Hybrid Modeling of ARIMA, ANN, And SYM for Macro Variables Forecasting in Pakistan

Rizwan Ali, Hafsa Hina*, Amena Urooj

Department of Economics and Econometrics, Pakistan Institute of Development Economics, Islamabad, Pakistan.

ABSTRACT

Time series forecasting remains a challenging task owing to its nonlinear, complex, and chaotic behaviour. The purpose of the current paper was to analyze the forecast performance of different models to determine Pakistan's macroeconomic variables, such as inflation, exchange rate, and stock return. These models included Linear Autoregressive Integrated Moving Average (ARIMA) model as well as nonlinear models, such as Artificial Neural Networks (ANN), and Support Vector Machine (SVM). Afterwards, a hybrid methodology was used to combine the linear ARIMA with nonlinear models of ANN and SVM. The forecasting performance of all the models, that is, ARIMA, ANN, SVM, ARIMA-ANN, and ARIMA-SVM was compared on the basis of RMSE and MAE. The results indicated that the best forecasting model to achieve high forecast accuracy was the hybrid ARIMA-SVM.

Keywords: Autoregressive Integrated Moving Average (ARIMA), Artificial Neutral Networks (ANN), hybrid models, Support Vector Machine (SVM), time series forecasting

JEL Classification: C22, C53

Introduction

Time series data does not seem very simple as it is subjected to many complexities, instabilities, and nonlinearities which emerge due to dynamic behavior of economy. Due to complexity, instability, and nonlinearity in most of time series data, it is not easy to make perfect and accurate predictions by using different conventional and advanced methods. Therefore, accurate and perfect future forecasting could not be accomplished by using different conventional and advanced methods.

* Corresponding Author: hafsahina@pide.org.pk

However, forecasting is crucial, since it also helps policy makers to take effective measures and plan better for the development of better future.

Autoregressive Integrated Moving Average (ARIMA) model has been used since long for time series forecasting to provide reliable results. It is well accepted owing to its statistical properties as well as popular Box-Jenkins (1976) methodology however, it has limitations in nonlinear data set handling (Zhu & Wei, 2013). Presently, novel neural network techniques have been used widely for nonlinear data sets because of their feature of fast learning and pattern recognizing for complex data sets. Artificial Neural Networks (ANN) is a famous and well accepted technique due to its flexibility while using nonlinear data sets and to provide reliable results for financial time series. However, it also has disadvantages of over fitting and not having much understanding regarding the data. Support Vector Machines (SVMs) is considered as a major breakthrough in machine learning and is widely applied to classification and regression analysis. It is high performing algorithm with little tuning. Most of the studies have shown that a hybrid model of different linear and nonlinear models has powerful tools to advance the forecast performance of each individual model (Khashei & Bijari, 2011).

Linear models, such as Ordinary Least Squares (OLS), Maximum Likelihood Estimation (MLE), Generalized Linear Model (GLM), Cholesky Decomposition, Analysis of Variance (ANOVA) methods and ARIMA are used to analyze the linear dependency among the dependent and independent variables. There are many limitations for these models in case of nonlinearity and complex data set. Often these models are used for nonlinear data set by making different prior assumptions and taking data transformation. However, in real life, data is not normal as assumed, therefore, it may lead to selection of the model which could be inappropriate and does not replicate the factual shape of the data set. However, linear models are preferred due to their simplicity, requirement of small memory space, and speed of convergence in comparisons with nonlinear models which usually have low convergence (Petruseva et al., 2017). When the influence of the nonlinearity to the overall specification of the model is very small, one cannot assume a nonlinear model to perform better than a linear model

Nonlinear models, such as SVM, ANN, K-Nearest Neighbors, Decision Trees like CART, Random Forest, and Naive Bayes take much attention for their importance to tackle the nonlinear and complex data set issues. Therefore, it is difficult to decide whether linear or nonlinear models are best, that's why both are chosen for forecasting and would make their hybrid for the sake of fruitful results.

In the current paper, forecasting performance of linear ARIMA, nonlinear models ANN, SVM, and hybrid ARIMA-ANN and ARIMA-SVM would be compared. An appropriate combination of linear and nonlinear models yields a more precise prediction than any separate linear and nonlinear models for forecasting time series data (Babu et al., 2014).

Literature review

Several studies have been conducted to examine the forecasting time series data arising from various fields. In literature different models such as linear, nonlinear, and hybrid models are suggested based on their forecasting performance. In this section, existing literature on ARIMA, ANN, SVM, and hybrid modeling along with their applications in different fields has been reviewed.

ARIMA models are theoretically more valid and could be unexpectedly robust as compared to the alternative modeling approaches. Kenny et al. (1998) forecasted Irish inflation using ARIMA model and suggested that there should be more focus to minimize out of sample forecasts errors rather than minimizing in sample 'goodness of fit'. Sultana et al. (2014) forecasted inflation and economic growth for Pakistan by using ARIMA and decomposition methodology on monthly series. They compared out of sample forecasts of both time series methods based on mean absolute deviation and sum of square of errors in which ARIMA gave better forecasts performance. Farooqi (2014) also employed ARIMA model to forecast the future annual values of imports and exports of Pakistan. Standard statistical techniques, such as AIC and diagnostic check were used to determine the validity of the fitted model.

Neural networks prove to be best as compared to their traditional counterparts. These models are able to capture more essential non linearities among financial variables and real output growth at longer horizons however, perform poor for 1-quarter forecasts (Tkacz & Hu, 1999). Neural networks are very robust to exploit the nonlinear relationships that exist between variables to provide the more precise forecasts of economic activities. Burney et al. (2005) forecasted Karachi stock exchange shares by ANN using pre-processed data. Levenberg Marquardt algorithm was used through which weights were adjusted during the back error propagation. It proved fairly accurate forecast when compared with weighted exponential method and ARIMA model.

