVM) algorithm. Consequently, in our subsequent paper, we intend to answer the question of: Can SVM offer a more accurate forecast at minimal cost to traditional or classical classification methods?

2.0 Empirical Data Analysis and Results

In order to study the relationships between the Direction of Exchange Rate and two very important macroeconomic factors such as Crude Oil Price and Inflation Rate influencing its movement in Nigeria, an empirical application of a simple and “*out of the box*” machine learning classification technique such as Support Vector Machine (SVM) is proposed. As earlier mentioned, support vector machines are computationally efficient and have been applied in various fields such as medicine, bioinformatics and astrophysics amongst several other fields. But its application to economic/financial data is an aspect that is not yet fully exploited mainly because of the data generating process underlying macroeconomic data. Most economic and financial data are in form of time series which require more sophisticated econometric techniques for analysis. Many researchers prove that fundamental models do not provide accurate exchange rate forecast. Hence, it is the goal of this current work to test the accuracy and forecast strength of SVMs on the direction of exchange rate data in Nigeria between 2004 and 2017.

2.1 *The Dataset*

This current work is based on 167 observations (from 2004 to 2017) on direction of exchange rate (based on the market rate), Crude Oil Price and Inflation Rate which were divided into two parts; training dataset (2004 to 2007; $n_1 = 48$) and test dataset (2008:2017; $n_2 = 119$). The reason is to ensure that the performances of the SVMs proposed are consistent in both datasets. Also, this will ensure that three observations in the "*unchanged"* class are captured in testing the performance of the SVMs. We postulated linear and non-linear models between direction of exchange rate (***er***) as the response variable, crude oil price (***cp***) and inflation rate (***ir***) as the two feature variables. Direction of Exchange Rate is a categorical variable with $k = 3$ classes (such as *"Up", "Unchanged" and "Down"),* Crude Oil Price and Inflation Rate are continuous variables. This data was analyzed based on the *one-against-one* classification using ${}^{k}\mathbf{C}_2 = 3$ Support Vector Machines (*SVMs*). The three all-pair SVMs are "*up-against-down*", "*up-against-unchanged*" and "*down-against-unchanged*". Support Vector Machine is an evolving machine learning technique developed in the mid-1990s to serve as alternative to classical methods of classification by using separating hyperplanes. They are based on structural risk minimization which perform better than the empirical risk minimization used in conventional neural networks. A hyperplane of a *p*-dimensional space is a flat subspace of *p*-1 dimension which need not pass through the zero origin on the *y*-axis. This method is an extension of Support Vector Classifier and a generalization of the Maximal Margin Linear Discriminant Classifier (a natural choice of classification resulting in a perfect separating hyperplane) which are classifiers requiring the classes be separable by a linear boundary. However, most datasets are separable by non-linear class boundaries as shown in this research work. Since the number of classes *k* of the Direction of Exchange Rate is three, this method is particularly desirable. We employed the *e1071* package [17] which was complemented by visualization and tuning functions in R-*Studio* throughout this research work. This package comprises the linear, polynomial, radial basis function (RBF) and sigmoid kernels which are all considered in this paper. In addition to these major features, it contains sequences for the calculation of probability of prediction in terms of training and test error rates using *q*=10-fold cross-validation, fixed sampling cross validation and bootstrapping.

2.2 *Linear decision boundary classifier (SVC)*

The *Support Vector Classifier* (an extension of maximal margin classifier) is based on the fact that we relax the condition that each observation falls on the correct side of the margin or hyperplane by allowing some observations to fall on the wrong side of the margin or the separating hyperplane instead. This is particularly necessary for robustness and a more improved classification of a larger proportion of the training observations. For our model on the Direction of Exchange Rate, we begin with a 167 by 2 data matrix **X** that consists of n_1 training observations in *p*-dimensional space. Here $n_1 = 48$ observations and $p = 2$ feature or output variables in x_1 = crude oil prices (***cp***) and x_2 = inflation rates (***ir***). So, the data matrix **X** is given as

$$X = \begin{bmatrix} x_{11} & x_{12} \\ . & . \\ . & . \\ . & . \\ . & . \\ x_{1671} & x_{1672} \end{bmatrix}$$

and each observation falls into one of two classes of the Direction of Exchange Rate (***er***) using the three *one-against-one* classification classes. A separating hyperplane defined by the perpendicular distance from the *i*th observation to the hyperplane is expressed as

