

Assimilation of Principal Component Analysis and Wavelet with Kernel Support Vector Regression for Medium-Term Financial Time Series Forecasting



Baba Gimba Alhassan, Fadhilah Binti Yusof, Siti Mariam Norrulashikin

Abstract: Entities and institutional financiers have gained a lot of growth from financial time series forecasting in recent times. But the major challenges of financial time series data are the high noise and complexity of its nature. Researchers in recent times have successfully engaged the application of support vector regression (SVR) to conquer this challenge. In this study principal component analysis (PCA) is applied to extract the low dimensionality and efficient feature information, while wavelet is used to pre-process the extracted features in order to nullify the influence of the noise in the features with a KSVR based forecasting model. The analysis is carried out based on the quarterly tax revenue data of 39 years from the first quarter of 1981 to the last quarter of 2016. The forecasting is made for ten quarters ahead. The initial empirical result shows that the multicollinearity has been reduced to zero (0), and the analytic result reveals that the proposed model PCA-W-KSVR outperforms KSVR, PCA-KSVR, and W-KSVR in terms of MAE, MAPE, MSE and RMSE

Keywords: Principal Component Analysis, Dimensionality, Financial time series, Forecasting, Tax revenue

I. INTRODUCTION

Scientific analysis is a prevalent method to study economic growth and development with extreme attention being given to financial time series forecasting from both the private and public institutional investors. This is because accurate forecasting induces several financial investment decisions. These decisions are largely based on analysis of tax revenue data with consideration of its intensity, hidden relationships, a high degree of uncertainty, unstructured nature, noise and non-stationarity [1, 2].

Tax revenue trends influence several factors such as economic condition, economic policies, political events, news related to foreign trade policy, etc. However, tax revenue data are highly non-linear, low dimensional and its generations are not random from academic investigations (Cao & Tay, 2003; Huang et al., 2005). Thus, the tax revenue forecasting task is highly challenging.

The short, medium-term or long-term forecast results obtained from technical analyses are not promptly reported and are based on economic and fiscal policies that affect tax revenue over time [3, 4]. The technical components used in this analysis were obtained from tax revenue. Researchers applied different artificial intelligence and machine learning approaches to analyze these technical components to predict future economic growth and development. The traditional statistical approach like a vector autoregression model by [5, 6]. Continuous researchers have introduced numerous approaches including Ensemble Empirical Mode Decomposition (EEMD), Artificial Neural Network (ANN), Fuzzy Logic (FL), Genetic Algorithm (GA) and others (Yao & Herbert, 2009; AlDarmaki et al., 2016; Leea et al., 2017).

These techniques suffer one of the different challenges like finding the optimum solutions, initializing a large number of control parameters, over-fitting, under-fitting, etc. To overcome most of these challenges, Support Vector Regression (SVR) has been extensively used in different nonlinear regression tasks. SVR is a functional estimation method for the structural risk minimization principle while the empirical risk minimization principle uses traditional methods for application. SVR having successfully applied in numerous time series problems [7-11] stimulated its adaption in financial time series forecasting.

The initial essential step in breeding an SVR based forecasting technique is to transform the original features into new ones (feature extraction) and to choose the most influential set of features (feature selection). The most widely used feature extraction method is Principal component analysis (PCA) for the framework of SVR [11, 12]. PCA converts dimensional input vectors (data) into uncorrelated principal components (PCs) by computing the eigenvectors of the covariance matrix of initial inputs [13].

Once again, the dormant noise existing in financial time series data impairs the performance of the forecasting system by over-fitting or under-fitting.

Revised Manuscript Received on March 13, 2020.

* Correspondence Author

Baba Gimba Alhassan*, Department of Mathematical Sciences, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Malaysia
Department of Statistics, School of Applied and Natural Sciences Federal Polytechnic Bida (FPB) Niger State Nigeria e-mail gazhigun@gmail.com

Fadhilah Binti Yusof, Department of Mathematical Sciences, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Malaysia. e-mail fadhilahy@utm.my

Siti Mariam Norrulashikin, Department of Statistics, School of Applied and Natural Sciences Federal Polytechnic Bida (FPB) Niger State Nigeria. e-mail sitimariam@utm.my

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

To negate the influence of such noise in data to improve the forecasting accuracy proposed the use of Wavelet for both linear and non-linear with SVR [12, 14-20]. In both approaches, wavelet is used at the beginning to extract the most important components from the technical variables and fed them to SVR for better forecasting purposes. Wavelet is a signal processing technique developed for unsighted source separation. In its bids to attain statistical components (Cs) from the transformed vectors. [21, 22] has shown that both PCA and Wavelet can improve the performance of KSVR in time series forecasting [23, 24] which inspired this research work to take on PCA and Wavelet with KSVR for forecasting tax revenue growth.

In this study, a KSVR based forecasting model is developed to blend both PCA and Wavelet to improve the forecast accuracy for tax revenue because a small improvement of this performance can have a significant influence on economic growth and development. The PCA is used to extract the influential components from input features which are then filtered to transform the high-dimensional input into low-dimension features. After that, the wavelet is applied to convert the reduced features into components. The KSVR then finally uses the filtered and transformed low-dimensional input variables to construct the forecasting model and predict tax revenue for the next ten quarters in advance. The performance of the proposed method is compared with three conventional methods: the combination of PCA with KSVR (PCA-KSVR), Wavelet with KSVR (W-KSVR) and single KSVR.

