Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(1) 2016 7-18

A Comparative Study on Time-delay Neural Network and GARCH Models for Forecasting Agricultural Commodity Price Volatility

Achal Lama[1], Girish K. Jha[2], Bishal Gurung[1], Ranjit Kumar Paul[1], Anshu Bharadwaj[1] and Rajender Parsad[1]

[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

[2]*ICAR-Indian Agricultural Research Institute, New Delhi*

Received 30 November 2015; Revised 27 February 2016; Accepted 29 February 2016

SUMMARY

In this paper, forecasting performance of time-delay neural network and GARCH models for predicting the volatility using monthly price series of edible oils in domestic and international markets is evaluated. An attempt has also been made to investigate whether the forecasting performance of two competing models can be improved by combining their individual forecasts. For this purpose, the individual models were combined to produce improved forecasts using non-parametric approach through the use of kernel. Further, the models were evaluated on their ability to predict the correct change of direction (CCD) for future values.

Keywords: Time-delay neural network, GARCH, Non-parametric, Combining forecasts.

1. INTRODUCTION

The presence of increased volatility in the agricultural commodity prices has become a common feature mainly due to globalisation. Volatility is of much concern as its presence disrupts the normal behaviour of any time-series data and agricultural commodity price series is no exception to it. Understanding the nature of agricultural commodity price volatility is required for improving agricultural market analysis and policy development. This has led to the development and application of many time-series models. As a result, modelling and forecasting of volatility by nonlinear models has emerged as an important tool for time-series analysis. The most commonly used statistical models are the Autoregressive Conditional Heteroscedastic (ARCH) models (Engle 1982), Generalised ARCH (GARCH) model (Bollerslev 1986), Bilinear (BL) time-series models (Granger and Anderson 1978), Threshold Autoregressive (TAR) model (Tong 1983) and Smooth Transition Autoregressive (STAR) models (Tsay 1989). All these models are nonlinear and parametric in nature. These models are widely used in time-series forecasting with a huge domain covering important business areas such as economics and agriculture. The researchers have used these models to explain various volatile phenomenon observed in stock market and prices of agricultural commodities (Sundaramoorthy *et al.* 2014, Lama *et al.* 2015). However, parametric nature of these nonlinear models makes them a little restrictive in application as it requires hypothesizing of an explicit relationship for the data series at hand

Corresponding author: Achal Lama
E-mail address: chllm6@gmail.com

with little knowledge of the data generating process. Essentially, the formulation of a nonlinear model to a particular data set is a very difficult task since there are too many possible nonlinear patterns and a pre-specified nonlinear model may not be adequate to capture all the important features of the data set (Zhang *et al*. 1998). Hence, in order to overcome such limitations, one may opt for non-parametric, data driven and self-adaptive computational methods such as Artificial Neural Network (ANN). Neural networks have also been extensively used in time series literature because of their ability to recognize and learn complex non-linear patterns (Zhang *et al*. 1998). ANN modelling technique has been used for estimation and forecasting in many fields of study including agriculture, economics and statistics (Jha and Sinha 2014). Most uses of ANN in economics have so far been in financial market, in part because traditional approaches have had low explanatory power and in part because the ANN approaches requires abundant data. The use of the ANN model in applied work is generally motivated by a mathematical result stating that under mild regularity conditions, a relatively simple ANN model is capable of approximating any Borel-measurable function to any given degree of accuracy (Fine 1999). Such an approximation would still contain a finite number of parameters.

Now-a-days with increasing horizons of time-series models, the researchers have lot many options to choose among the available models. This has unquestionably increased the domain of applicability of the time-series models, but also has opened up a new way of thinking whether one can combine the forecasts of two competing models and yield a better result. The literature has shown that a weighted average of forecasts is often more accurate than any of the individual forecasts (Clemen 1989, Zhang 2003). The motivation for combining models comes from the assumption that either one cannot identify the true data generating process (Terui and van Dijk 2002) or that a single model may not be sufficient to identify all the characteristics of the time series (Zhang 2003). Further, in Section 2 brief details of the ARMA, GARCH and ANN models have been described, followed by discussion for the need of combining of ANN and GARCH models in Section 3. Section 4 of this paper deals with the different statistics used for forecast evaluation and their appropriateness. Lastly, the paper is concluded by highlighting the empirical results in Section 5. In Section 6, conclusion and future works are delineated.