Haider and Hanif (2009) forecasted inflation on the basis of monthly data for Pakistan through ANN and concluded that ANN was more precise when compared with ARIMA and AR (1) on the basis of out-of-sample forecast. They built the architecture of 'feed forward with back propagation' by using standard Levenberg Marquadt algorithm. To reduce the error volatility 12 hidden layers were trained and model learning rate was kept 0.25.They evaluated the forecast by calculating RMSE. Awan et al. (2012) proposed a hybrid non-linear Autoregressive Exogenous model (NARX) established with feed Forward Network (FFNN), SVR, and Neural Network Models. They forecasted the long term industrial load and compared all three techniques based on MAPE. All of the three models showed accuracy with respect to results with acceptable MAPE in which proposed hybrid model NARX based FFNN and which remained more attractive in forecasting.

Cortes and Vapnik (1995) extended the SVMs to nonlinear regression problems by introducing the idea of mapping input vectors into higher dimensional space. Linear decision surface was constructed in this space which resulted in higher generalization ability. SVMs used structural risk minimization instead of empirical risk minimization which reduced the generalization error by minimizing its upper bound, resulting in better generalization as compared to the conventional techniques. Cao and Tay (2001) evaluated the SVR as a promising alternative for time series forecasting. They noted that SVMs were better when compared with multilayer perceptron trained with back propagation (BP) based on the criteria of normalized root mean square error (NRMSE) and mean absolute error (MAE). SVMs were faster and had small number of parameters as compared to BP. Calveria et al. (2015) examined the regional forecasting for tourism using Support Vector Regression (SVR) with three different kernels and two ANN models of Radial Basis Function (RBF) and multilayer perceptron. SVR with Gaussian kernel outperformed other models of AAN and SVR with linear and polynomial kernels. They concluded that the choice of kernel is important in SVR for better results and machine learning techniques are better suitable for long term forecasts.

The hybrid methodology of ARIMA and ANN could be a productive way to improve forecasting performance by their unique feature of linearity and nonlinearity. Zhang (2001) proposed the hybrid methodology by estimating first linear part using ARIMA and then its residuals through ANN, and compared the results based on mean square error (MSE). He inferred that the hybrid model may perform worse than ANN and ARIMA in some data points however, its overall performance was better. Pai and Lin (2005) stated hybrid methodology of ARIMA whereas, SVM showed satisfactory results in stock price forecasting. They used the real data sets of ten stocks in order to inspect the accuracy of proposed methods on the basis of four statistical indices. These indices included Mean Absolute Error (MAE), MSE, MAPE, and RMSE to measure the performance of the proposed model. Chen and Wang (2007) also proposed hybrid model of SARIMA and SVM and concluded that hybrid methods gave reliable results as compared to individual methods. They forecasted the Taiwan's machine industry production values and investigated the proposed methodology on the basis of Normalized Mean Square Error (NMSE) and Mean Absolute Percentage Error (MAPE).

Many authors have combined statistical models and suggested that hybrid statistical modeling improves predictive performance as compared to stand-alones (Stone, 1974; Breiman, 1996; Leblence & Tibshirani, 1996; Mojirsheibani, 1999; Khashei & Bijari, 2010). Papatla et al. (2012) presented two classes of hybrid models, that is, linear and nonlinear and suggested that mixed NN had higher probability in performance. Kumar (2014) made hybrid model of SVM, ANN, and Random Forest with ARIMA and compared all the models with each other along with their individual forms, in which SVM-ARIMA hybrid model took first place. The results were measured on the basis of RMSE, MAE, and NMSE based criteria to identify the best hybrid model. Among others ANN-ARIMA performed better to forecast the stock index returns. It was suggested that ARIMA- SVM could be helpful to improve the forecasting and profit-making returns were assured for policy makers to forecast economic and financial data.

Al-hnaity and Abbod (2016) explored the predictability of stock index time series. They stated that a single classical model would not yield the accurate prediction results. In their study, they used the most reliable data pertaining to mining techniques such as SVR, SVM, and BPNN. It was concluded that the proposed hybrid model outperformed all other single models along with the bench mark traditional model AR based on MSE, RMSE, and MAE. Karathanasopoulos and Osman (2019) also proposed a hybrid model to forecast the Dubai Financial general index by combining the momentum effect with a novel methodology of deep belief networks. They compared the empirical results based on MAE, MSE, MAPE, and RMSE with other three linear models named as moving average convergence divergence, naïve strategy, and ARMA model. They concluded that the proposed hybrid model outperformed all other models significantly and provided auspicious results for further usage in financial forecasts.

Methodology and Data Description

Autoregressive Integrated Moving Average (ARIMA) Modeling

Box and Jenkins (1976) introduced the ARIMA model as one of the most popular approaches to forecast time series data. In an ARIMA model, the future value of a variable is assumed to be a linear combination of several past values and past errors. Mathematically, it could be written as:

$$y_t = \sigma + \varphi_1 y_{t-1}, + \varphi_2 y_{t-2}, + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1}, - \theta_2 \varepsilon_{t-2}, - \cdots - \theta_q \varepsilon_{t-q} \tag{3.1}$$

Where y_t represents the series at time t, ε_t represents the error terms at time t, φ, θ and p are the coefficients, and q represents the lag lengths of AR and MA respectively. Box-Jenkins gave a systematic procedure to choose ARIMA model which involved iterative steps such as identification, estimation, and diagnostic checks on model adequacy. Under ARIMA modeling, it is necessary to have a stationary series. The Augmented dickeyfuller (1981) is used for the identification of order of integration in time series.