II. METHODOLOGY

2.1 Research Data

The study is conducted to evaluate the performance of the proposed model using 36 years of historical data of the tax revenue from the first quarter of 1981 to the last quarter of 2016. This data is obtained from the National Bureau of Statistics (NBS) and the Central Bank of Nigeria (CBN). The data covers 144 quarters which comprises of company tax, customs and exercise duty tax, personal income tax, property tax, sales tax and total tax revenue for the period. 80% of the total sample data point (123 quarters) was used as a training sample and the remaining 20% of the data points (21 quarters) were holdout as the testing sample.

2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA), designed by [11, 25], is a famous statistical technique for feature extraction. It realizes a minor number of uncorrelated components from high dimensional original inputs by computing the eigenvectors of the covariance matrix. Given a set of m dimensional input vectors $x_i=(x_i(1), x_i(2), \dots, x_i(m))^T$ where $i=1, 2, \dots, n$. PCA is a transformation of x_i into a new vector y_i as given in Equation (1).

$$y_i = U^T x_i \tag{1}$$

where U is the $m \times m$ orthogonal matrix whose j^{th} column u_j is the j^{th} eigenvector of the sample covariance matrix $C = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$. In other words, PCA solves the eigenvalue problem of Equation (2)

$$\lambda_j u_j = C u_j, j = 1, 2, \dots, m \tag{2}$$

where λ is one of the eigenvalues of C and u_j is the corresponding eigenvector. Based on the estimates, u_j , the components of y_i are then calculated as the orthogonal transformation of x_i : using Equation (3).

$$y_i(j) = u_j^T x_i, j = 1, 2, \dots, m \tag{3}$$

The new components are called principal components. By using only, the first several eigenvectors sorted in descending order of the eigenvalues, the number of principal components in y_i can be reduced [26, 27].

2.3 Discrete Wavelet Transform

Wavelets have become a well-known tool for time series forecasting. The basic aim of the wavelet transform is to analyze the time-series data in both time and frequency domains. This is obtained by decomposing the original time series into different frequency bands using wavelet functions. This is against the Fourier Transform (FT) in which sine and cosine functions are used to analyze time series. Wavelet transformations provide a decomposition of the original time series by capturing information on the various decomposition levels.

To obtain a number of decomposition level, the following formula (Equation (4) by Wang & Ding (2003) is applied

$$L = \text{int}[\log(N)] \tag{4}$$

Where L is the decomposition level and N is the number of the tax revenue data series. In this formula, the original tax revenue data is decomposed into L components from A to D_{L-1} ($A, D_1, D_2, \dots, D_{L-1}$) which stands for different frequency components of the original data. Each of these components plays a unique role in the original tax revenue data and has different effects on the original data. Instead of using the D 's component separately, as an input model, suitable values of D 's component which are more useful and capable of increasing the forecast performances of the models are added. If a continuous time series $x(t)$, $t \in [-\infty, \infty]$, a wavelet function can be written as given in Equation (5).

$$\psi(t, s) = \frac{1}{\sqrt{s}} \psi\left[\frac{t-\tau}{s}\right] \tag{5}$$

Where t represents time, τ signifies the time step where the window function is iterated and $s \in [0, \infty]$ stands for the wavelet scale: $\psi(t)$ called the mother wavelet can be defined as $\int_{-\infty}^{\infty} \psi(t) dt = 0$. The continuous wavelet transforms (CWT) is therefore given in Equation (6).

$$W(\tau, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t) \bar{\psi}\left[\frac{t-\tau}{s}\right] dt \tag{6}$$

where $\bar{\psi}(t)$ stands for complex conjugation $\psi(t)$. $W(\tau, s)$ represents the sum over all the time of time series multiplied by scale and the shifted time version of the wavelet function $\psi(t)$. As stated previously, using CWT in forecasting is not practically possible since the wavelet coefficient must be calculated at every possible scale which is time-consuming, and it generates a lot of data.

Consequent upon this, therefore, Discrete Wavelet Transformation (DWT) is adopted in most of the forecasting problems due to its simplicity and ability to compute using less time [28-37]. The methodology that is adopted in using the DWT is the choice of scales and positions on powers of 2, which make the analysis to be more efficient and more accurate. The benefit of using the DWT is because of its robustness since it does not involve any erroneous assumption and or testing procedure of parameters said (Kisi, 2010). Therefore, DWT is given in Equation (7)

$$\psi_{m,n} \left[\frac{t-\tau}{s} \right] = \frac{1}{\sqrt{S_0^{m/2}}} \psi \left[\frac{t-n\tau_0 S_0^m}{S_0^m} \right] \quad (7)$$

Where m and n are integers which control the scale and time respectively; S_0 is a specified, fixed dilation step greater

than 1; and τ_0 is the location parameter, which must be greater than zero. The most common choices for the parameters $S_0=2$ and $\tau_0=1$. For a discrete-time series $x(t)$

where $x(t)$ occurs at discrete time t , the DWT becomes:

$$W_{m,n} = 2^{-m/2} \sum_{t=0}^{N-1} \psi(2^{-m}t - n) x(t) \quad (8)$$

Where $W_{m,n}$ is the wavelet coefficient for the discrete wavelet at scale $s = 2^m$ and $\tau = 2^m n$?