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) > 0 \text{ for all } i = 1, 2, 3, \dots, n_1 \qquad (1)$$

A training observation x_i is classified depending on the sign associated with

$$f(x) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \qquad (2)$$

if f(x) is positive, we assign it to one class and if it is negative we assign it to the other class. A Support Vector Classifier is the solution to the convex optimization problem

$$\max_{\beta_0, \beta_1, \beta_2, \epsilon_1 \ldots\ldots \epsilon_{n1}} imize \;\; M$$

subject to

$$y_i(\beta_0 + \sum_{j=1}^{2} \beta_{j1} x_{ij} + \sum_{j=1}^{2} \beta_{j2} x_{ij}{}^2) \geq M(1 - \in_i)$$

$$\sum_{i=1}^{n_1} \in_i \leq C, \in_i \geq 0, \sum_{j=1}^{2} \sum_{k=1}^{2} \beta^2{}_{jk} \qquad (3)$$

where C is a nonnegative tuning or "*cost*" parameter (also called the regularization term), M is the width of the margin, ϵ_i are slack variables which is the proportional amount by which individual observations fall on the wrong side of the hyperplane. Therefore, the sum of all the slacks bound the total proportional amount by which observations fall on the wrong side of the hyperplane. By placing a limit on the sum of the slacks ϵ_i, we bound the total proportional amount by which observations are misclassified. An observation is misclassified if $\in_i > 1$. Hence by bounding $\sum_{i=1}^{n_1} \in_i$, at a constant C, we limit the number of training observations to be misclassified at C. The solution to the convex optimization problem which is solved via quadratic programming using Lagrange multipliers measures the overlap in relative distance, which changes with the width of the margin M leading to a "*standard*" support vector classifier. In order to inform the Lagrange (primal) function for computational convenience, the convex optimization is expressed as

$$\min_{\beta_0, \beta_1, \beta_2} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^{n_1} \in_i \qquad (4)$$

subject to

$$\in_i \geq 0, y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \geq (1 - \in_i), \forall i$$

whose primal function was minimized with respect to β_0, β_1 and β_2 by setting the corresponding derivatives to zero. Subject to the following Karush-Kuhn-Tucker conditions

$$\alpha_i [y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) - (1 - \in_i)] = 0 \qquad (5)$$

$$\mu_i \epsilon_i = 0 \qquad (6)$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) - (1 - \in_i) \geq 0 \qquad (7)$$

The resulting Wolfe objective function was maximized subject to $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{n_1} \hat{\alpha}_i y_i = 0$. The estimates of the model parameters β_0, β_1 and β_2 is the vector

$$\hat{\beta} = \sum_{i=1}^{n_1} \hat{\alpha}_i y_i x_i \qquad (8)$$

The $\hat{\alpha}_i$'s are nonzero parameters only for the support vectors for which $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) - (1 - \in_i) \geq 0$. Some of these support vectors fall on the edge of the margin with slack value of zero while the rest of the support vectors have $\hat{\alpha}_i = c$. The intercept β_0 was obtained from the support vectors on the margin. An observation either in the training dataset or the test dataset was classified using the following decision rule:

$$\hat{\beta}(x) = sign[\hat{f}(x)] = sign[\beta_0 + \beta_1 x_1 + \beta_2 x_2] \qquad (9)$$

2.3 *Non-linear decision boundary classifiers (SVM)*

As earlier mentioned in this research work, the decision boundary between the three classes may be non-linear making the classification using a support vector classifier inadvisable. Support vector machine is an extension of the support vector classifier which involves increasing the number of output variables using kernels. Kernels are functions used to quantify the connection between two observations. There are different types of kernels but this research work focuses mainly on the polynomial basis kernels, sigmoid kernel and the radial basis kernels. These kernels are expressed below:

