# COMPARISON OF HYBRID SUPPORT VECTOR MACHINES AND HOLT WINTERS EXPONENTIAL SMOOTHING MODELS IN AIRLINES PASSENGERS' TIME SERIES FORECASTING

**[1]MOHAMMED SALISU ALFA, [2]ABUBAKAR MUHAMMAD AUWAL, [2]ALHAJI ISMAILA SULAIMAN**

*[1]Department of Statistics, The Federal Polytechnic, Bida. Niger State. [2]Department of Statistics, Nasarawa State University, Keffi. Nasarawa State*

**ABSTRACT**

Hybrid Support vector machines (SVMs) and Holt Winters Exponential Smoothing (HW) model is a promising method for the time series forecasting. This paper therefore attempts to compare the potentiality of the hybrid model with the single SVM and HW models using the Airline passengers' data for the analysis to obtain the desired results. The performance measure of evaluation criteria shows that HW model has MSE of 996.10, MAE of 44.64 and coefficient of correlation (R) of 0.75; SVM model has MSE of 925.92, MAE of 25.24 and R of 0.89 and the hybrid model produced 463.72 for MSE, 17.15 for MAE and 0.96 for R. The hybrid model produced the overall minimum error and maximum correlation coefficient which outperform both SVM and HW models in the study. The hybrid model also recorded 149.3% and 35.9% improvement over HW and SVM models respectively. Various results show that hybrid model likewise provides a promising alternative to airlines passengers' data for the time series forecasting.

**Keywords: -** Hybrid, SVM, HW, Time series, Forecasting

## Introduction

Time series forecasting is an important practical problem with a diverse range of applications in many observational disciplines, such as finance, economics, meteorology, biology, medicine, hydrology, oceanography, physics, engineering, and geomorphology. Forecasting can assist them to make a better development and decision-making for most of the organization. The identification of highly accurate and reliable time series forecasting models for future time series is a precondition for successful planning and management for applications in variety of areas. Forecasting involves making projections about future performance on the basis of historical and current data (Kalekar & Bernard, 2004). Forecasting methods are key tools in decision making process in many areas. Time series forecasting is an important area of forecasting in which past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. The model is then used to extrapolate the time series into the future. This modeling approach is particularly useful when little knowledge is available on the underlying data generating process. Much effort has been devoted over the past several decades to the development and improvement of time series forecasting models ( Zhang, 2003).

Exponential smoothing is a procedure for continually revising a forecast in the light of more recent experience. Exponential Smoothing assigns exponentially decreasing weights as the observation get older. In other words, recent observations are given relatively more weight in forecasting than the older observations. (Gelper et al., 2001) described exponential smoothing as a well-known method for smoothing and predicting univariate time series. Similarly, exponential smoothing has a major disadvantage of being strongly influenced by the presence of outliers in the time series (Gelper et al., 2001). Forecasting in time series of a certain phenomenon or variable under study is one of the main reasons for applying time series models. The choice of the forecasting model depends on the data structure and study objectives. Exponential smoothing methods are the most used in time series modeling and forecasting due to their flexibility and the massive model option they integrate. The Holt-Winters (HW) method is an extension of the Holt method, and is applied whenever the data behavior is trendy and is seasonal (Costa et al., 2015). In the view of Bolarinwa et al., (2021), the Holt-Winters' method usually perform better than the double exponential method having recorded lower values of the performance indicators. Khan et al., (2017) said Holt–winters method is applicable to forecast the future state and fitted to forecast monthly revenue in Bangladesh shows that the model is highly applicable.