2.4 Support Vector Regression

The SVR broadens the essential principles of Vapnik's support vector machines (SVM) for classification [38, 39] by setting a leeway of tolerance ϵ in approximation and up until the threshold ϵ , 0 error is considered. Given a training set (x_i, y_i) , $i=1, 2, \dots, n$, where the $x_i \in R^m$ is the m -dimensional input vector and $y_i \in R$ is the response variable. SVR generates the linear regression function in the form:

$$f(x, w) = w^T x + b \quad (9)$$

Vapnik's linear ϵ -Insensitivity loss (error) function is:

$$|y - f(x, w)| = \begin{cases} 0 & \text{(if } |y - f(x, w)| \leq \epsilon) \\ |y - f(x, w)| - \epsilon & \text{Otherwise} \end{cases} \quad (10)$$

On this note, linear regression $f(x, w)$ is projected by instantaneously minimizing $\|w\|^2$ and the sum of the linear ϵ Tactlessness losses as shown in equation (12). The constant C restrains a trade-off amongst an estimated error and the weight vector norm $\|w\|$ is a design parameter chosen by the user.

$$R = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n |y - f(x, w)| \epsilon \quad (11)$$

Minimizing the risk R is equivalent to minimizing the following risk under the constraints mentioned in equations (13) - (15).

$$R = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (12)$$

$$(w^T x_i + b) - y_i \leq \epsilon + \xi_i \quad (13)$$

$$y_i - (w^T x_i + b) \leq \epsilon + \xi_i^* \quad (14)$$

$$\xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, m \quad (15)$$

Here ξ_i and ξ_i^* are slack variables, one for exceeding the target value by more than ϵ and other for being less than ϵ below the target. As used in SVM, the above constrained optimization problem is solved using Lagrangian theory and the KarushKuhn-Tucker conditions to obtain the desired weight vector of the regression function.

SVR maps the input vectors $x_i \in R^m$ into a high dimensional feature space $\phi(x_i) \in H$. A kernel function $K(x_i, x_j)$ performs the mapping $\phi(\cdot)$. The most popular kernel function that is used in this study is Radial Basis Function (RBS) as shown in equation (16).

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (16)$$

where γ is the constant of the kernel function. The RBF kernel function parameter γ and regularization constant C are the design parameters of KSVR.

Table 1 Technical indicators and their formulas.

S/N	Feature	Description	Calculation Formula
1	10-qtrs SMA	Simple 10-qtrs moving average	$\left(\frac{1}{2}\right) \sum_{i=t-n+1}^t C_i$, where C_i is the total tax revenue
2	10-qtrs EMA	Exponential 10-qtrs moving average	$EMA(n)_{t-1} + \alpha \times (C_t - EMA(n)_{t-1})$, where α is a smoothing factor and $\alpha = \frac{2}{n+1}$
3	10-qtrs WMA	Weighted 10-qtrs moving average	$\frac{(n)C_t + (n-1)C_{t-1} + \dots + C_{t-n+1}}{n+(n-1)+\dots+1}$
4	A/D Oscillator	Accumulation/distribution oscillator. It is a momentum indicator that relates changes in tax revenue generation	$\frac{H_t - C_{t-1}}{H_t - L_t}$, where L_t is the minimum tax value at and H_t is the total tax revenue
5	Stochastic %K	Stochastic %K. it compares where total tax relative to its component over a given period	$\frac{C_t - LL_{t-(n-1)}}{HH_{t-(n-1)} - LL_{t-(n-1)}} \times 100$, LL_t and HH_t mean lowest and highest high in time t.
6	Disparity 5	5-qtrs disparity	$\frac{C_t}{MA_5} \times 100$
7	Disparity 10	10-qtrs disparity	$\frac{C_t}{MA_{10}} \times 100$

To evaluate the performance of the proposed model, Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), relative Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) are used. Formulas of these evaluation measures are shown in equations (17) – (20)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{|A_t|} \times 100 \quad (17)$$

$$MAP = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{|A_t|} \quad (18)$$



$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{A_t - F_t}{A_t} \right)^2} \quad (19)$$

$$MSE = \frac{1}{2} \sum_{i=1}^n (A_i - F_i)^2 \quad (20)$$

where A_t is the actual value and F_t is the predicted value. To calculate all these measures, a 5-fold cross-validation method is used. These are the measures of deviation between actual and predicted prices. The prediction model should produce lower values for all four measures.

2.5 Proposed Principal Component Analysis and Wavelet with Kernel Support Vector Regression PCA-W-KSVR

The proposed model PCA-W-SVR in this research scheme is depicted in Figure 1. In the first stage, data are imputed into PCA to extract features which were then reduced into the low-dimensionality and multicollinearity in the feature space. The extracted features were filtered through the Wavelet has pre-processing. Finally, the pre-processed extracted features are used in the SVR to make forecasting for the future.

First of all, technical analysis will be conducted on the dataset and 6 technical indicators (TIs) will be calculated that would be used by financial experts [40, 41]. Some important technical indicators and their formulas are shown in Table 1. All values of these constructed features are scaled into the range of [0, 1] to eliminate the biases towards larger value attributes. Then PCA is applied to the normalized data to extract the PCs containing the most influential information. These PCs are filtered according to the corresponding variance and thus the irrelevant features are discarded to construct a reduced feature space. The Wavelets model is then used in the low-dimensional data to estimate Wavelets containing the hidden and effective information of the prediction variables. Finally, the Wavelets are used as input variables to construct the SVR tax revenue, forecasting model.

Table 2 Model Selection for Kernel Support Vector Regression

Model Selection for Kernel Support Vector Regression (KSVR)				
	Models Linear	Polynomial	Radial Tangent	Laplace Bessel Anova
MSE	37.6%	80.1%	27%	18%
	99.08%	33.1%	9.35%	
R^2_{train}	96.7%	57.5%	11.4%	
	43.30%	29.4%	71.3%	7.34%
R^2_{test}	99.6%	78.9%	34%	
	47.11%	60.0%	60.8%	6.88%
RMSE	6.13%	8.91%	5.19%	4.24%
	9.95%	5.75%	3.05%	

Table 3: Grid search results for linear based function LBF of Kernel parameters

Grid search results for linear based function LBF of Kernel parameters.			
Parameter	Indirect Tax	Manufacturing Tax	Value Added Tax (VAT)
III. C	$e^{9.0}$	$e^{8.4}$	$e^{5.8}$
γ	$e^{2.8}$	$e^{2.0}$	$e^{2.4}$

As mention in Section, the linear based function (LBF) kernel para is incorporated in this study because it is the most widely used and well-performing kernel function for forecasting purposes. But the performance of SVR is highly influenced by the selection of the parameters: γ and C. A very popular method to select the best values of these parameters is the grid search approach with cross-validation Hsu et al. (2003). Equation (17-20) was used to checking the model performance.