- Radial Basis Function (RBF) kernel

$$k(x, x') = \exp(-\gamma \sum (x_{ij} - x_{i'j})^2)) \qquad (10)$$

- Polynomial kernel

$$k(x, x') = (1 + \sum_{j=1}^{2} x_{ij} x_{i'j})^{degree} \qquad (11)$$

- Sigmoid kernel

$$k(x, x') = \tanh(offset + a \sum_{j=1}^{2} x_{ij} x_{i'j}) \qquad (12)$$

Support vector classifier is a polynomial kernel of degree one (*degree* = 1) which can be expressed in terms of the inner product of the $n_1 = 48$ training observations as

$$f(x) = \beta_0 + \sum_{i=1}^{n_1} \alpha_i K(x, x_i) \qquad (13)$$

There are n_1 parameters α_i, $i = 1, 2, \ldots n_1$ per training observation which were estimated using 10,296 inner products between all pairs of the training observations. If we let G be the collection of indices of all support vectors, the support vector classifier above can be expressed as

$$f(x) = \beta_0 + \sum_{i \in G} \alpha_i K(x, x_i) \qquad (14)$$

The mixture of a linear classifier such as the support vector classifier expressed in equation (14) and a non-linear kernel such as the RBF or polynomial basis kernel of higher degree results in the support vector machine (SVM). The empirical analysis of the data produced the following descriptive statistics and subsequent SVM results on direction of exchange rate, crude oil price and inflation rate in Nigeria during the review period.

Table 1: *Descriptives on direction of exchange rate, crude oil price and inflation rate*

	cp ($/barrel)	er (₦/$US)	ir (per cent)	Direction of Exchange Rate
Minimum	30.66	118.7	3	"*Down*" = 81
First Quarter	51.87	140.9	8.5	"*Unchanged*" = 3
Median	71.8	158.3	11.3	"*Up*" = 83
Mean	76.83	187	11.82	

Third Quarter	106.31	171.4	14.25	
Maximum	141.86	494.7	28.2	

Source: : *Authors' Computations using R-Studio*

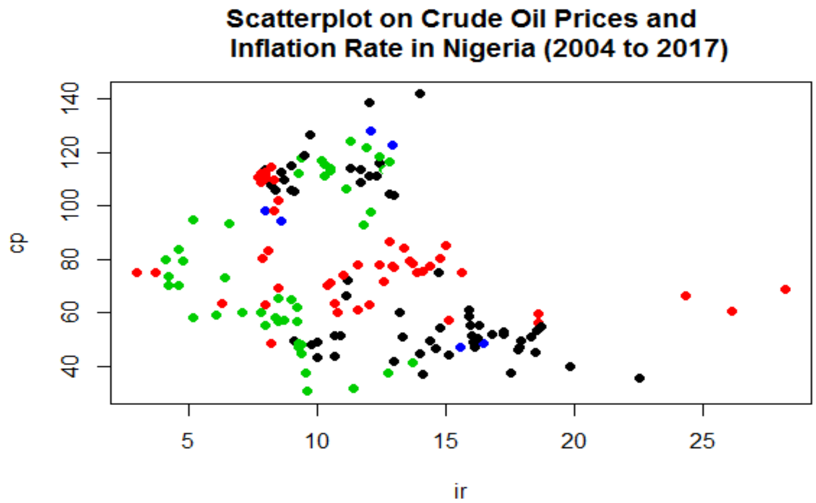


Figure 2: *Scatterplot on crude oil price (**cp**) and inflation rate (**ir**) which will be used for the linear and nonlinear classifications. This plot exhibits pattern of non-linearity between **cp** and **ir** in the feature space.* Source: *Authors' Computations using R-Studio*