Support Vector machine is currently a hot topic in the statistical learning area and is now widely used in data classification and regression modeling (Hong et al., 2017). More advanced, Artificial Intelligence (AI) is support vector machine (SVM) is proposed by Vapnik and his co-workers in 1995 through statistical learning theory, has gained the attention of many researchers. SVMs reduce most machine learning problems to optimization problems and optimization lies at the heart of SVMs (Tian et al., 2012). SVM is nonlinear model which is one of the soft computational techniques is a powerful methodology and has been successfully applied to solve various problems (Xie et al., 2006). The standard SVM is solved using quadratic programming methods. This method needs very few assumptions to learn patterns of input variables for prediction variables since it deals with nonlinear models. According to Nayak et al., (2015), the use of SVM in various data mining applications makes it an essential tool in the development of products that have implications for the human society. SVM on the other hand has advantage of not necessarily creating dummy variables while dealing with categorical variables but also there is no limit in the number of independent variables. Wong et al., (2007) stated that the foundation of SVMs was developed by Vapnik, who was a Russian Mathematician at the beginning of 1960s popularly referred to as "Vapnik 1965" whose principle was based on the structural Risk Minimization principle from the statistical learning theory and this gained popularity due to its many attraction features and promising empirical performance. Tian et al., (2012) also viewed that SVM has been proved to be effective in classification by many researchers in different fields of Science, Engineering, Medical and Financial studies Vapnik, (1999) and extended to the domain of regression problems, (Hu & Xiaotong 2017). SVM offers remarkable generalization performance in many areas such as pattern recognition, text classification and regression estimation Evgeniou & Pontil, (2014) and also that some researchers deal with the application of SVM in time series forecasting. According to Wang et. al., (2013) in recent years, SVM has become a popular tool for pattern recognition and machine learning. SVM is used for classification problems and its goal is to optimize ''generalization'' (Chao & Horng, 2015). However, applications of the SVM models in seasonal time series data have not been widely studied (Hsu et al., 2016). Therefore, this study attempts to use the SVM model in the seasonal time series forecasting problems. The support vector machines (SVMs) have been extensively used in several applications which includes the decision-making application Li et. al., (2011), forecasting malaria transmission Kumar, (2016), live fibrosis diagnosis Boži & Stojanovi, (2011) and pattern classification (Liu et. al., 2014).

In this paper, we propose a hybrid approach to time series forecasting using both SVM and HW models. The motivation of the hybrid model arises from the following perspectives. First, it is often difficult in practice to determine whether a time series under study is generated from a linear or nonlinear underlying process or whether one particular method is more effective than the other in out-of-sample forecasting. Therefore, it is difficult for forecasters to select the right method for their distinctive situations. Usually, a number of different models are tried and the one with the most accurate result is selected. However, the final selected model is not necessarily the best for future uses due to many potential influencing factors such as sampling variation, model uncertainty and structure change. By combining different methods, the problem of model selection can be eased with slight extra effort. Second, real-world time series are rarely pure linear or nonlinear because they often contain both linear and nonlinear patterns. If this is the case, then neither SVM nor HW can be suitable in modeling and forecasting time series since the SVM model cannot deal with linear relationships while the HW model alone is not able to handle both linear and nonlinear patterns alike hence, by combining SVM with HW models, the data can then be modeled more accurately. Third, it is almost universally agreed in the forecasting literature that no single technique is best in every situation (Majid, 2018). This is largely due to the fact that a real-world problem is often complex in nature and any single model may not be able to capture different patterns equally well.

Therefore, combining different models can increase the chance to capture different patterns in the data and improve forecasting performance (Zhang, 2003). Several empirical studies have already suggested that by combining several different models, forecasting accuracy can often be improved over the individual model (Zhang et. al., 2015). In addition, the combined model is more robust with regard to the possible structure change in the data.

## Problem Statement and the objective

Exponential smoothing models have been found to be amongst the most effective forecasting models (Fried, 2009). It has been useful in many fields of human activities. Nevertheless, it suffers from the limitation of being able to capture only linear features in time series data. Support vector machines (SVM) on the other hand though new has made remarkable in roads in the field of time series forecasting. The current trend in forecasting practice is the combination of two or more technique into one. This is done to harness the advantages of the practices involved. It is this unique trend that has motivated this study in which we propose to develop a hybrid model which is a mixture of Holt winters (HW) and SVM to help overcome the shortcomings of using either of the two methods. The research objective is to obtain HW and SVM models for airlines passengers' time series data which will be combined to gain the hybrid model and to compare the forecasting performance of these single models with the hybrid model in order to choose the most effective model

## Methodology

The methods used in this paper to achieve the desired objective is the combination of HW and SVM models to obtain the hybrid model in time series forecasting of the airlines passengers'.

## The Holts-Winters exponential smoothing (HW) model

The HW model is an extension of the Holt method and applied when the data is seasonal which can be additive or multiplicative depending on the oscillatory movement along the time period (Makatjane &

Moroke, 2016). In both versions the forecast will depend on the three components of a seasonal time series: its level, its trend and its seasonal coefficient. The additive version is to be considered whenever the seasonal pattern of a series has constant amplitude over time whose series can be written as:-

$$Y_t = T_t, S_t, \varepsilon_t \tag{1}$$

Where $T_t$ represents the trend (the sum of the level and slope of the series at time $t$), $S_t$ is the seasonal component and $\varepsilon_t$ are error terms with mean zero and constant variance. In the case of multiplicative version, the series can be represented by equation (1) above. The recursive equations of additive and multiplicative HW methods for the level, trend and seasonal factors and forecasts can also be given where $Y_t$ is the observed data at time $t$, $S$ is the length of seasonality (the number of months in a season), $F$ is the number of forecasts ahead and $\theta = (\alpha, \beta, \gamma)^T$ represent the vector of the smoothing parameter (R. J. Hyndman, 2008).