If time series has a seasonal component in case of seasonal fluctuations , the model would become seasonal ARIMA (SARIMA). Mathematically the general expression of the SARIMA model could be inscribed as follows:

$$\phi_p(B)\Phi_P(B)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{3.2}$$

Where P and Q are seasonal lag lengths, D is the order of seasonal differencing and s is the number of seasons per year.

Artificial Neural Networks (ANN) Methodology

A computational model for neural networks based on mathematics and algorithms known as 'the threshold logic' has been created by McCulloch and Pitts (1943). Connectionist systems in ANN are referred to as computing systems that are motivated by biological neurons works which resembles human brain, such as to learn by experience and then analyze it. Such systems "learn" (that is, gradually upgrade performance on) tasks by evaluating examples, usually without task specific programming.

ANN is established on a number of attached units or nodes termed as artificial nodes. Each connected unit between artificial nodes can transfer a signal from one to another. Commonly, ANN executes the signals at a connection among artificial neurons and the output of each artificial neuron is computed by a nonlinear function of the summation of its inputs. ANN nodes and units consist of layers. Usually three layer perceptron is enough to train a model, however, it could be changed depending on the structure of data set. ANN performs in a very analogous way as it takes several inputs, processes it in and out of multiple neurons from multiple hidden layers and yields the results using an output layer. This outcome estimation procedure is technically known as "Forward propagation". The basic objective of ANN is to solve problems in the same way as human mind does and to conduct different tasks more quickly in a better way than the traditional systems.

Artificial neurons and networks generally have weights and biases that are modified as learning procedures. The target is to generate the output to neural networks as close as possible to get the desired output. The mathematical association among the output (y_t) and the inputs $(y_{t-1}, \dots, y_{t-p})$ has the following representation:

$$y_t = w_0 + \sum_{j=1}^{q} w_j . g \left(w_{0.j} + \sum_{i=1}^{p} w_{i.j} . y_{t-p} \right) + e_t \qquad (3.3)$$

Where $w_{i.j} = (i = 0,1,2, \dots, p\, ; j = 1,2,3, \dots, q)$ and $w_j = (j = 0,1,2, \dots, q)$ are model constraints known as neuron weights, p represents the input nodes and q is the number of hidden nodes. The different activation functions have been used to transmit an input signal of a neuron to an output signal. The values of neurons which contribute more to the errors are minimized and this happens while moving back to the neurons of the neural network and detecting where the error lies. This process is known as “back propagation”. The neural networks use a common algorithm known as “Gradient Descent” in order to reduce the numbers of iterations while minimizing the error. This helps the task to be executed quickly and efficiently. One round of onward and back propagation iteration is known as, one training iteration also known as Epoch.

Selection of Input Parameters

The most decisive part of the ANN modeling is the selection of input lags because these are helpful to determine (non-linear) autocorrelation form of the time series. Neural networks have no hard and fast rule for parameter selection. Therefore, determining the number of input lags p and number of hidden nodes q is data dependent. There exist many approaches for parameter selection in ANN modeling but none of these techniques assure to perform best for actual time series analysis (Khashei & Bijari, 2010). Therefore, many experiments have been conducted to select the suitable p and q parameters that minimize the general criteria for accuracy, such as mean square error (Zhang, 2001). Once an appropriate ANN parameter is selected for a data set, the model is ready for final forecast analysis.

Estimation of Parameters and Activation Function

Resilient Back Propagation (RPROP) is a well-known gradient descent algorithm and is used to train a neural network. RPROP only uses the sign of gradients to compute the parameter updates. If the sign of parameter remains in the same direction for several iterations then the step size of update value would be increased, otherwise in case of oscillation it would be minimized. The main advantage of RPROP over standard back propagation is that it does not need any free parameter value like learning rate and momentum term and is much faster. RPROP also has separate step size for each weight. Thus, even if one weight is much closer to optimal value while the other needs more updates, it does not create any issue like other gradient descent variants cause problem in such scenarios. However, the drawback of RPROP is that it is more complex to be applied than Standard Back Propagation. Activation function sums the multiplication of inputs and their respective weights and then feeds it as input to the next layer of neurons through transfer function. The sigmoid or logistic function is used as transfer function and typically it could be written as:

$$Sig(x) = \frac{1}{1 + e^{-x}} \tag{3.4}$$

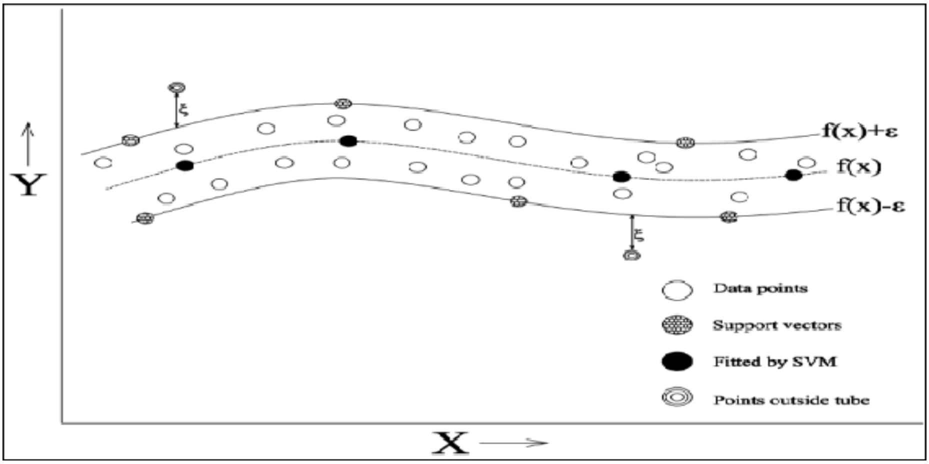
Support Vector Machine (SVM)

Support Vector Machine (SVM) was developed in 1990's by Vapnic *et al.* SVM is basically a classification method that executes classification tasks by formulating hyperplane in a multidimensional space that splits the cases of unalike class labels. SVM works by transforming the data to a higher dimensional space, and then performing separation in the resulting hyperplane. In SVM a line is constructed that best separates the points to their class 0 or 1 in input space variable. The best or optimal hyperplane is the line which separates the two classes by largest margin. Margin is calculated as the vertical distance from line to the closest points only. These points are important to explain the line and to construct the classifier. These are called support vectors and they define or support the hyperplane.