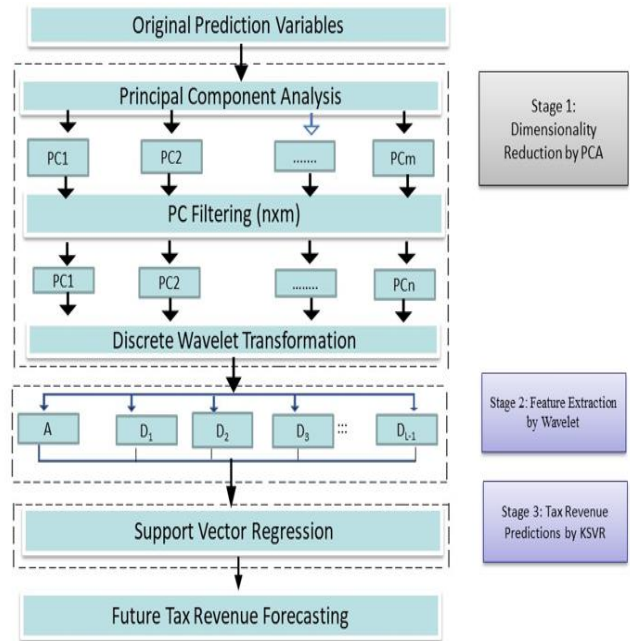


Figure 1 the PCA-W-VSR model

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The principal component analysis reveals from the original data that there is a high correlation in the data

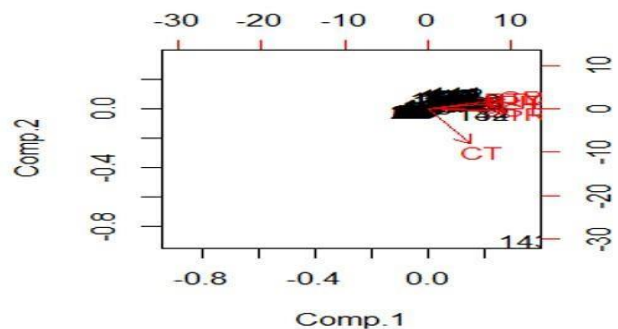


Figure 2 Correlation amongst the independent variables

Figure 4.4 reveals how the correlation amongst the independent variables of the original data, where the red line arrows are the independent variables (components) and the black areas are the concentration of the observations in the components which speak out the degree of multicollinearity in the data that is observed to be very high except CT (company tax) variable that is on negative side of the relationships with other variables meaning that multicollinearity exists in the data



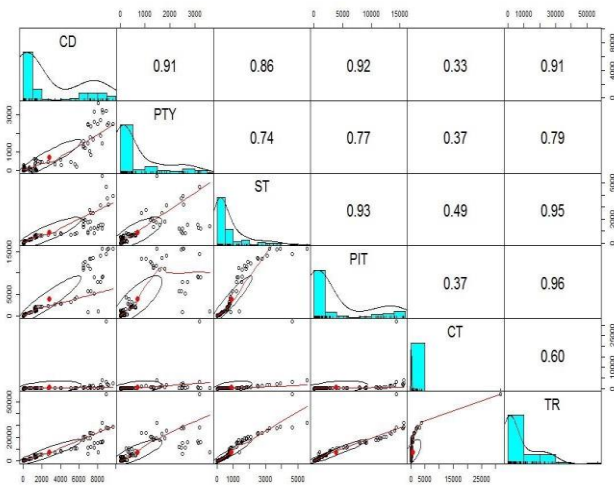


Figure 3 Multicollinearity amongst the variables

Figure 4.5 indicates the degree of correlation amongst the independent variables. They are all positive and very high above 70% except company tax that is below 40%. However, when the independent variable (component) with high correlation leads to a multicollinearity problem [42-44] which will cause instability in the data or cannot allow accurate forecast. More so, the lower diagonal of the graph shows individual graphs of each component and also how the relationship of various components continues to raise except the company tax which is on the negative side despite the increase on the x-axis the y-axis remains constant.

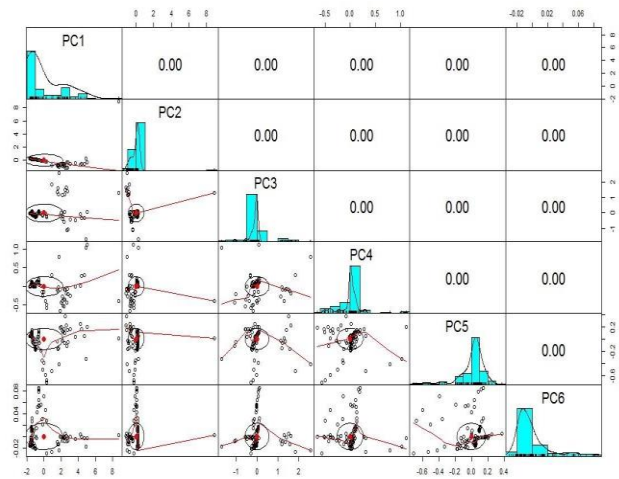


Figure 4 Transformation of the original data

Having applied PCA on the data, Figure 4.6 indicates the degree of correlation amongst the independent variables are all zero (0). That is by implication there's no more multicollinearity in the data which means forecasting tax revenue using the multivariate time series technique would yield a better result.