Firstly, we fit the support vector classifier (which is equivalently an SVM using a polynomial kernel of degree = 1), then a polynomial basis kernel of degree = 2, to the training dataset (n_1=48 training observations). By using the *one-against-one* procedure when the number of classes *k* is greater than 2, SVC classifier calculated fitted values which are numerical scores of the form

$$\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1(cp) + \hat{\beta}_2(ir)$$

for each training observation. For any given cutoff *t*, we classify observations into the "*down*", "*unchanged*" or "*up*" categories depending on whether $\hat{f}(X) < t$ or $\hat{f}(X) \geq t$. We set the cutoff value at zero, so if the fitted value is less than zero, the observation is assigned to one class and if it greater than zero, it is assigned to the other class. Generally, for all non-linear kernel SVMs including polynomial basis kernel of degree = 2 and higher, radial basis kernel and sigmoid kernel, the fitted values are obtained using equation 13 and the classification of an observation is done in a similar manner with the SVC. The 10-fold cross-validation selected the best support vector classifier and polynomial basis kernel of degree = 2 with cost C = 0.5 and 10^5 respectively and training error rates of 41.26 per cent and 40.72 per cent respectively. **Table** 2 below is the detailed performance result of cost = 0.1, 0.5, 1, 10, 10^2, 10^3, 10^4, 10^5,10^6. One-against-one procedure generated 3(3-2)/2 = 3 SVMs (i.e., "*up-against-down*", "*up-against-unchanged*" and "*down-against-unchanged*" SVMs) by coding one class +1 and the other -1.

Table 2: *Performance report on the 10-fold cross validation for the Support Vector Classifier (SVC) and Polynomial basis with degree = 2, Sigmoid and Radial Basis Kernels*

Cost	Training error rate			Dispersion		
	SVC	*Degree = 2*	*Sigmoid*	*SVC*	*Degree = 2*	*Sigmoid*
0.1	0.4198	0.418	0.4379	0.0206	0.0201	0.1391
0.5	0.4126	0.4233	0.5562	0.0182	0.0144	0.0754
1	0.4144	0.421	0.5375	0.0197	0.0169	0.0929
10	0.415	0.4204	0.5371	0.0179	0.0185	0.1099
100	0.4144	0.4204	0.5371	0.0174	0.0174	0.1099
1000	0.4144	0.4168	0.5489	0.0174	0.0175	0.1152
10000	0.4138	0.4174	0.5489	0.0182	0.0167	0.1152
100000	0.4168	0.4072	0.5489	0.0186	0.0147	0.1152
1000000	0.4701	0.4138	0.5489	0.0649	0.017	0.1152

Source: : Authors' Computations using R-Studio

Note that all the support vectors are plotted as "***X***", other observation as circles, the true classes are highlighted via symbol color and the predicted class regions are visualized using colored background. Whilst the support vector classifier has a total of 155 support vectors in which 75 come from the "*up"* class, 77 from the "*down*" class and 3 support vectors come from the *"unchanged"* class, the polynomial kernel of degree = 2 has a total of 148 support vectors out of which 73 belong to "*up"* class, 72 to the "*down*" class and the remaining 3 support vectors belonging to *"unchanged"* class. We then classify a test observation using each of the three classifiers (i.e. *up-against-down*, *up-against-unchanged* and *down-against-unchanged*), and we tally the number of times that the test observation is assigned to each of the three classes. The final classification is performed by assigning the test observation to the class to which it was most frequently assigned in the three pairwise SVMs.