The Ɛ -insensitive SVR has been used for forecasting. In Ɛ -insensitive SVR, the main goal is to find a function f(x) that has an Ɛ -deviation from the actually acquired target y_i for whole training data and simultaneously as flat as possible. The sketch of SVR has been presented below in Figure 1 as:

Figure 1

Detailed Epsilon Tube with Slack Variables



Source : Kuntoji et al. (2017)

Assume f(x) takes the following form as under taken by Smola and Scholkopf (2004):

$$f(x) = wx + b \qquad w \in x, b \in R \qquad (3.5)$$

Here w is the regularizing term which provides optimization problem central over the flatness of the solution and b is the bias. This problem has been solved as follows:

$$\min \frac{1}{2} ||w||^2 \qquad (3.5.1)$$

Subject to,

$$y_i - wx_i - b \leq \mathcal{E} \qquad (3.5.2)$$
$$wx_i + b - y_i \leq \mathcal{E} \qquad (3.5.3)$$

In such a case where the constraints are infeasible, to protect against outliers, and to discern how many points could be tolerated outside the tube. One can insert slack variables $\xi_{i\,,}\xi^*$ in this case, called soft margin formulations according to Cortes & Vapnik (1995) and is explained by the following problem (also see Figure 1 for slack variables which lie outsides the $\mathcal{E}$ -insensitive tube). After solving the optimization as,

$$f(x) = \sum_{i=1}^{n} (\lambda_i - \lambda_i^*) K(x_{i,} x) + b \qquad (3.6)$$

Where K represents the kernel function, any symmetric semi-definite function, which fulfills the Mercer's conditions, can be used as a kernel function in the SVMs situation (Cortes & Vapnik, 1995). Different kernel functions may be used to achieve better generalization for a specific problem. However, one cannot say that a specific kernel outperforms others. Therefore, some validation techniques could be used to fix good kernel. Commonly used kernel functions include polynomial, Gaussian, and RBF.

The Selection of Input Lags and Hyperparameters for SVM model

In literature many authors have selected different input lags depending on the nature and frequency of the data. Twelve lags were used for the analysis of the current research as it made sense because of monthly data and would be appropriate for forecasting. The selection of hyper parameters plays a crucial role in any analysis of SVM which is sort of kernel function, regularization constant C, and the maximum allowable loss function ε. Many authors (Kumar, 2014; Cao & Tay, 2003; Chen & Wang, 2007) found that Gaussian radial basis function is superior to other kernel types as they take more time to train the model and gives adverse results as compared to Gaussian radial basis function. Therefore, Gaussian radial basis function was used because of its better prediction performance.

In the tuning of SVM model for parameter selection, kernel parameter γ and regularization parameter C plays an important role pertaining to model performance. The improper selection of these parameters could lead to under and over fitting of the trained model, that is, too large value of constant C could cause over fitting of the train data, while small values could under fit and vice versa for gamma parameter. Therefore, the combination of optimal C and γ parameters were chosen through tuning the SVM model which tested several different values and yielded the ones which had minimum error for 10-fold cross validation. The 10-fold cross validation means splitting the data randomly into ten equal parts in which each fold was used as a testing set at some point. In order to attain optimal parameters, a suitable range of parameter C and γ was provided to training data while tuning of the model which chose the best parameter values based on cross validation. A reasonable value for ε was insensitive to SVM modeling (Kumar, 2014).

Hybrid Model

The inspiration to mix the models comes from the consideration that a single model may not be enough to identify the true data, generating process, or cannot capture the entire features of the time series (Khashei & bijari, 2010). Therefore, the idea behind hybrid methodology was to acquire advantages of individual models and best traits to gain best possible results (Mojirsheibani, 1999; Paptla et al., 2002).

Let's assume that we have a series y_t consisting of linear and nonlinear components of L_t and N_t given as:

$$y_t = L_t + N_t \tag{3.7}$$

ARIMA model is first used on given data sets to obtain linear forecasts. Then by using the residuals acquired from the ARIMA, models were estimated through ANN. After combining the resulted output from ARIMA and ANN, the final required output was yielded. The same procedure was repeated for ARIMA and SVM. If e_t denotes residuals from the linear model at time t then,

$$e_t = y_t - \hat{L}_t \tag{3.8}$$

Where $\hat{L}_t$ represents the forecasting value by ARIMA model at time t, and then the residuals of ARIMA are estimated through SVM and ANN as follows:

$$e_t = f(e_{t-1}, e_{t-2}, \dots \dots\ e_{t-n}) + \mathcal{E}_t \tag{3.8.1}$$

Where f is a non-linear function computed by ANN and SVM and $\mathcal{E}_t$ is random error. The resulted combined forecast is given below as:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{3.9}$$

Where $\hat{N}_t$ is the forecast value of equation (3.8.1).

Data Description

For empirical analysis stock return, exchange rate and inflation data set was used. The detailed description of all the variables along with their sources are given in Table 1.