2. A BRIEF DESCRIPTION OF MODELS: ARMA, GARCH AND ANN

The autoregressive moving average (ARMA) model has been used to model the conditional mean of the price series due to its simple structure and statistical properties (Hamilton 1994). In ARMA model, variable of interest is assumed to be a linear function of past actual values and a random shock. An ARMA (*p, q*) model is defined by the equation (Box *et al.* 1994).

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad (1)$$

that is, $$\left(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p\right) y_t = \left(1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q\right) \varepsilon_t \quad (2)$$

or $\varphi(B) y_t = \theta(B) \varepsilon_t$, t=1,2,3,…,n (3)

where B is the backshift operator defined by $B^r y_t = y_{t-r}$, $\phi(B)$ and $\theta(B)$ are polynomials of degree p and q in B, r is the degree of polynomial p and n is the number of observations. A generalization of ARMA models, which incorporates a wide class of nonstationary time-series, is obtained by introducing ‘differencing’ in the model. The most suitable ARMA model is selected using the smallest Akaike Information Criterion (AIC) or Schwarz-Bayesian Criterion (SBC) value and the lowest root mean square error (RMSE).

Volatility measures the second moment of the price distribution. The first type of model that has been used for modelling and forecasting the time varying conditional variance is the simple

GARCH (*p*, *q*) model. The conditional variance V_t in a GARCH (*p*, *q*) model is a deterministic function of past squared innovations and lags of itself and is given by

$$V_t = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} b_j V_{t-j} \quad (4)$$

A sufficient condition for the conditional variance to be positive is

$$a_0 > 0,\ a_i \geq 0,\ i = 1,2,...,q.\ \ b_j \geq 0,\ \ j = 1,2,...,p$$

The order of the GARCH model suitable for the data was determined using Portmanteau Q test which is based on squared residuals. The estimates of GARCH model were obtained using maximum likelihood. The log-likelihood function is computed from the product of all conditional densities of the prediction errors.

The second type of model that has been used for modelling and forecasting the conditional variance of price series is artificial neural network. Neural network models are computational methods that mimic the behaviour of the human brain's central nervous system. They are considered as a class of generalized non-linear, nonparametric, data driven statistical methods. General neural networks architecture consists of an input layer that accepts external information; one or more hidden or middle layer that provides non-linearity to the model and an output layer that provides the target value. Each layer contains one or more nodes. All the layers in a multi-layer neural network are connected through an acyclic arc.

Time series data can be modelled using neural network in two possible ways. The first way is to explicitly represent time in the form of recurrent connections from output nodes to the preceding layer (Elman 1990). The second way is to provide the implicit representation of time; where by a static neural network like multilayer perceptron is bestowed with dynamic properties (Haykin 1999). A neural network can be made dynamic by embedding either long-term or short-term memory, depending on the retention time, into the structure of a static network. An example of such architecture is a time-delay neural network (TDNN), which has been employed for the present study.

In this study, we used neural network with only one hidden layer, as it is capable of producing a better modelling performance. In TDNN, determination of number of input nodes plays a crucial role as it helps in modelling the autocorrelation structure of the data. The determination of number of output nodes is relatively easy. In this study, one output node is used and multi-step ahead forecasting is done using the iterative procedure as used in Box-Jenkins method. This involves use of forecast value as an input for forecasting the future value. It is always better to select the model with small number of nodes at hidden layer as it considerably improves the out of the sample forecasting performance and also avoids the problem of over fitting. Data pre-processing has significant impact on neural network learning and generalization ability. The relationship between the conditional variance V_t and its past lags which serves as inputs $(V_{t-1}, V_{t-2}, ..., V_{t-p})$ has the following mathematical representation.

$$V_t = f\left(\alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} V_{t-i}\right)\right) + \varepsilon_t \quad (5)$$

where, $\alpha_j\,(j = 0, 1, 2, ..., q)$ and $\beta_{ij}(i = 1, 2, ..., p; j = 1, 2, ...q)$ are the model parameters often called the connection weights; p is the number of input nodes, q is the number of hidden nodes, f and g are the activation function. Hence, the TDNN model of equation (5) in fact performs a nonlinear functional mapping from the past observations $(V_{t-1}, V_{t-2}, ..., V_{t-p})$ to the future value V_t, i.e.

$$V_t = f(V_{t-1}, V_{t-2}, ..., V_{t-p}, w) + \varepsilon_t \quad (6)$$

where, w is a vector of all parameters and f is a function determined by network structure and connection weights. Thus neural network is equivalent to a nonlinear autoregressive model. Here, the expression (5) implies one output node in the output layer which is typically used for one-step-ahead forecasting. The simple network given by equation (5) is surprisingly powerful in that it is able to approximate arbitrary function as the number of hidden nodes q is sufficiently large.