3.1 Forecasting Results of the Error from the Models

The results in table 4 prove the evidence that the proposed PCA-W-SVR model has produced lower MAPE (%), MAE, MSE and RMSE (%) for the tax revenue. The assimilation of PCA and wavelet has improved the performance of KSVR in most of the cases but the proposed PCA-W-SVR model outperforms the other three compared methods. This validates that the proposed PCA-W-SVR method can generate lower prediction errors than the other three compared approaches. Again, it could be noticed from the results that, the forecasting performance of all the approaches decreases as the predictions are made for a greater number of quarters in advance, which may be obvious for any prediction system. The robustness of the proposed PCA-W-SVR method is evaluated by comparing its performance with PCA-SVR, Wavelet-SVR and single SVR methods using different ratios of training and testing sample sizes. The performance is compared in terms of MAPE (%) and RMSE (%)

Table 4: Prediction performance of quarters ahead for Total Tax Revenue

Quarter No;	Performance Measures				
	Prediction models	MAPE%	MAE	MSE	RMSE
1 st Qtr	PCA-Wavelet-KSVR	0.13	0.0013	0.2052	0.4613
	PCA-KSVR	0.201	0.00201	0.4119	0.6417
	Wavelet-KSVR	0.205	0.0025	0.3857	0.621
	KSVR	0.19	0.0019	0.1901	0.436
2 nd Qtr	PCA-Wavelet-KSVR	0.119	0.00119	0.1952	0.4721
	PCA-KSVR	0.211	0.00211	0.3123	0.6501
	Wavelet-KSVR	0.285	0.00285	0.3217	0.5201
	KSVR	0.201	0.00201	0.2001	0.4265
3 rd Qtr	PCA-Wavelet-KSVR	0.109	0.00109	0.2152	0.4763
	PCA-KSVR	0.311	0.00311	0.3219	0.6917
	Wavelet-KSVR	0.194	0.00194	0.3677	0.684

4 th Qtr	KSVR	0.145	0.00145	0.1921	0.429
	PCA-Wavelet-KSVR	0.111	0.00111	0.2142	0.4513
	PCA-KSVR	0.281	0.00281	0.4269	0.6547
	Wavelet-KSVR	0.195	0.00195	0.3317	0.611
5 th Qtr	KSVR	0.179	0.00179	0.1721	0.433
	PCA-Wavelet-KSVR	0.121	0.00121	0.1802	0.4521
	PCA-KSVR	0.331	0.00331	0.3999	0.6617
	Wavelet-KSVR	0.24	0.0024	0.4157	0.561
6 th Qtr	KSVR	0.161	0.00116	0.1911	0.326
	PCA-Wavelet-KSVR	0.126	0.00126	0.2153	0.4433
	PCA-KSVR	0.222	0.00222	0.4213	0.6457
	Wavelet-KSVR	0.231	0.00231	0.3567	0.672
7 th Qtr	KSVR	0.145	0.00145	0.2466	0.446
	PCA-Wavelet-KSVR	0.102	0.00102	0.1992	0.4513
	PCA-KSVR	0.191	0.00191	0.3299	0.6411
	Wavelet-KSVR	0.198	0.00198	0.3661	0.6731
8 th Qtr	KSVR	0.132	0.00132	0.201	0.411
	PCA-Wavelet-KSVR	0.11	0.001	0.1782	0.3433
	PCA-KSVR	0.179	0.00179	0.3449	0.5267
	Wavelet-KSVR	0.183	0.00183	0.3537	0.621
9 th Qtr	KSVR	0.121	0.00121	0.1801	0.335
	PCA-Wavelet-KSVR	0.13	0.0013	0.2052	0.4103
	PCA-KSVR	0.201	0.00201	0.4119	0.6521
	Wavelet-KSVR	0.205	0.00205	0.3857	0.521
10 th Qtr	KSVR	0.19	0.0019	0.1901	0.436
	PCA-Wavelet-KSVR	0.091	0.00091	0.1952	0.4613
	PCA-KSVR	0.121	0.00121	0.3249	0.5617
	Wavelet-KSVR	0.144	0.00144	0.3457	0.5811
	KSVR	0.109	0.00109	0.1991	0.385