Table 1

Data Description and Source

Variables	Definition	Frequency	Time period	Source
Stock return	$S_t = logp_t - logp_{t-1}$	Monthly	1994-2018	Business Recorder & KSE
Exchange rate	The rate at which currency of two countries can be exchanged. $E_t = \frac{Rs}{Dollar}$	Monthly	1990-2018	SBP Monthly Statistical Bulletin
Inflation	$\pi_t = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}}$	Monthly	1990-2018	SBP Monthly Statistical Bulletin

Results and Discussion

Before estimations, the dataset was divided into two parts. The last 24 observations of total data points were characterized as the test set, while remaining as training set, so as to compare out of sample forecasts with the actual values. Statistical description of all variables is discussed in Table 2.

Table 2

Descriptive Statistics

Data	Mean	Std. Dev	Skewness	Kurtosis	JB	Probability	Obs.
π_t	0.007	0.007	0.441	0.415	13.9	0.001	343
E_t	4.408	0.495	-0.386	-0.933	21.05	0.001	348
ΔE_t	0.005	0.013	1.825	5.932	712	0.001	347
S_t	0.009	0.084	-0.972	4.697	323.1	0.001	294

In Table 2, JB (Jarque Bera) test indicates that none of the data set was normal. The series that had positive values of skewness, its mean data was skewed to right and kurtosis was less than 3 which implied that the data distribution was platykurtic. The series with negative skewness value depicted non normal distribution of data that was left skewed and kurtos was greater than 3, which was said to be leptokurtic or fat tailed. ADF statistic was implemented on the series to get the statistical counter part of stationary and non-stationary.

Table 3

ADF Test Statistics for all Series

Series	Deterministic part	Lags	T_{cal}	Integration order
π_t	None	6	-5.38	$I(0)$
E_t	Drift	7	-2.24	$I(1)$
ΔE_t	None	7	-5.42	$I(0)$
S_t	None	6	-6.08	$I(0)$

According to Table 3, the calculated values for inflation and stock return at level were less than critical values, so the null hypothesis was rejected and it was concluded that the series had no unit root at 5% significance level. However, in case of exchange rate it had unit root at level because its calculated value was greater than its critical value but after taking the 1st difference it became stationary.

Analysis of Series through ARIMA Modeling

The analysis of ARIMA modeling was proceeded by following the iterative steps of Box and Jenkins methodology accordingly. The statistics of the estimated models for all series are given below in Table 4.

Table 4

ARIMA Model for all data series

Series	Lags	Coefficients	Standard errors	t-statistics
Inflation	AR(1)	0.584	0.142	4.112
	AR(3)	0.170	0.077	2.208
	AR(9)	0.113	0.047	2.404
	MA(1)	-0.399	0.160	-2.493
	SAR(1)	0.967	0.023	42.043
	SMA(1)	-1.008	0.065	-15.508
	SMA(2)	0.205	0.068	3.015
	σ2 = 3.737e-05, Log Likelihood = 1167.14, AIC = -2318.27			
Return on exchange rate	Intercept	0.005	0.001	4.08
	AR(1)	0.422	0.049	8.67
	AR(11)	-0.348	0.126	-2.78
	MA(11)	0.478	0.131	3.65
	SMA(1)	0.212	0.070	3.02
	σ^2 = 0.00013: log likelihood = 983.85, AIC = -1955.69			

Stock return	Intercept	0.011	0.005	2.46
	AR(15)	-0.137	0.06	-2.82
	σ^2 = 0.0074: log likelihood = 279.02, AIC =-552			

Where sigma^2 is the value of constant variance assumed by the model, while the value of log likelihood shows the probability to derive the data. AIC was used to select the best possible model among others. After diagnosing, 24 months ahead forecasts were being calculated for all series by using their ARIMA model.

Analysis of all Series through ANN Modeling

The ANN architecture for each in-sample training set for all the series was selected after a careful consideration with minimum MSE by following steps in section 3.2. The description of the trained ANN architectures with 100 repetitions for all series is given below in Table 5.

Table 5

Multilayer Perceptron inputs for all data series

Series	Inflation	Return on exchange rate	Stock return
Neural network model	21-3-1	10-3-1	2-1-1

1 to 24 ahead forecasts have been made for all series from selected ANN architectures.

SVM Modeling on all data series

The optimum selection of parameters was made by tuning the SVM model by providing a range of values to parameters based on 10 fold cross validation. Afterwards, that model was selected with the best performance based on minimum training mean square error (MSE) is selected.

Table 6

Performance of SVM model for all data series and resulted parameters

Series	Inflation	Return on exchange rate	Stock return
parameters	c =5, γ=0.04, ε=0.1	c =10, γ=0.17, ε=0.1	c =10, γ=0.2, ε=0.1

After the selection of best suited model 24 months ahead forecasts were made.

Analysis of all Series by Hybrid Models

Analysis of all Series through Hybrid ARIMA-ANN

Hybrid model is a sum of linear forecasts from ARIMA and forecasts of its residuals from ANN. The ANN architecture of residuals of ARIMA for all data sets is presented in Table 7.

Table 7

Multilayer perceptron inputs for residuals of all data series

Series	Inflation	Return on exchange rate	Stock return
Neural network model	12-7-1	12-5-1	8-4-1

Where Table 4.6 represents the inputs and hidden nodes used to forecast the residuals from ARIMA. Forecasts of residual were obtained from fitted ANN model, now combining the both forecasts from ARIMA and ANN which would have equation as:

$$\hat{y}_t = \hat{L}_{t,ARIMA} + \hat{N}_{t,ANN} \tag{4.1}$$

The combined forecasts values of all series from 1 to 24 steps ahead were obtained from hybrid ARIMA-ANN model.