In addition to choosing an appropriate number of hidden nodes, another important task of TDNN modelling of a time series is the selection of the number of lagged observations, p that is the dimension of the input vector. This is perhaps the most important parameter to be estimated in a TDNN model because it plays a major role in determining the (nonlinear) autocorrelation structure of the time series. However, there is no theory that can be used to guide the selection of p. Hence, experiments are often conducted to select an appropriate p as well as q. It is the lagged input which makes the neural network dynamic thus resulting in time-delay neural network (TDNN). For p tapped delay nodes, q hidden nodes, one output node and biases at both input and hidden layer, the total number of parameters (weights) in a three layer feed forward neural network is $q(p + 2) + 1$. In fact, for building an appropriate neural network model, the available data is often divided into two portions. The first part is used for model training, i.e. parameter estimation and model selection. The second part of the sample is then used for true forecasting evaluation. The parameters of the model are estimated via least squares method through a nonlinear optimization routine. The objective of training is minimization of an error function that measures the misfit between the predicted value and the actual value for any given value of w. The error function which is widely used is given by the sum of the squares of the error between the predicted value $\widehat{V}_t$ for time t and the corresponding target value V_t at time t, so that we minimize.

$$E(w) = 1/2\sum_{t=p+1}^{n}\left[V_t - \widehat{V}_t\right]^2 \quad (7)$$

where the factor 1/2 is included for mathematical simplification. The error surface for multilayer feed forward neural network with non-linear activation function is complex and believed to have many local and global minima.

3. COMBINING FORECAST FROM TDNN AND GARCH MODELS

Considerable literature has accumulated over the years regarding the combination of forecasts. It has been found in many instances that weighted average of forecasts is often more accurate than any individual forecasts (Dunis and Huang 2002). The method of combing the forecast by simple average is used widely due to its ease of calculation (Clemen 1989). Another method of combining forecast is deciding the weights with reference to cost of an error, and minimising the cost by assuming it to be proportional to square of the errors (Lupoletti and Webb 1986). Further, forecast can be combined by simple regression procedure, which results in smallest mean squared error and has an unbiased combined forecast even if individual forecasts are biased (Granger and Ramanathan 1984). In general, some drawbacks exist for all the previously mentioned ways of weighting the individual forecasts.

In this study we have combined the forecasts from the GARCH and the TDNN models using a weighting scheme based on non-parametric smoothing method (Tsangari 2007). The process of finding the optimum weights is based on kernels. Major advantage of using this method is its freedom from the assumptions of functional form of the weights as well as the errors of the individual models are serially or cross-uncorrelated. Later one being more important as it is violated mostly in the economic or price data. Moreover, this technique is robust to the presence of outliers and structural breaks. A kernel function K is a continuous, bounded and symmetrical function which satisfies the condition.

$$\int_{-\infty}^{\infty} K(x)dx = 1 \quad (8)$$

Usually, K is a symmetric probability density function, for instance the normal density function. We used the regression relationship of the actual value V_t with the predictors' f_{1t} and f_{2t} as:

$$V_t = m(f_t) + e_t \\ = m(f_{1t}, f_{2t}) + e_t \quad (9)$$

where $m(f_t)$ is the mean response function, and f_{1t}, f_{2t} are the individual forecasts of the conditional variance as obtained by the GARCH and TDNN models. We have estimated $m(f_t)$ non-parametrically, which is a flexible functional form of the regression curve, where neither the error distribution nor the functional form of the mean function is pre-specified and thus predictions of observations can be made without a fixed parametrical model. To estimate the function, we have used the Nadaraya–Watson kernel weights. In this study we have used Gaussian kernel for smoothing the forecasts before combining them. Finally after minimising the Averaged Squared Error (ASE) asymptotically, we got an optimal bandwidth h for our kernel function. Then we estimated the mean response function by the following equation.

$$\hat{m}_h(f_t) = \frac{n^{-1}\sum_{t=1}^{n} K_h(f - f_t)V_t}{n^{-1}\sum_{t=1}^{n} K_h(f - f_t)} \quad (10)$$

where $k_h(f_t) = h^{-2}k(f_t / h)$, $f_t = (f_{1t}, f_{2t})$, f is the functional form of kernel and $k_h(f)$ is the kernel function at bandwidth h.