Using the Holt-Winters exponential smoothing algorithm, two seasonal adjustment techniques which are additive and multiplicative are available. With respect to additive method, given $X_1, X_2, ..., X_t$ of a time series, the Holt-Winters additive seasonality algorithm calculates a developing trend equation with a seasonal adjustment that is additive. Additive means that the amount of the adjustment is constant for all levels (average value) of the series. The forecasting algorithm uses the following formula:

$$a_t = \alpha(X_t - F_{t-s}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{2}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{3}$$

$$F_t = \gamma(X_t - a_t) + (1-\gamma)F_{t-s} \tag{4}$$

In these equations (2), (3) and (4), $\alpha$ , $\beta$ and $\gamma$ are smoothing constants which are between zero and one. Again, $\alpha_t$ gives the $y$ -intercept (or level) at time $t$, while $b_t$ is the slope at time $t$. The letter $s$ represents the number of periods per year, so the quarterly data is represented by $s = 4$ and monthly data is represented by $s = 12$.

The forecast at time $T$ for the value at time $T+k$ is $a_T + b_T k + F_{[(T+K-1)/S]+1}$. Here $[(T+K-1)/S]$ is the remainder after dividing $T+k-1$ by $s$. That is, this function gives the season (month or quarter) that the observation is obtain from.

Similarly, given observations $X_1, X_2, ..., X_t$ of a time series, the Holt-Winters multiplicative seasonality algorithm computes an evolving trend equation with a seasonal adjustment that is multiplicative. *Multiplicative* means that the amount of the adjustment varies with the level (average value) of the series.

The forecasting algorithm makes use of the following formulas:

$$a_t = \alpha(X_t / F_{t-s}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{5}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{6}$$

$$F_t = \gamma\left(X_t / a_t\right) + \left(1-\gamma\right) F_{t-s} \quad (7)$$

It should be noted that the explanations given with respect to the additive also apply to this (multiplicative).

For the smoothing constants, it should be noted that the *smoothing constants* determines how fast the weights of the series decays. The values may be chosen either subjectively (instinctively) or objectively (empirically). Values of a smoothing constant near one put almost all weight on the most recent observations. A value of a smoothing constant near zero allows the distant past observations to have a large influence. In all the equations for both additive and multiplicative methods, the term $\alpha$ is associated with the level of the series, $\beta$ is associated with the trend, and $\gamma$ is associated with the seasonality factors.

When selecting the smoothing constant subjectively, one uses his own experience with this, and similar, series. Also, specifying the smoothing constant one tunes the forecast to his beliefs about the future of the series. If one believes that the mechanism generating the series has recently gone through some fundamental changes, a smoothing constant value of 0.9 which will cause distant observations to be ignored can be used. If, however, one thinks the series is fairly stable and only going through random fluctuations, a value of 0.1 can be used.

To select the value of the smoothing constants objectively, one searches for values that are best in some sense. The program used here searches for the values that minimize the size of the combined forecast errors of the currently available series. Three measures of performance of the amount of error in the forecasts are in existence: the mean square error (MSE), the mean absolute error (MAE), and the mean absolute percent error (MAPE). The forecast error is the difference between the forecast of the current period made at the last period and the value of the series at the current period. This is written as

$$e_t = X_t - F_{t-1} \quad (8)$$

**The Support Vector Machines (SVM) model**

The role of Support Vector Machine (SVM) model in data mining during the last two decades cannot be over emphasized as it plays a substantial amount of research efforts in the application of time series forecasting. Data Mining is a pioneering and attractive research area due to its huge application areas. SVM is playing a decisive role as it provides techniques that are especially well suited to obtain results in an efficient way and with a good level of quality (Nayak et al., 2015). SVM is originally proposed by Vapnik, (1999) within the area of statistical learning theory and structural risk minimization, have demonstrated to work successfully on various classification and forecasting problems. Similarly, in the view of Nayak et al., (2015), SVM has the prospective to capture very large feature spaces, due to the generalization principle which is based on the Structural Risk Minimization. Support Vector Machine is a speedily increasing field with promise for greater applicability in all domain of research.