Analysis of all Series through Hybrid ARIMA-SVM

The forecasts of ARIMA and its residuals forecasts from SVM are to be combined to obtain hybrid forecasts. Mathematically, it could be presented as:

$$\hat{y}_t = \hat{L}_{t,ARIMA} + \widehat{N}_{t,SVM} \qquad (4.2)$$

Where $\hat{y}_t$ is the forecasted value at time t from hybrid ARIMAPSVM model, $\hat{L}_{t,ARIMA}$ represents the linear forecasts from ARIMA, and $\widehat{N}_{t,SVM}$ represents the nonlinear forecasts of residuals of ARIMA by SVM model.

SVM Modeling on residuals of all data series

Table 8: Performance of SVM model for residuals of all data series and resulted parameters

Series	Inflation	Return on exchange rate	Stock return
parameters	c =5, γ=5, ε=0.1	c =10, γ=10, ε=0.1	c =10, γ=1, ε=0.1

After the selection of best suited ARIMA-SVM model, 24 months ahead forecasts were made to analyze the preciseness of forecasted values against actual values. A graphical comparison was made as well as loss functions, such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used. The statistical equations for used loss functions are given as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}} \qquad (4.3)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \qquad (4.4)$$

Forecasts performance comparison of all used models

In this section performance of all models, based on RMSE and MAE is discussed systematically. The graphs signify how far a model succeeded to capture the right trend or direction of a data set. The Table 9 contains the used loss functions values for inflation series based on test data from all the models:

Table 9

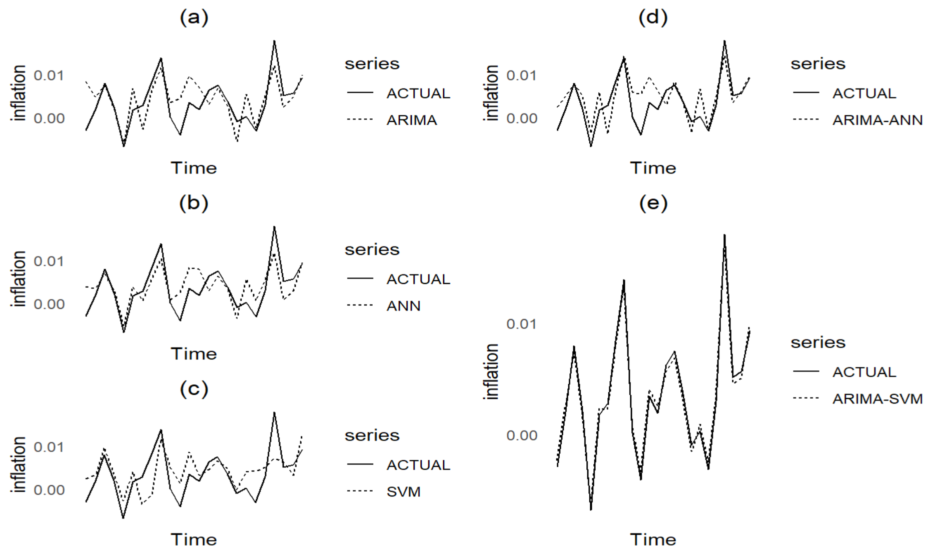
Loss errors of all models for inflation series

Models	ARIMA	ANN	SVM	ARIMA-ANN	ARIMA-SVM
RMSE	0.004367	0.003654	0.004495	0.003977	0.000567
MAE	0.003363	0.003038	0.003603	0.003143	0.000562

Table 9 shows ARIMA-SVM hybrid model which performed excellent as compared to other models based on RMSE and MAE. ANN has minimum RMSE and MAE after ARIMA-SVM model, while the hybrid model of ARIMA-ANN remained at 3rd position. ARIMA shows that it has better performance than SVM based on used loss functions. The graphical representation of inflation series is given in Figure 2.

Figure 2

Actual and Forecast Values of Inflation for All Models



In Figure 2, solid lines indicate the actual values and dashed lines represent the forecast values for respective models. It can be seen that all models seemed to be good to capture the direction of the actual values of solid line. However, the ARIMA-SVM Hybrid model outperformed all other models as it overlapped the actual values. Additionally, the hybrid model for ARIMA-ANN proved to be good to catch the trend. After these hybrid models, comes the performance of SVM followed by ANN, whereas ARIMA remained at last position. One important thing to be noted is that the ANN and ARIMA both performed poorly in catching the real direction of actual values in start as compared to the SVM model however had lower loss errors than SVM.

Next series to evaluate the forecast performance evaluation of various models is return on exchange rate. The Table 10 for its loss functions is given below as:

Table 10

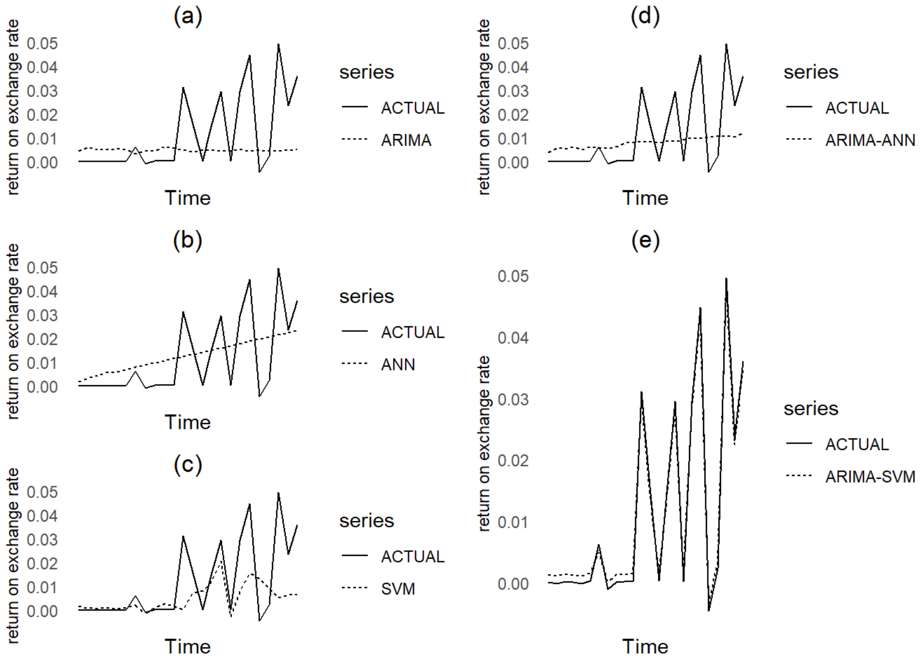
Loss Errors of All Models for Exchange Rate Series

Models	ARIMA	ANN	SVM	ARIMA-ANN	ARIMA-SVM
RMSE	0.017707	0.013309	0.015868	0.015421	0.001221
MAE	0.012744	0.010720	0.010139	0.011975	0.001221

The table above shows that ARIMA-SVM hybrid model had lowest loss errors as compared to other models and ANN stood at second place because it had comparatively least RMSE and MAE compared to other models. The hybrid ARIMA-ANN and SVM relatively had mixed results as one performed better in RMSE, while the other had good performance in case of MAE. However, ARIMA had the highest loss errors as compared to all other models. It could be concluded from the Table 4.10 that a suitable hybrid model is able to provide better forecasts based on RMSE and MAE for return on exchange rate series than most of linear and nonlinear models. The next graphical picture of return on exchange rate by all used models is given below in Figure 3.

Figure 3

Actual Versus Forecast Values of Return on Exchange Rate for all Models



The hybrid model of ARIMA-SVM performed very well in capturing the real direction of return on exchange rate. The dashed line throughout remained close to actual values of solid line as can be seen in part (e) of the Figure. After this, single model of SVM could be seen in part (c) of the Figure which captured the actual direction of real values to some extent. However, all other models performed very poor in catching the real trend of return on exchange rate. Part (a) represents straight dashed line of ARIMA, while the other models like ANN and hybrid ARIMA-ANN in part (b) and (d) respectively had the upward straight line. One more thing could be noted that the SVM model comparatively remained good in capturing the real trend of return on exchange rate than the other models that had lower loss errors as compared to SVM except hybrid ARIMA-SVM.

The different models that forecast performance for stock return series are discussed as following. The Table 11 had the loss functions of all models for stock return:

Table 11

Loss Errors of All Models for Stock Return Series

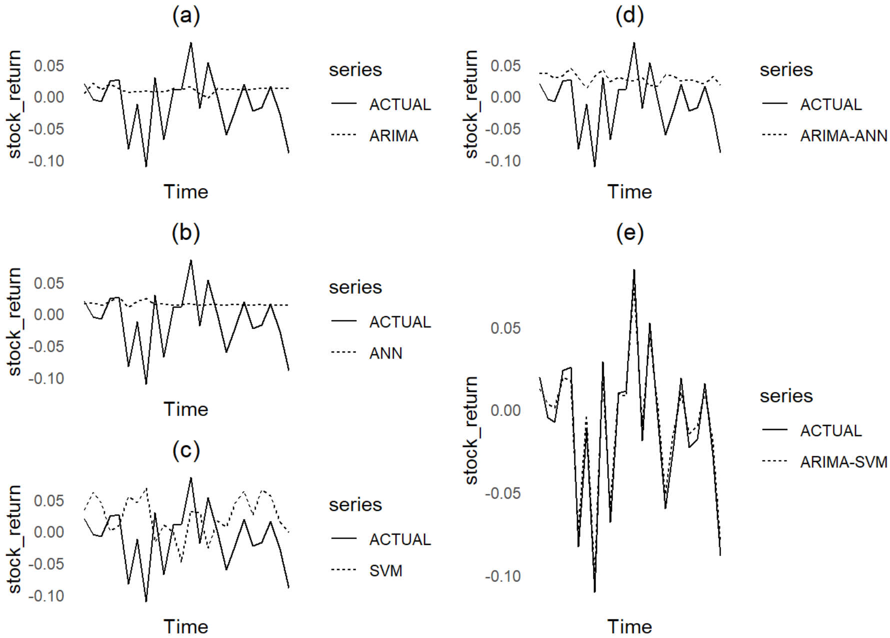
Models	ARIMA	ANN	SVM	ARIMA-ANN	ARIMA-SVM
RMSE	0.049153	0.051883	0.069622	0.059136	0.007787
MAE	0.036776	0.374887	0.058995	0.046278	0.007502

It can be seen from above Table 4.26 that ARIMA-SVM hybrid model showed minimum loss errors which represented that it performed better than others. The RMSE and MAE of ARIMA was lowest after hybrid ARIMA-SVM, however from Figure 4 part (a), it could be seen that ARIMA performed worse in capturing the direction of return than any other. While in case of SVM model, it showed higher loss errors when compared to other models except it superseded ANN in case of MAE but performed better after ARIMA-SVM hybrid in capturing the direction of stock return. Graphical representation of actual versus forecast values of stock return for all models is given in Figure 4.

Figure 4 further indicates that the hybrid model of ARIMA-SVM also outperformed other models in this data series as the dashed line of ARIMA-SVM in part (e) traced the actual values of the stock return. All other models had poor performance while capturing the direction of return except SVM model which performed better as compared to others.