4. FORECAST EVALUATION METHODS

To evaluate the forecasting performance of different models, the following statistics have been used. The mean square error (MSE) measures the overall performance of a model and is given by

$$\text{MSE} = \sum_{t=1}^{n} ((V_t - \hat{V}_t)^2 / n)$$

where V_t and $\hat{V}_t$ is the actual and predicted conditional variance respectively at time t. This statistic is scale dependent and small value of MSE indicates good forecasting accuracy. The root mean square error (RMSE) is the square root of MSE. The second criterion, the Theil's inequality coefficient or Theil U-statistic is scale independent and is expressed as

$$U = \sqrt{(1/n)\sum_{t=1}^{n}(\hat{V}_t - V_t)^2} \Bigg/ \left[\sqrt{(1/n)\sum_{t=1}^{n}\hat{V}_t^2} + \sqrt{(1/n)\sum_{t=1}^{n}V_t^2}\right]$$

The value of U ranges between 0 and 1 and value close to 0 indicates efficient model. The third measure, correct directional change (CDC) provides the direction of change and is given by

$$CDC = (100/n)\sum_{t=1}^{n} D_t$$

where $D_t = \begin{cases} 1, & \text{if } (V_t - V_{t-1}).(\hat{V}_t - \hat{V}_{t-1}) > 0 \\ 0, & \text{elsewhere.} \end{cases}$

It is used to check whether the direction given by the volatility forecast is the same as the actual change that has subsequently occurred. Higher the value of CDC better is the forecasting accuracy of the model concerned.

5. EMPIRICAL RESULTS

5.1 Data and Implementation

Monthly wholesale price index of edible oils in India, obtained from the Office of the Economic Adviser, Ministry of Commerce and Industry and comparable international monthly price data of edible oils published by the World Bank have been used. The price series under consideration covers a total of 360 months (April, 1982 to March, 2012). The first 348 observations were used for training purpose and the last 12 months price data was retained for testing purpose. Fig. 1 and 2 exhibit the time plot of domestic and international edible oils series respectively. In order to compare the performance of TDNN and GARCH models, we followed the usual modelling procedures namely identification, estimation, diagnostic checking and evaluation, for both price series. In most of the previous studies (Tsangari 2007), data have been log-transformed, so there remains very little non-linearity for a neural network model to capture and to improve on a linear model. Accordingly, in this study, we did not carry out logarithmic transformation to the data in order to preserve inherent non-linearity. We modelled the conditional mean of both domestic and international price series using autoregressive (AR) model of order 2. Given the ARMA (2, 0) structure for the level price series, we obtained the residuals (actual observations minus predicted values) for both series. The square of the residuals was created as a new variable (esquare) which was used as input to TDNN model. The squared residuals from both series exhibited heteroscedastic structure which was tested using ARCH-LM test, results are presented in Table 1. This clearly highlighted the need to model the mean and variance of the series simultaneously using GARCH model.