The SVM has been one of the more extensively used data learning tools in recent years. It is generally used to address a dualistic pattern classification problem. The dualistic SVM constructs a set of hyperplane in an infinite dimensional space which can then be divided into two kinds of representations, such as the linear and nonlinear SVM model. In order to construct an optimal hyperplane, SVM employs the iterative training process which is used to minimize an error function (Nayak et al., 2015). This can be classified into four groups according to the form of the error function. In explaining the procedure for the SVM model, we consider the separation of two sets of separable data points, given as: $(x_i, y_i)$, $i = 1, 2, 3, \ldots, N$ $y_i \in (+1, -1)$, $x_i \in R^d$, which is by a linear predictor. If we have some hyperplane that separate these two sets of points, the point $x_i$ which lie on the hyperplane satisfy $w^T x_i + \beta = 0$, where w is normal to the hyperplane. We then have

$w^T x_i + \beta \geq +1$ for $y_i = +1$,

$w^T x_i + \beta \leq -1$ for $y_i = -1$,

The margin can then be calculated as $2/\|w\|$, where $\|w\|$ is called the Euclidean norm of w. Maximizing $2/\|w\|$ is the same as Minimizing $2/\|w\|^2$. The problem is then formulated as:

Min $\frac{1}{2} w^T w$ by adding $\frac{1}{2}$ to the objective function for easy computation subject to the two constraints above. positive slack variable $\xi$ is further added to the cost C in the objective function and the adjusted model is according to Cortes & Vapnik, (1995) and Vapnik, (1999) referred to as soft margin SVM which is used to solve

$$\text{Min } \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \tag{9}$$

subject to $y_i \left( w^T x_i + \beta \right) \geq 1 - \xi_i$ ,

$\xi_i \geq 0$

For $i = 1, 2, \ldots, N$,

Where $x_i = (1, 2, \ldots N)$ are the N training points, $y_i$ is the label of each point with values +1or -1 and C is the penalty cost for the sample points which are not classified correctly by the SVM, however, a large C corresponds to a higher penalty to errors.

SVM models can be classified into four different groups as follows:

**a. Classification SVM**

i. Classification SVM Type1:- this is known as C-SVM classification

In this type of SVM, training involves minimization of the error function usually given by:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \tag{10}$$

Subject to the constraints: $y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i$ *and* $\xi_i \geq 0$, $i = 1 \ldots N$

Where $C$ is the capacity constant; $w$ is the vector of coefficients, $b$ a constant, $\xi_i$ is a parameter for handling data (inputs). $i$ is the index which label the $N$ training cases. $y \in \pm 1$ is the class labels $x_i$ is the independent variables. The kernel $\phi$ is used to transform data from the input (independent) to the feature space. It should be noted that the larger the $C$, the more the error is penalized. Thus, $C$ should be chosen with care to avoid over fitting.

ii. Classification SVM Type2:- this is known as nu-SVM classification

Similar to classification SVM type1, the classification SVM type2 model minimizes the error function:

$$\frac{1}{2}w^T w - v\rho + \frac{1}{N}\sum_{i=1}^{N}\xi_i \qquad (11)$$

Subject to the constraints $y_i\left(w^T\phi(x_i)+b\right) \geq \rho - \xi_i,\ \xi_i \geq 0,\quad i=1...N$ and $\rho \geq 0$

**b. Regression SVM**

In regression, SVM which is the functional dependence of the independent variable x has to be estimated. Its assumption is like other regression problems, which states that the relationship between the independent and dependent variables are given by a deterministic function $f$ in addition to some additive noise:

$$y = f(x) + noise \qquad (12)$$

The issue is now to find functional form for $f$ that can correctly predict new cases for SVM. This is achieved by training the SVM model on a sample set. This process is like classification and the sequential optimization of the error function. Two types of SVM models can be applicable depending on the definition of this error function. These are regression SVM type1 and regression SVM type 2. This continues as follows:

iii. Regression SVM Type1:- this is known as epsilon-SVM regression

This type of SVM uses the error function:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi_i^* \qquad (13)$$

This is minimized subject to: $\left(w^T\phi(x_i)+b\right) - y_i \leq \in + \xi_i;\quad y_i - \left(w^T\phi(x_i)+b_i\right) \leq \in + \xi_i^*$ and $\xi_i, \xi_i^* \geq 0,\ i=1,...,N$

iv. Regression SVM Type2:- this is known as nu-SVM regression

In this type of model, the error function is given by:

$$\frac{1}{2}w^T w - C\left[v \in + \frac{1}{N}\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right] \qquad (14)$$

This minimizes subject to: $\left(w^T\phi(x_i)+b\right) - y_i \leq \in + \xi_i;\quad y_i - \left(w^T\phi(x_i)+b_i\right) \leq \in + \xi_i^*$ and $\xi_i\ \xi_i^* \geq 0, i=1,...,N \in \geq 0$

Parameters that control the regression quality are the cost of error C, the width of the tube $\in$ and the mapping function $\phi$ . According to Wang, (2005) and Tan et al., (2004), intuitively, ε value should be small for large sample size than for small sample size and the SVM model value of ε in the ε- insensitive loss function is to be selected and it also affect the smoothness of the SVM response which affect the number of support vectors. It is useful to fix parameter ε by specifying the desired accuracy in advance (Hu, 2017).