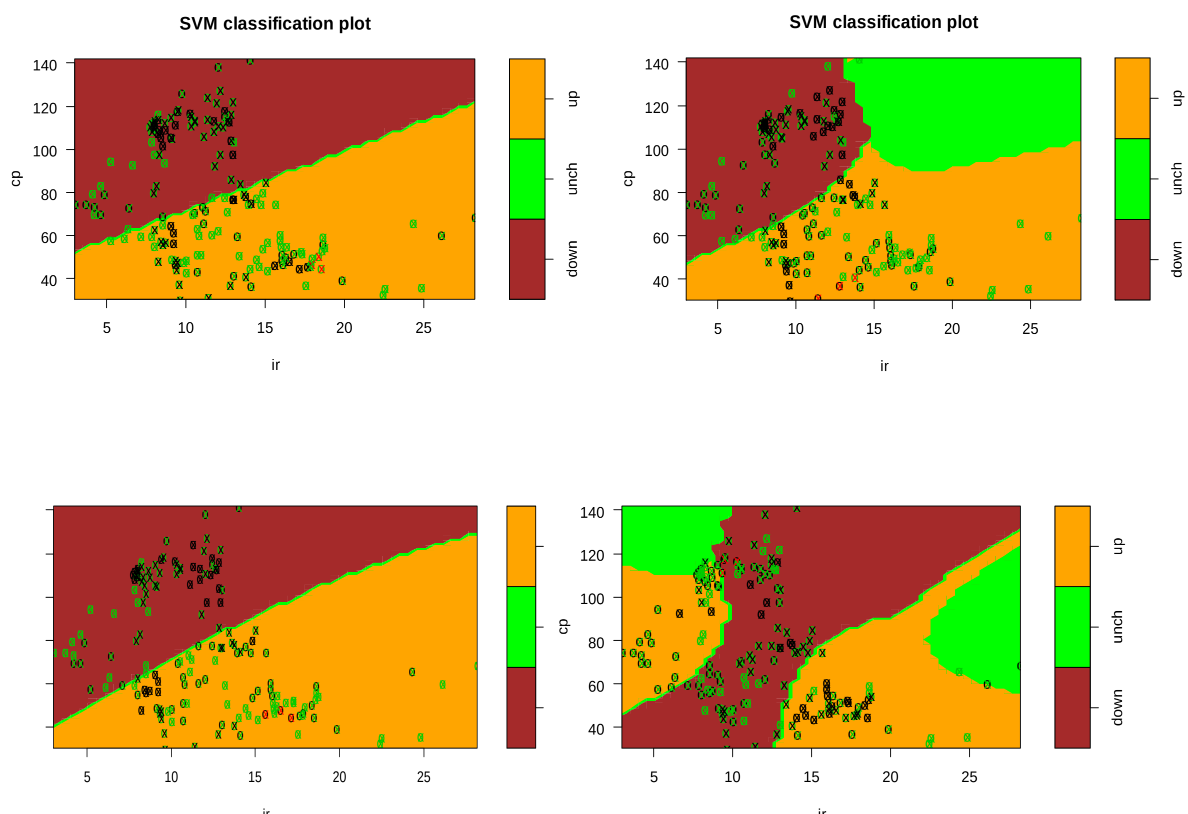


Figure 3: *A support vector classifier, the polynomial kernel of degree = 2, radial basis kernel and sigmoid kernels were fitted using nine different values of the tuning parameter C (i.e. C = 0.1, 0.5, 1, 10, 100, 1,000, 10,000, 100,000, 1,000,000) in equations (3) & (4)*. ***Upper left panel:*** *Support vector* *classifier with a polynomial kernel of degree = 1 classifying the non-linear data from Figure 2 into down, unchanged and up classes with an estimated training error rate of 0.4126 or 41.26 per cent. The decision boundary for the classification is linear but not maximal.* ***Upper right panel****: A polynomial kernel of degree = 2 classifying the same dataset into the three classes with an estimated training error rate of 0.4072 or 40.72 per cent. The decision boundary in the upper right panel seems to be better than on the upper left panel. A total of 48 observations from the test dataset were misclassified and 71 observations were correctly classified which accounts for 40.34 per cent test error rate for both the support vector classifier and the polynomial kernel of degree = 2 SVM (****Tables*** *4 & 5). 10-fold cross validation ensures that the risk of overfitting the data due to higher cost value is minimized at the same time keeping the training error rate as low as possible.* ***Lower left panel****: Support Vector Machine with radial basis kernel, $C = 10^3$ and $\gamma = 0.001$ classifying the non-linear data from Figure 2 into down, unchanged and up classes with an estimated training error rate of 41.32 per cent. The test error rate recorded was 42.02 per cent. The decision boundary for the classification is nearly linear.* ***Lower right panel*****:** *Support Vector Machine with sigmoid kernel with cost parameter $C = 0.1$ classifying the non-linear data from Figure 2 into down, unchanged and up classes with an estimated training and test error rates of 43.79 per cent and 40.34 respectively. Note that all the support vectors are plotted as crosses and all other observations as circles*. Source: : *Authors' Computations using* ***R****-Studio.*