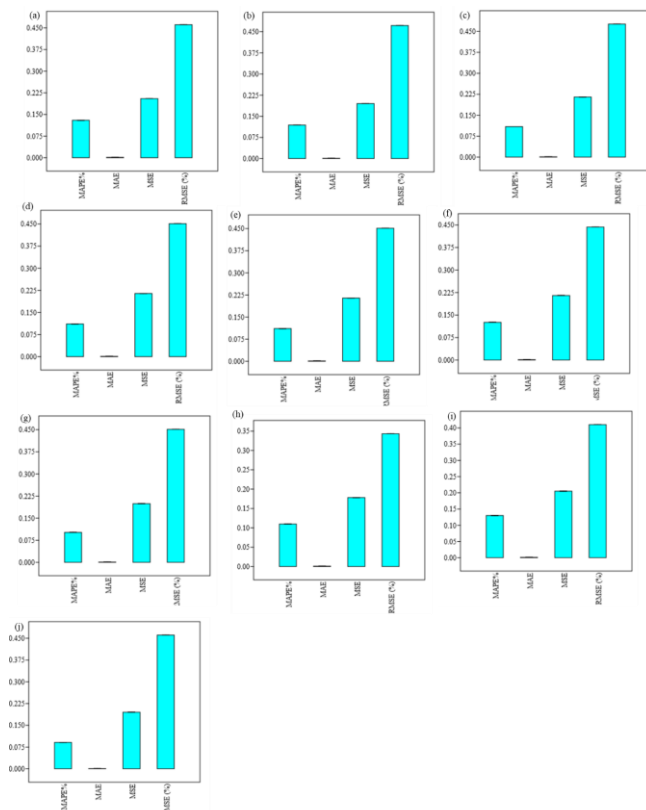


Figure 4 Indicates performance measures of the Errors 1st and 2nd Quarters

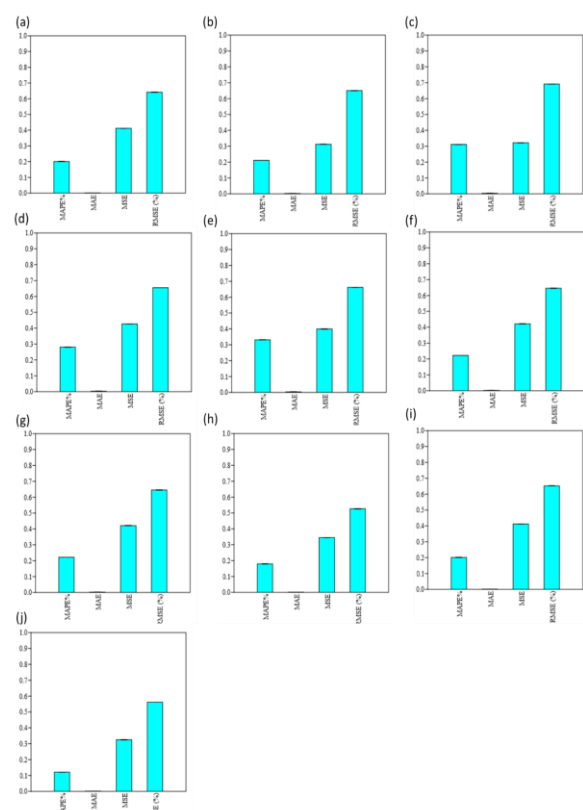


Figure 4 Indicates performance measures of the Errors 3rd and 4th Quarters

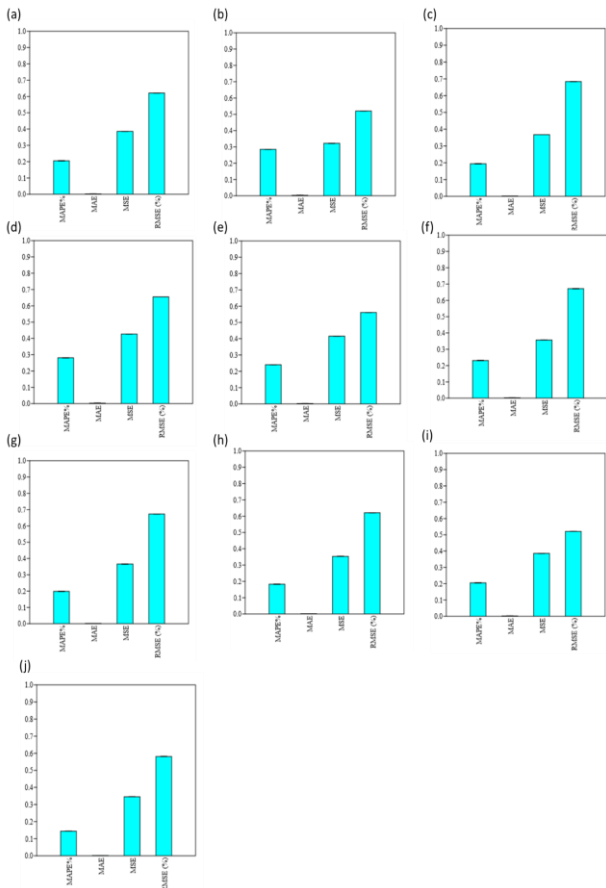


Figure 4 Indicates performance measures of the Errors 5th and 6th Quarters

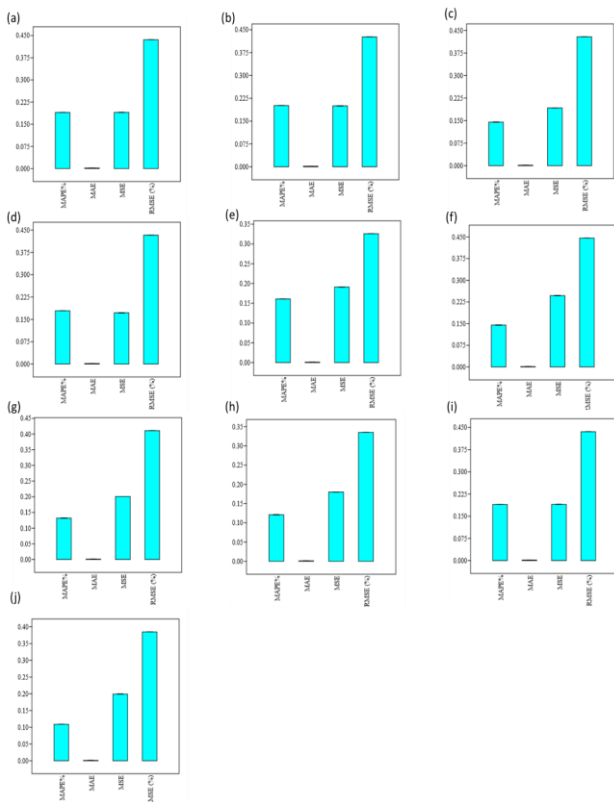


Figure 4 Indicates performance measures of the Errors 7th and 8th Quarters

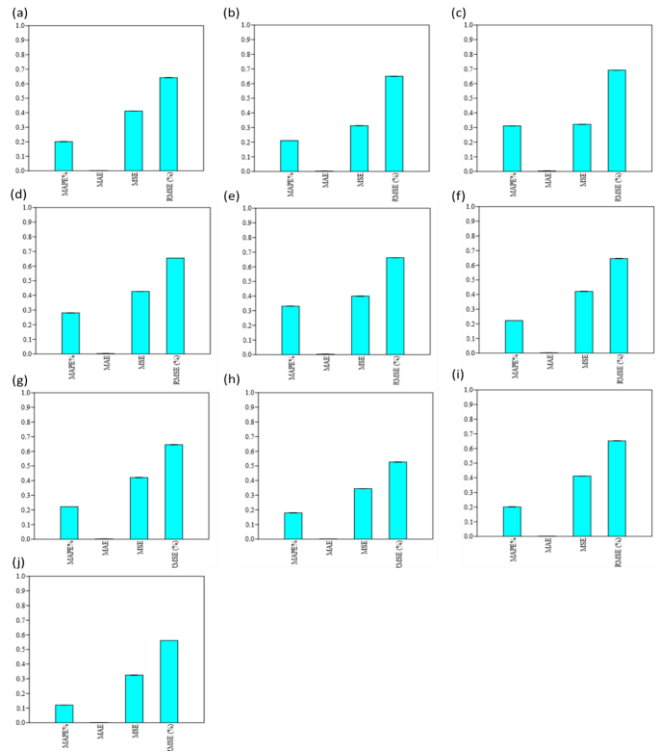


Figure 4 Indicates performance measures of the Errors 9th and 10th Quarters

V. CONCLUSION

This study proposed a tax revenue forecasting model assimilating PCA and Wavelet with SVR for financial time series. This PCA-W-SVR model first applied PCA to extract the best significant components from the input features in order to overcome the problem of over-fitting or under-fitting caused by the noisy sort of financial time series data. The filtered PCs were then processed by wavelet to estimate DWTs which were finally used in KSVR with LBF kernel function as input variables. The grid search for the best kernel parameters is performed to enhance KSVR's performance. The investigations have evaluated 36 years' data for tax revenue data in Nigeria. The proposed model performance is compared with PCA-KSVR, W-KSVR, and single KSVR for medium-term durations (10 quarters) in terms of prediction error.

Experimental results show that the proposed PCA-W-SVR model outperforms all three other approaches by producing less predictive errors. The empirical results can infer that the PCA and Wavelet, employed together, can effectively unwrap the significant information from the original data and improve the performance of KSVR in tax revenue forecasting. As the proposed model helps to predict tax revenue with less error, tax administrators and management can use this to generate more revenue or create more ways that can bring changes such that more people can be brought into the tax net or database. Again, this proposed approach can also be used in other domains like weather forecasting, energy consumption forecasting or stock market index forecasting.

Future research may integrate Kernel PCA, non-linear CCA and other signal processing techniques like affinity propagation clustering with SVR to further enhance the forecasting performance. This paper focuses on medium-term tax revenue prediction. Its applicability might be examined for long-term forecasting in the future and appropriate methods could be integrated to enhance performance in the future. However, only the price related historical data is used here to predict future prices. But it is well known that various other aspects like general economic conditions, government policies, company performance, investor's interest, stock market, etc. also play vital roles in tax revenue generation. In the future, these aspects can also be incorporated as input features for prediction which may buttress the accurate prediction.