## The Hybrid model

The hybrid model is a combination of exponential smoothing HW and SVM models. Such combination can potentially provide a better result based on the result of HW and SVM models. According to Li et al., (2011) and Zhang et al., (2015) most researches show that combining forecasts improve the accuracy than the individual forecasts and forecasts from a combined linear and nonlinear model can improve forecasting accuracy; therefore a hybrid method that involves linear and non-linear modeling can provide a good alternative for forecasting seasonal time series. Consequently, the hybrid model for linear and non-linear can improve the overall performance of the forecast (Li et al., 2010). Armstrong, (2001) in discussing the number of techniques to be considered in the combination with respect to efficiency limits it to five. He based his argument or suggestion on experimental behavior of the combination gains. The combination of up to a maximum of five forecasting models reduces the number of errors. However, when it is more than five techniques, the gains get diminishing.

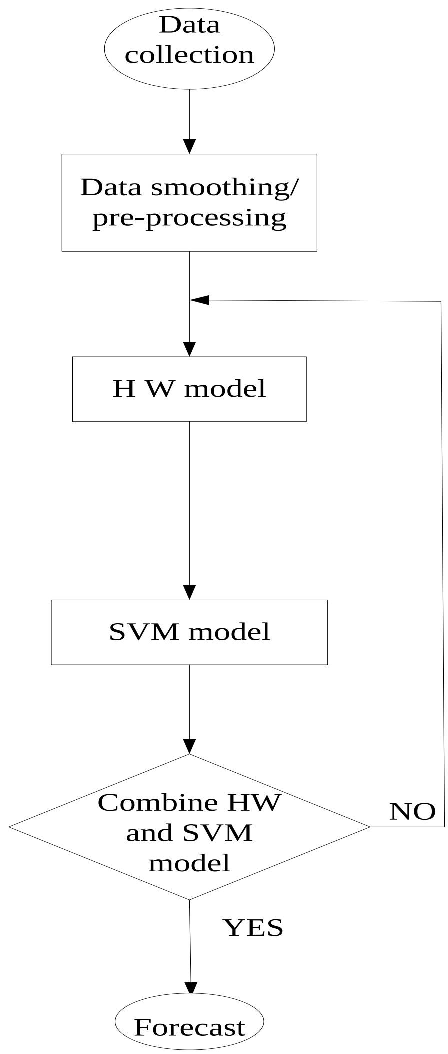


Figure1: Research Methodology Flowchart

Figure1 explains the combination of HW exponential smoothing model and SVM model to produce a hybrid model in which exponential smoothing model is used for linear characteristics of the airlines passengers' time series data, while SVM is used for nonlinear patterns of the time series data. In other to get more satisfactory forecasting results a hybrid system was developed when SVM is combined with exponential smoothing model. The study of our hybrid model is divided into three phases. The first phase of the study is the use of exponential smoothing model which has been discussed earlier. The second phase is the SVM, explained earlier also while the third phase is the hybrid system Zhang, (2003) of the time series data which will be obtained by feeding the predicted data of exponential smoothing model into the SVM model for forecasting time series.
For hybrid model, the input values are the predicted values of H-W model, however, the hybrid model is the same as SVM process. Some researchers believe that it may be reasonable to consider a time series comprising of a linear and nonlinear components so as to get the hybrid model of the two models (Zhang, 2003). The combined model is given by

$$X_t = L_t + N_t \qquad (15)$$

Where $L_t$ , is the linear component (exponential smoothing model) and $N_t$ is the nonlinear component (SVM) of the hybrid model. Both $L_t$ and $N_t$ have to be estimated from the data set. First the H-W model will model the linear component, and then the errors (residuals) from the linear model will contain only the nonlinear relationship. Let $\varepsilon_t$ denote the residual at time t from the linear model, then

$$\varepsilon_t = X_t - \hat{L}_t \qquad (16)$$

Where $\hat{L}_t$ represents the forecast value for time *t* from the estimated relationship. By modeling residuals (errors) using SVM nonlinear relationships can be discovered. Therefore, with n input, the SVM model for the residuals will be

$$\varepsilon_t = f(error\ from ES) + e_t, \qquad (17)$$

Where $f$ a nonlinear function determined by the SVM and $\varepsilon_t$ is the random error. Hence, the combined forecast will be

$$\hat{X} = \hat{L} + \hat{N} \qquad (18)$$

Where $\hat{L}_t$ the forecast value of HW is model at time t; $\hat{N}_t$ is the forecast from the forecasting result using SVM model; $\hat{Z}_t$ is the forecast of the combine model.
The result of equation (16) is then tested and compared using the performance evaluation measures. The best model is then selected out of the six inputs based on the smallest value of these measures. The hybrid model exploits the unique feature and strength of exponential smoothing model as well as SVM model in determining different patterns using different models as advantage of modeling the linear and nonlinear forms separately. In order to improve the overall modeling and forecasting performance, it will be better to combine the forecasts.