Next, we employed another flexible non-linear classifier with radial basis kernel which takes the form in equation 10. Again, *one-against-one* approach and 10-fold cross validation resampling technique was used to select amongst 9 main cost parameters *C* and 18 positive γ values. A total of 153 training errors and dispersions were generated out of which we selected a cost $C = 10^3$ and $\gamma = 0.001$ with training error rate of 41.32 per cent. With this kernel, training observation which are far from the test observation plays no role in the classification of the test observation because $\exp(-\gamma\sum_{i=1}^{n_1}(x_{ij} - x_{i'j})^2)$ will be small making the radial basis kernel of equation 10 to have a *local* behavior. The radial basis kernel is as shown in lower left panel of **Figure** 3 generated a total of 150 support vectors in which 73 come from the "*up"* class, 74 from the "*down"* class and 3 support vectors come from the "*unchanged*" class. Low cost parameter means the radial basis kernel has no problem of overfitting as shown by the near linear nature of the decision boundary.

Finally, we applied the sigmoid kernel of equation 12 shown in the lower right panel of **Figure** 3, using the usual *one-against-one* and 10-fold cross-validation. This method yield a best performance model with cost parameter $C = 0.1$ and training error rate of 43.79 per cent (**Table** 2). A total of 158 support vectors were generated in which 77, 78 and 3 support vectors were associated with the "*up", "down"* and *"unchanged"* respectively. On the test dataset involving 119 observations, 48 observations were misclassified while 71 were correctly classified. This amount to about 40.34 test error rate for this SVM. In all the four SVMs, the three observations from the unchanged class are all support vectors.

2.4 *Choosing an optimal support vector machine for the direction of exchange rate in Nigeria*

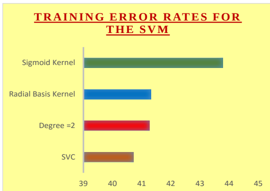
In general, our interest is not how well the SVMs perform on the training dataset. Rather, we are interested in the accuracy of the predictions that we obtain when we apply them to previously unseen test dataset**. Table** 3 shows the summary of the performances of the various SVMs in terms of the training and test error rates. This is essential because most often than not, a SVM may do well on the training dataset but perform poorly on the test dataset. However, how well a classifier perform is judged by the performance on the test dataset. In order to compare the performances of our four SVMs, we used the computed training and test error rates corresponding to each classifier. Recall that we used data between 2004 and 2007 to train the four SVMs and data between 2008 and 2017 to test them accordingly.

Table 3: Training error rate versus test error rate on the four SVMs

SVM	**Training error rate (**per cent**)**	**Test error rate (**per cent**)**
SVC	40.72	40.34

Degree =2	41.26	40.34
Radial Basis Kernel	41.32	42.02
Sigmoid Kernel	43.79	40.34

Source: : *Authors' computations using* ***R*** *Studio*



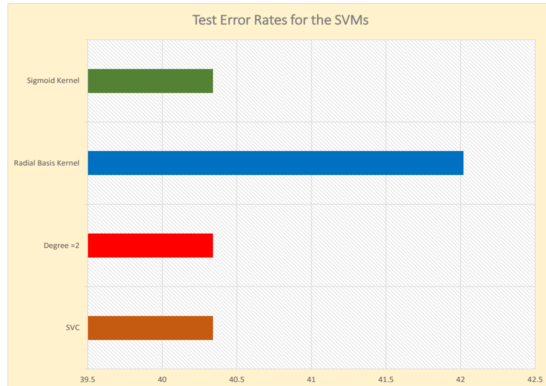


Figure 4: Bar Charts on the Training and Test error rates for the SVMs

The confusion matrices, shown in **Table** 4, **Table** 5, **Table** 6, **Table** 7, are conveniently displaying the number of correctly classified and misclassified observations in the test dataset for Support Vector Classifier, polynomial basis kernel of degree = 2, radial basis kernel and sigmoid kernel respectively. The **Table** 3 reveals that whilst SVC, degree = 2 polynomial basis kernel and sigmoid kernel performed better on the test dataset, radial basis kernel classifier had its performance reduced on the test dataset. Only the polynomial basis kernel of degree = 2 was able to classify one of the observations into "*unchanged*" class correctly. Other kernels classified the three observation in this class as "*down*".