Figure 4

Actual and Forecast Values of Stock Return for all Models



Conclusion and Recommendations

The ARIMA, ANN, and SVM models represented an effective flexibility in forecasting time series data; however, none of these individually gave a satisfactory performance. In the current study, an attempt was made to explore the joint models for inflation, exchange rate, and stock return, in which linear ARIMA was combined with the nonlinear models, such as ANN and SVM. The forecast performance of two hybrid models with the three independent models was evaluated through RMSE, MAE, and graphical representation. Empirical results clearly indicated that the ARIMA-SVM hybrid model outperformed all other models including the hybrid model of ARIMA-ANN.

The outstanding achievement of ARIMA-SVM above other models was due to the fact that SVM applied the structural risk minimization principle, which minimized an upper bound of the generalization error rather than minimizing the training error. This ultimately led to superior generalization performance as compared to other nonlinear models. The hybrid ARIMA-SVM model superiority showed that it could be utilized by policymakers and investors to forecast economic and financial data.

References

Al-hnaity, B., & Abbod, M. (2016). Predicting financial time series data using hybrid model. In *Intelligent systems and applications* (pp. 19-41). Springer.

Awan, S. M., Khan, Z. A., Aslam, M., Mahmood, W., & Ahsan, A. (2012). *Application of NARX based FFNN, SVR and ANN Fitting models for long term industrial load forecasting and their comparison* [Paper presentation]. Industrial Electronics (ISIE), IEEE International Symposium on (pp. 803-807).

Babu, C. N., & Reddy, B. E. (2014). A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. *Applied Soft Computing*, *23*, 27-38. https://doi.org/10.1016/j.asoc.2014.05.028

Box. G. E. P., & Jenkins, G. M. (1976). Time series analysis, forecasting arid control (2nd Ed.). Holden Day.

Breiman, L. (1996). Stacked regressions. *Machine learning*, *24*(1), 49-64. https://doi.org/10.1007/BF00117832

Burney, S. M. A., Jilani, T. A., Ardil, C., & Muhammad, S. (2005). Levenberg-Marquardt algorithm for Karachi Stock Exchange share rates forecasting. *International Journal of Computational Intelligence*, *1*(3), 144-149.

Cao, L. J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, *14*(6), 1506-1518. https://doi.org/10.1109/TNN.2003.820556

Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. *Neural Computing & Applications*, *10*(2), 184-192. https://doi.org/10.1007/s005210170010

Chen, K. Y., & Wang, C. H. (2007). A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications*, *32*(1), 254-264. https://doi.org/10.1016/j.eswa.2005.11.027

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297. https://doi.org/10.1007/BF00994018

Farooqi, A. (2014). ARIMA model building and forecasting on Imports and Exports of Pakistan. *Pakistan Journal of Statistics and Operation Research*, *10*(2), 157-168. https://doi.org/10.18187/pjsor.v10i2.732

Haider, A., & Hanif, M. N. (2009). Inflation forecasting in Pakistan using artificial neural networks. *Pakistan Economic and Social Review*, *47*(1), 123-138.

Karathanasopoulos, A., & Osman, M. (2019). Forecasting the Dubai financial market with a combination of momentum effect with a deep belief network. *Journal of Forecasting*, *38*(4), 346-353. https://doi.org/10.1002/for.2560

Kenny, G., Meyler, A., & Quinn, T. (1998). *Forecasting Irish inflation using ARIMA models* (No. 3/RT/98). Central Bank of Ireland.

Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, *37*(1), 479-489. https://doi.org/10.1016/j.eswa.2009.05.044

Khashei, M., & Bijari, M. (2011). Which methodology is better for combining linear and nonlinear models for time series forecasting? *Journal of Industrial and Systems Engineering*, *4*(4), 265-285.

Kumar, M., & Thenmozhi, M. (2014). Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *International Journal of Banking, Accounting and Finance*, *5*(3), 284-308. https://doi.org/10.1504/IJBAAF.2014.064307

Kuntoji, G. Rao, S., & Mandal, S. (2017). Application of support vector machine technique for damage level prediction of tandem breakwater. *International Journal of Earth Sciences and Engineering*, *10*, 633-638

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115-133. https://doi.org/10.1007/BF02478259

Mojirsheibani, M. (1999). Combining classifiers via discretization. *Journal of the American Statistical Association*, *94*(446), 600-609.

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*(6), 497-505. https://doi.org/10.1016/j.omega.2004.07.024

Papatla, P., Zahedi, M., & Zekic-Susac, M. (2002). Leveraging the strengths of choice models and neural networks: A multiproduct comparative analysis. *Decision Sciences*, *33*(3), 433-461. https://doi.org/10.1111/j.1540-5915.2002.tb01651.x

Petruseva, S., Zileska-Pancovska, V., Žujo, V., & Brkan-Vejzović, A. (2017). Construction costs forecasting: comparison of the accuracy of linear regression and support vector machine models. *Technical Gazette*, *24*(5), 1431-1438.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, *36*(2), 111-133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

Sultana, K., Rahim, A., Moin, N., Aman, S., & Ghauri, S. P. (2013). Forecasting inflation and economic growth of Pakistan by using two time series methods. *International Journal of Business and Economics Research*, *2*(6), 174-178. https://doi.org/10.11648/j.ijber.20130206.17

Tkacz, G., & Hu, S. (1999). *Forecasting GDP growth using artificial neural networks* (Working Paper No. 1999-3). Bank of Canada.

Zhang, G. P. (2001). An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research*, *28*(12), 1183-1202. https://doi.org/10.1016/S0305-0548(00)00033-2

Zhu, B., & Wei, Y. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega, *41*(3), 517-524.