**Forecast evaluation methods**

Generally, the performance evaluation criteria are widely used in evaluating the results of time series forecasting as described in various literature, Woschnagg & Cipan, 2004). Hyndman, (2014); Shcherbakov & Brebels, (2013) and Propper & Wilson, (2003) refer to these errors as 'standards' statistical measures of the forecast accuracy are made up of MSE and MAE. The criteria for judging the best model are how relatively small these values are in both the training and testing of the data. This is necessary to quantify the amount by which an estimator differs from the original (true) value. That is why the measure with smaller values is usually selected as the best. These measures include, the Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Average Error (MAE), Mean Average Percentage Error (MAPE) and Correlation Coefficient (R). These measures are defined as follows: -

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{19}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|} \tag{20}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{21}$$

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\left[\sum_{i=1}^{n}(y_i - \bar{y}_i)^2\right]^{1/2}\left[\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}}_i)^2\right]^{1/2}} \tag{22}$$

where $\hat{y}_i$ are the predicted values at time *I*; $y_i$ are the actual values at time *i*, and *n* is the number of Predictions; $\bar{y}_i$ are the mean values at time $i$ ; $\bar{\hat{y}}_i$ are the predicted mean values at time $i$ and $R$ is the coefficient of correlation.

**Training and testing (forecasting) Airline passengers'**

We now obtain the analysis involving the actual data, the forecast and errors all from the HW. This error is made up of both positive and negative values, hence it is normalized using this relation :

$$y = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{23}$$

Where $y$ is the normalized value; $x_i$ is the error value of the series; $x_{\min}$ is the minimum value in the series; $x_{mix}$ is the maximum value in the series

The next step is to input the normalized values into the SVM using the same procedure as already outlined and then continue with the analysis that include the forecast from HW, SVM, the obtained SVM values are transformed using:

$$Z = x_{\min} + y*\left(x_{\min} - x_{\max}\right) \qquad (24)$$

The hybrid is now obtained by adding HW to the transformed data and the errors by taking the difference between the hybrid values and the actual data, after which the measures of performance are calculated. The software used for the prediction performance is the STATISTICA software.

**Results and Discussion**

Table1 showing the results of performance of Airline passengers using hybrid model

| *Input* | Measures | Training | Testing |
|---|---|---|---|
| *2* | MSE | 797.1574 | 2053.5460 |
| | MAE | 22.28007 | 36.013330 |
| | R | 0.964070 | 0.8344000 |
| *4* | MSE | 512.2239 | 1780.1720 |
| | MAE | 17.34570 | 31.717430 |
| | R | 0.976570 | 0.8419000 |
| *6* | MSE | 517.2903 | 1959.1890 |
| | MAE | 16.53990 | 32.448630 |
| | R | 0.976620 | 0.8268000 |
| *8* | MSE | 296.6476 | 1370.1390 |
| | MAE | 13.80679 | 30.961720 |
| | R | 0.986320 | 0.8767000 |
| *10* | MSE | 155.7703 | 664.69680 |
| | MAE | 9.883210 | 21.122600 |
| | R | 0.992650 | 0.9474000 |
| *12* | MSE | 84.70710 | **463.71770** |
| | MAE | 7.514370 | **17.147390** |
| | R | 0.995950 | **0.9577000** |

From table1, the best model for the forecast is input 12 with MSE = 463.7177, MAE = 17.14739 and R =0.9577

Table 2 showing the performance of HW, SVM and hybrid models for Airlines passengers'

| ***Models*** | **MSE** | **MAE** | **R** |
|---|---|---|---|
| *HW* | 996.10 | 44.63513 | 0.7502 |
| *SVM* | 925.9196 | 25.24209 | 0.8979 |
| ***Hybrid*** | **463.7177** | **17.14739** | **0.9577** |

The results in table 2 show that the best performance can be obtained by the hybrid model with the minimum values of MSE, MAE and maximum value of correlation coefficient, R, followed by SVM model and lastly HW model. The results obtained in this study also indicate that the hybrid model is a powerful tool as well as an alternative method of forecasting the time series for the airline passengers'.