Table 4: Confusion Table on Support Vector Classifier (cost = 0.5) using test dataset (2008 to 2017)

CLASSIFIED AS	**TRUTH**			
	"Down"	*"Unchanged"*	*"Up"*	**Total**
"Down"	32	3	25	60
"Unchanged"	0	0	0	0
"Up"	20	0	39	59
Total	52	3	64	**119**

Source: : *Authors' computations using* ***R*** *Studio*

Table 5: Confusion Table on Polynomial Kernel degree = 2 (cost = 10^5) using test dataset (2008 to 2017)

CLASSIFIED AS	**TRUTH**			
	"Down"	*"Unchanged"*	*"Up"*	**Total**
"Down"	30	2	24	*56*
"Unchanged"	0	1	0	*0*
"Up"	22	0	40	*62*
Total	*52*	*3*	*64*	**119**

Source: : *Authors' computations using* ***R*** *Studio*

Table 6: Confusion Table on Radial Basis Kernel, $\gamma = 0.001$, cost = 10^3, using test dataset (2008 to 2017)

CLASSIFIED AS	**TRUTH**			
	"Down"	*"Unchanged"*	*"Up"*	**Total**
"Down"	30	3	25	*58*

"Unchanged"	0	0	0	*0*
"Up"	22	0	39	*61*
Total	*52*	*3*	*64*	**119**

Source: : *Authors' computations using **R** Studio*

Table 7: Confusion Table on Sigmoid Kernel, (cost = 0.1) using test dataset (2008 to 2017)

CLASSIFIED AS	**TRUTH**			
	"Down"	*"Unchanged"*	*"Up"*	**Total**
"Down"	37	3	30	*70*
"Unchanged"	0	0	0	*0*
"Up"	15	0	34	*49*
Total	*52*	*3*	*64*	**119**

Source: : *Authors' computations using **R** Studio*

Further analysis reveals that by increasing the degree in equation (11) and employing the 10-fold cross-validation resampling procedure, the training error rates reduce but the SVMs perform poorly on the test dataset. Specifically, when the degree = 3, 4, and 5, the training error rates were 39.52 per cent, 39.38 per cent and 38.68 per cent with cost parameters C = 1, 1000, and 10,000 respectively. However, the test error rate for each of these higher-degree polynomial basis kernel SVMs was constant at 40.02 per cent. Hence polynomial basis kernel of degree = 2 is considered to be sufficient for our data during the period under review.

3.0 Concluding remarks

In summary, we have considered mainly four different support vector machines (SVM): support vector classifier (polynomial kernel with degree = 1), polynomial kernel of degree = 2, radial basis kernel and sigmoid kernel in modeling and forecasting the direction of exchange rate in Nigeria between 2004 and 2017 using two macroeconomic variables such as the crude oil price and inflation rate. We analysed the data using these four statistical learning methods of classification. Firstly, direction of exchange rate which was classified into $k = 3$ levels i.e. "*down*", "*unchanged*" and "*up*", revealed that during the period under review, it rose in 83 months, decline in 81 months and remained unchanged for three consecutive months (between August 2008 through October 2008) out of the 167 months (**Table** 1). This research work favours the polynomial basis kernels and the sigmoid kernel due to the equal lower test error rates when compared with the training error rates and better decision boundaries at the same time avoiding overfitting the models. However, polynomial basis kernel with degree = 2 was able to classify one of the observations in the "*unchanged*" class correctly. Hence this SVM is recommended for the data on the direction of exchange rate in Nigeria during the period under review. The results of the SVMs discussed in this paper indicated that though these macroeconomic variables are theoretically important in forecasting the direction of exchange rate in Nigeria, high training and test error rates show that the four SVMs considered are generally weak classification techniques when applied to the exchange rate data.