REFERENCES

1. Vedachalam, N., G.A.a. Ramadass, and M.A. Atmanand, Importance of Lithium-Based Energy Storage in Achieving India's Climate Goals. *IEEE India* 2019. **14**(4): p. 54-134.
2. Jones, M.a. and B.C. Love, Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral And Brain Sciences* 2011. **34**: p. 169-231.
3. Tay, F.E.H.a. and L. Lijuan Cao, Application of Support Vector Machines in Financial time series forecasting. *Omega International Journal of Management Sciences*, 2001. **29**: p. 309-317.
4. Jua, X., et al.,(2014) SupportVectorRegression and time series Analysis for the forecasting of Bayannur'sTotal WaterRequirement. Available online at www.sciencedirect.com ScienceDirect Procedia Computer Science 2014. **31**: p. 523 - 531.
5. Kim, K.a. and I. Han, Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of a stock price index. *Expert Systems with Applications*, 2000. **19**: p. 125-132.
6. Patel, J., et al., Predicting stock market index using the fusion of machine learning techniques. *Expert Systems with Applications* journal homepage: www.elsevier.com/locate/eswa, 2015.
7. Elattar, E.E., J.Y.a. Goulermas, and Q.H. Wu, Electric load forecasting based on locally weighted support vector regression. *IEEE Transactions on Systems Man and Cybernetics-Part C (Applications and Reviews)*, 2010. **40**(4): p. 439-450.
8. Zhang, X., J.a. Wanga, and K. Zhang, Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. *Electric Power Systems Research* journal home page: www.elsevier.com/locate/epsc, 2017. **146**: p. 270-285.
9. Hong, W.C., et al., Hybrid Evolutionary Algorithms in an SVR Traffic Flow. *Applied Mathematics and Computation* journal homepage: www.elsevier.com/locate/amc, 2011. **217**: p. 6733-6747.
10. Li, M.W., D.F.a. Han and W. Wang, Vessel traffic flow forecasting by RSVR with chaotic cloud simulated annealing genetic algorithm and KPCA. *Neurocomputing* journal homepage: www.elsevier.com/locate/neucom, 2015. **157**: p. 243-255.
11. Lin, L., et al., On-line prediction of border crossing traffic using an enhanced Spinning Network method. *Transportation Research Part C* journal homepage: www.elsevier.com/locate/trc, 2014. **43**: p. 158-173.
12. Sheta, F.A., S.E.M.a. Ahmed, and H. Faris, A comparison between regression, artificial neural networks and support vector machines for predicting stock market index. (*IJARAI*) *International Journal of Advanced Research in Artificial Intelligence*, 2015. **4**(7): p. 55-64.
13. Utpala, U.C., C.S.a. Kumar, and H.M. Tanvir, Short-Term Financial Time Series Forecasting Integrating Principal Component Analysis and Independent Component Analysis with Support Vector Regression. *Journal of Computer and Communications*, 2018. **06**(03): p. 51-67.
14. Traveling, T.B.a. and H. Ince, Support Vector Machine for Regression and. *IEEE*, 2000: p. 348-353.
15. Tay, F.E.H.a. and L. Cao, Application of support vector machines in financial time series forecasting. *Omega International Journal of Management Sciences*, 2001. **29**: p. 309-317.
16. Okasha, M.K., Using support vector machines in financial time series forecasting. *International Journal of Statistics and Applications* 2014. **4**(1): p. 28-39.
17. Cao, L.J.a. and F.E.H. Tay, Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 2003. **14**(6): p. 1506-1520.
18. Law, T.a. and J.S. Taylor, Practical Bayesian support vector regression for financial time series prediction and market condition change detection. *Quantitative Finance* ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: <http://www.tandfonline.com/loi/rqf20>, 2017. **17** (9): p. 1403-1416.
19. Cao, L., Support Vector Machines Experts for Time Series Forecasting. *Neurocomputing* 2003. **51**: p. 321-339.
20. Das, S.P.a. and S. Padhy, A novel hybrid model using teaching-learning-based optimization and a support vector machine for commodity futures index forecasting. *Springer-Verlag Berlin Heidelberg. J. Mach. Learn. & Cyber.*, 2018. **9**: p. 97-111.
21. Chen, Y.a. and Y. Hao, A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems With Applications* journal homepage: www.elsevier.com/locate/eswa, 2017. **80**: p. 340-355.
22. Ling-Jing, K., et al., Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. *Neurocomputing*, 2013. **99**: p. 534-542.
23. Chowdhury, U.N., Integration-of-Principal-Component-Analysis-and-Support-Vector-Regression-for-Financial-Time-Series-Forecasting. *International Journal of Computer Science and Information Security*, August 2017, 2017. **15**(8): p. 27-32.
24. Huang, W., Y. Nakamoria, and, S.Y. Wang, Forecasting stock market movement direction with support vector machine. *Computers & Operations Research* 2005. **32**: p. 2513-2522.
25. Wang, W., et al., Generalized autoencoder A neural network framework for dimensionality reduction. *Working Paper IEEEE*, 2014: p. 490-497.
26. Fauve, M., J.a. Chanussot, and J.A. Benediktsson, Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP Journal on Advances in Signal Processing*, 2009(783194): p. 1-14.
27. Du, Q.a. and J.E. Fowler, Low-Complexity Principal Component Analysis for Hyperspectral Image Compression. *International Journal of High-Performance Computing Applications*, 2008. **22**: p. 438-448.
28. Liu, K., et al., Prediction of mild cognitive impairment conversion using a combination of independent component analysis and the cox model. *Frontiers in Human Neuroscience* | www.frontiersin.org, 2017. **11**(33): p. 1-11.
29. Bell, A.J.a. and T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 1995. **7**(6): p. 1004-1034.
30. Akhtar, M.T., et al., Recursive independent component analysis for online blind source separation. *IEEE*, 2012: p. 2813-2817.
31. Cao, L.J., et al., A comparison of PCA, KPCA, and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, 2003. **55**: p. 321-336.
32. Hyvarinen, A.a. and E. Oja, Independent component analysis: algorithms and applications A. *Neural Networks* 13 (2000), 2000. **13**: p. 411-430.
33. Girolami, M.a. and C. Fyfe, Generalised independent component analysis through unsupervised learning with emergent bussgang properties. *IEEE*, 1997: p. 1788-1791.
34. Karhunen, J., et al., A class of neural networks for independent component analysis.
35. Aziz, R., C.K.a. Verma, and N. Srivastava, A fuzzy-based feature selection from independent component subspace for machine learning classification of microarray data. *Genomics Data* journal homepage: www.elsevier.com/locate/gdata, 2016. **8**: p. 4-15.
36. Giannakopoulos, X., J.a. Karhunen, and E. Oja, An experimental comparison of neural algorithms for independent component analysis and blind separation. *International Journal of Neural Systems*, 2014.
37. Hsu, S.H., et al., Real-time adaptive EEG source separation using online recursive independent component analysis. *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING*, 2016. **24**(3): p. 109-210.
38. and, C.C. and V. Vapnik, Support-Vector Networks. *Machine Learning*, 1995. **20**: p. 273-297.

39. Bui, D.T., et al., models for shallow landslide hazards a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and.
40. Shin, K.S., T.S.a. Lee, and H.J. Kim, An application of support vector machines in the bankruptcy prediction model. *Expert Systems with Applications* 2005. **28**: p. 127-135.
41. Hsu, C.W., C.C.a. Chang, and C.J. Lin, A practical guide to support vector classification. <https://www.researchgate.net/publication/272039161>, 2015: p. 1-17.
42. Nollet, J., G.a. Filis, and E. Mitrokostas, Corporate social responsibility and financial performance A non-linear and disaggregated approach. *E Mitrokostas - Economic Modelling*, 2016: p. 1-27.
43. Saeid, E., et al., A Principal Component Regression Method for Estimating Low Flow Index. *Water Resources Management*, 2010. **24**(11): p. 2553-2566.
44. Yu, H., S.a. Jiang, and K.C. Land, M