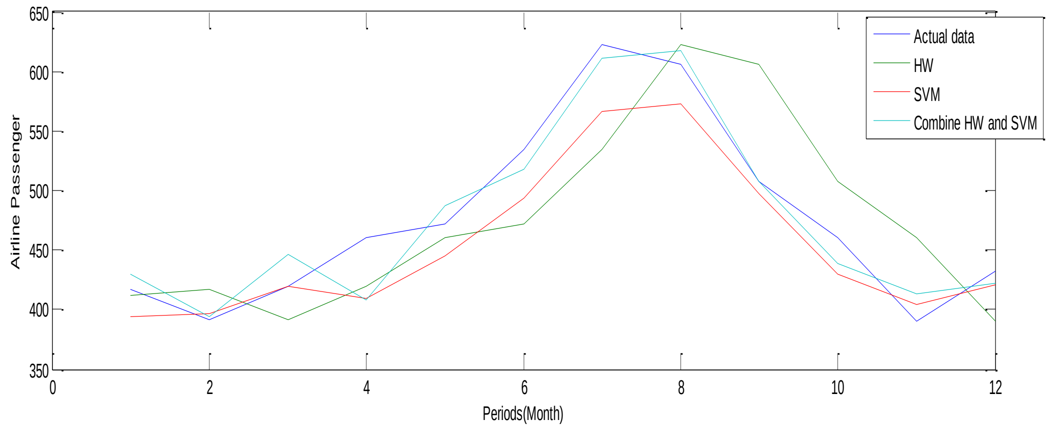


Figure 2: Time series plot of actual airline passengers data and the three models

Figure 2 shows the results from the three models HW, SVM and hybrid compared with the actual airline passengers' data for the 12 periods in the testing phase. HW shows that the values are not stable; SVM is better than the HW and hybrid has forecasted the values that are close to the actual data which make its forecast the best.

Table 3 showing MAPE improvement of the models (Airline passengers')

| *Models* | MAPE | Percentage improvement |
|---|---|---|
| *HW* | 9.3082 | 149.3 % |
| *SVM* | 5.0685 | 35.9 % |
| *Hybrid HW-SVM* | 3.7425 | - |

From Table 3, it can be observed that the combined model presented in this paper recorded 149.3% improvement over HW model and 35.9% improvement over the traditional SVM model.

## Conclusion

The study is meant to investigate the potentiality of the HW, SVM and hybrid methods in the airlines passengers' time series forecasting. The results of the hybrid model in the analysis indicate that it can be applied successfully to establish time series forecasting model which can provide accurate forecasting and modeling time series. Therefore, the application of HW, SVM and hybrid models in the field of forecasting time series data is studied and the accuracy of the forecasts in terms of MSE, MAE and R have been computed and compared among the three models. The results of comparison demonstrate that the hybrid model is the best when compared with HW and SVM models for the airlines passengers' data series used in this paper. From table 3 it can be seen that the hybrid model recorded 149.3% and 35.9% improvement over HW and SVM models respectively.

## References

Armstrong, J. S. (2001). *Selecting Forecasting Methods*. 365–386. https://doi.org/10.1007/978-0-306-47630-3_16

Bolarinwa, I. A., Bolarinwa, B. T., & Polytechnic, T. F. (2021). *UNDERSTANDING FEATURES OF TIME SERIES*. *7*(8), 15–19.

Boži, M., & Stojanovi, M. (2011). *Application of SVM Methods for Mid-Term Load Forecasting **. *8*(1), 73–83.

Chao, C., & Horng, M. (2015). *The Construction of Support Vector Machine Classifier Using the Firefly Algorithm*. *2015*.

Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. *297*, 273–297.

Costa, M., Gonçalves, A. M., Silva, J., Costa, M., Gonçalves, A. M., & Silva, J. (2015). *Forecasting Time Series Combining Holt-Winters and Bootstrap Approaches*. *110004*. https://doi.org/10.1063/1.4912411

Evgeniou, T., & Pontil, M. (2014). *Support Vector Machines : Theory and Applications WORKSHOP ON SUPPORT VECTOR MACHINES : THEORY AND APPLICATIONS*. *January 2001*. https://doi.org/10.1007/3-540-44673-7

Fried, R. (2009). *Exponential and Holt-Winters Smoothing*. *January 2015*. https://doi.org/10.1007/978-3-642-04898-2

Gelper, S., Fried, R., & Croux, C. (n.d.). *Computational aspects of robust holt-winters smoothing based on m -estimation*. *2001*, 1–13.

Hong, H., Pradhan, B., Bui, D. T., Xu, C., Youssef, A. M., & Chen, W. (2017). Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: a case study at Suichuan area (China). *Geomatics, Natural Hazards and Risk*. https://doi.org/10.1080/19475705.2016.1250112

Hsu, C., Chang, C., & Lin, C. (2016). *A Practical Guide to Support Vector Classification*. *1*(1), 1–16.