References

1. Lawal Ganiyu Omoniyi, Aweda Nurudeen Olawale (2015), "An Application of ARDL Bounds Testing Procedure to the Estimation of Level Relationship between Exchange Rate, Crude Oil Price and Inflation Rate in Nigeria", *International Journal of Statistics and Applications*, DOI: 10.5923/j.statistics.20150502.06, Vol 5(2): 81-90.

2. Nasution B.P and Agah A. (2000), "Currency Exchange Rate Forecasting with Neural Networks", *Journal of Intelligent System,* Vol 10, No 3, pg. 219-253.
3. Avav Quandlous (2006), "Performance of Deterministic and Stochastic Trends Model in Forecasting the Behaviour of the Canadian dollar and the Japanese Yen against The Us Dollar", *International Business & Economics Research Journal,* Vol. 5, No. 12, pg. 1-16.
4. Clements K. W and Lan Y. (2006), *"A New Approach to Forecasting Exchange Rates"*: The University of Western Australia, Business School.
5. Barbara Rossi (2013), "Exchange Rate Predictability", *Journal of Economic Literature*. J.E.L. Codes F3, C5, pg. 1- 74.
6. Beckmann, J., and Czudaj, R. (2013a). Oil prices and effective dollar exchange rates. *International Review of Economics & Finance*, 27(1), 621-636.
7. Beckmann, J., and Czudaj, R. (2013b). Is there a homogeneous causality pattern between oil prices and currencies of oil importers and exporters? *Energy Economics*, 40(1), 665-678.
8. Onasanya O.K. and Adeniji O.E. (2013), "Forecasting the Direction of Exchange rate between Nigerian naira and US dollar using Time Domain Model", *International Journal of Development and Economic Sustainability*, Vol. 1, No. 1, pp. 45-55.
9. Michele Ca'Zorsi, Marcin Kolasa, Michał Rubaszek (2016), *"Exchange Rate Forecasting using DSGE Models."*, European Central Bank, Working Paper.
10. Nwankwo Steve (2014), "Autoregressive Integrated Moving Average (ARIMA) Model for Exchange Rate (Naira to Dollar)", *Academic Journal of Interdisciplinary Studies*, Vol. 3, No. 4, pp. 429-433.
11. Jamal Bouoiyour, Refk Selmi, Aviral Kumar Tiwari, Muhammad Shahbaz (2015), "The nexus between oil price and Russia's real exchange rate: Better paths via unconditional vs conditional analysis", *Journal of Energy Economics,* DOI: 10.1016/j.eneco.2015.06.001, Vol. 51, pg. 54-66.
12. Tariq Mahmood Ali, Muhammad Tariq Mahmood, Tariq Bashir (2015), "Impact of Interest Rate, Inflation and Money Supply on Exchange Rate Volatility in Pakistan." *World Applied Sciences Journal*, Vol. 33, No. 4, pp. 620-630.
13. Yin-Wong, Menzie D. Chinn, Antonio Garcia Pascual, Yi Zhang (2017), "*Exchange Rate Prediction Redux: New Models, New Data, New Currencies."* European Central Bank, Working Paper.
14. Cheung, Yin-Wong, Menzie Chinn, and Antonio Garcia Pascual (2005), "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?" *Journal of International Money and Finance* 24 (November): 1150-1175. *NBER Working Paper* no. 9393 (December 2002).
15. Cheung, Yin-Wong and Menzie Chinn (1998), "Integration, Cointegration, and the Forecast Consistency of Structural Exchange Rate Models," *Journal of International Money and Finance 17*(5): 813-830.
16. Ni Yang (2011), "Prediction of Chaotic Exchange Rate Time Series Applying Dynamic Component Predicting Model", Published in**:** Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011 2nd International Conference, DOI: 10.1109/AIMSEC.2011.6009803, INSPEC Accession Number: 12259077.
17. ***R*** Core Team, ***R***: A language and environment for statistical computing. ***R*** Foundation for statistical computing, (2018) Vienna, Austria. URL http://www.R-project.org/.