Hu, X. (2017). *Support Vector Machine and Its Application to Regression and Classification TO REGRESSION AND CLASSIFICATION*.

Hyndman, L. R. J. (2014). *Forecasting : Principles & Practice*. *September*.

Hyndman, R. J. (2008). *Automatic Time Series Forecasting : the forecast Package for R*. *27*(3). https://doi.org/10.18637/jss.v000.i00

Kalekar, P. S., & Bernard, P. (2004). *Time series Forecasting using Holt-Winters Exponential Smoothing Under the guidance of*. *04329008*, 1–13.

Khan, S. A., Hussain, I., Hussain, T., Faisal, M., Muhammad, Y. S., & Mohamd Shoukry, A. (2017). Regional Frequency Analysis of Extremes Precipitation Using L-Moments and Partial L-Moments. *Advances in Meteorology*, *2017*. https://doi.org/10.1155/2017/6954902

Kumar, K. S. (2016). *Performance Variation of Support Vector Machine and Probabilistic Neural Network in Classification of Cancer Datasets*. *11*(4), 2224–2234.

Li, S., Kuo, S., Cheng, Y., & Chen, C. (2010). Deterministic vector long-term forecasting for fuzzy time series. *Fuzzy Sets and Systems*, *161*(13), 1852–1870. https://doi.org/10.1016/j.fss.2009.10.028

Li, S., Kuo, S., Cheng, Y., & Chen, C. (2011). A vector forecasting model for fuzzy time series. *Applied Soft Computing Journal*, *11*(3), 3125–3134. https://doi.org/10.1016/j.asoc.2010.12.015

Liu, Z., Zhou, P., Chen, G., & Guo, L. (2014). Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. *Journal of Hydrology*, *519*(PD), 2822–2831. https://doi.org/10.1016/j.jhydrol.2014.06.050

Majid, R. (2018). Advances in Statistical Forecasting Methods: An Overview. *Economic Affairs*, *63*(4). https://doi.org/10.30954/0424-2513.4.2018.5

Makatjane, K. D., & Moroke, N. D. (2016). *COMPARATIVE STUDY OF HOLT-WINTERS TRIPLE EXPONENTIAL SMOOTHING AND SEASONAL ARIMA : FORECASTING SHORT TERM SEASONAL CAR SALES IN SOUTH AFRICA*. *6*(1), 71–82.

Nayak, J., Naik, B., & Behera, H. S. (2015). A comprehensive survey on support vector machine in data mining tasks: Applications & challenges. *International Journal of Database Theory and Application*, *8*(1), 169–186. https://doi.org/10.14257/ijdta.2015.8.1.18

Propper, C., & Wilson, D. (2003). The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy*, *19*(2), 250–267. https://doi.org/10.1093/oxrep/19.2.250

Shcherbakov, M. V., & Brebels, A. (2013). *A Survey of Forecast Error Measures*. *24*(4), 171–176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032

Tan, Y., Member, S., Wang, J., & Member, S. (2004). *A Support Vector Machine with a Hybrid Kernel and Minimal Vapnik-Chervonenkis Dimension*. *16*(4), 385–395.

Tian, Y., Shi, Y., & Liu, X. (2012). *Recent advances on support vector machines research*. *18*(1), 5–33. https://doi.org/10.3846/20294913.2012.661205

Vapnik, V. N. (1999). *An Overview of Statistical Learning Theory*. *10*(5), 988–999.

Wang, L. (2005). *Support Vector Machines : Theory and Applications*.

Wang, S., Feng, B., & Liu, G. (2013). Application of seasonal time series model in the precipitation forecast ☆. *Mathematical and Computer Modelling*, *58*(3–4), 677–683. https://doi.org/10.1016/j.mcm.2011.10.034

Wong, K. K. F., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management*, *28*(4), 1068–1078. https://doi.org/10.1016/j.tourman.2006.08.003

Woschnagg, E., & Cipan, J. (2004). *Evaluating Forecast Accuracy*. 17.

Xie, W., Yu, L., Xu, S., & Wang, S. (2006). *A New Method for Crude Oil Price Forecasting*. 444–451.

Zhang, G. P. (2003). *Time series forecasting using a hybrid ARIMA and neural network model*. *50*, 159–175.

Zhang, X., Peng, Y., Zhang, C., & Wang, B. (2015). Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting ? Some experiment evidences. *JOURNAL OF HYDROLOGY*, *530*, 137–152. https://doi.org/10.1016/j.jhydrol.2015.09.047