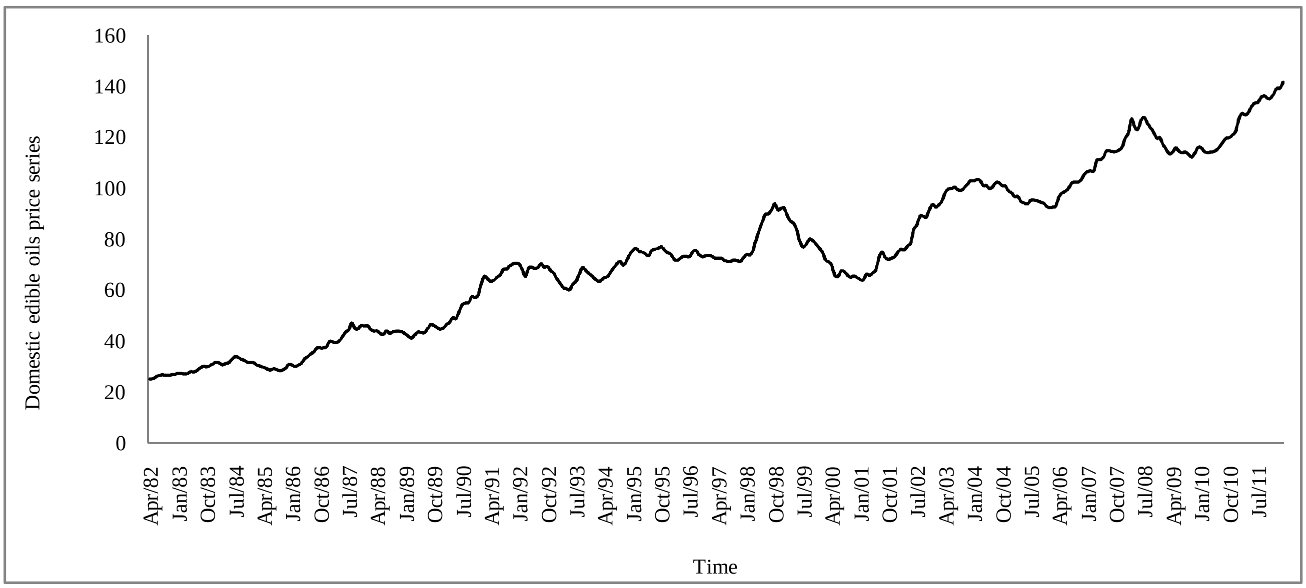

Fig. 1. Time plot of the domestic edible oils price series

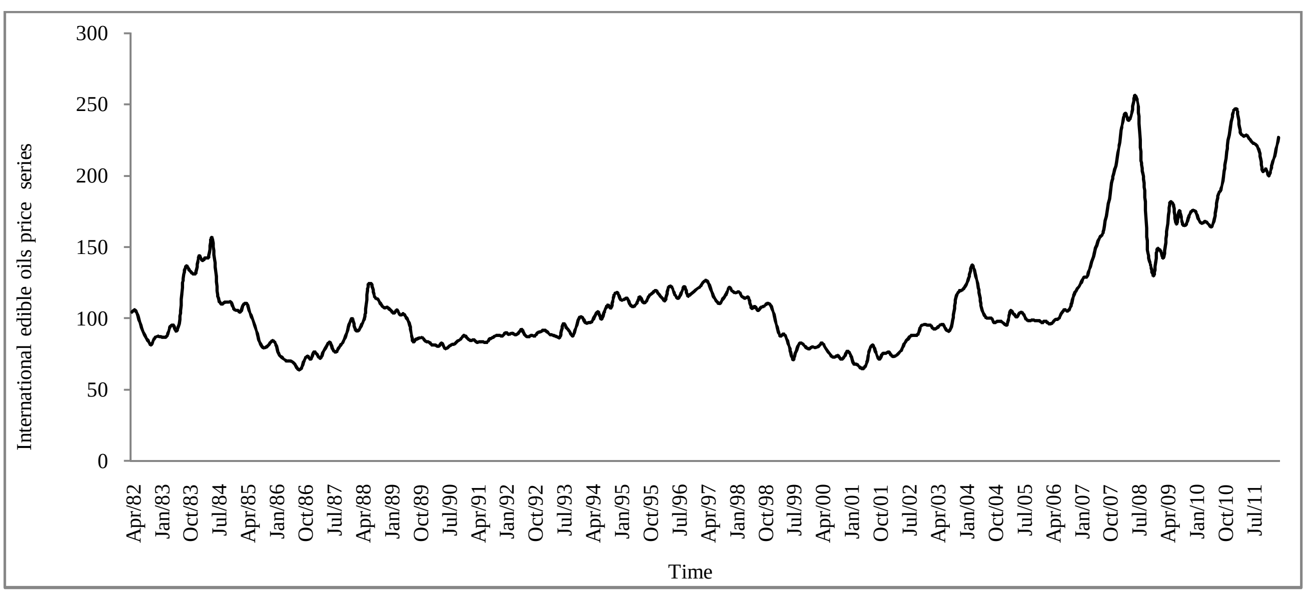

Fig. 2. Time plot of the international edible oils price series

Table 1. ARCH-LM test

Lags	Domestic Edible Oils Price Series		International Edible Oils Price Series	
	Q	P Value	Q	P Value
1	336.32	<.0001	309.32	<.0001
2	656.57	<.0001	550.00	<.0001
3	962.50	<.0001	724.16	<.0001
4	1256.64	<.0001	843.79	<.0001
5	1538.86	<.0001	921.25	<.0001
6	1808.99	<.0001	970.81	<.0001

We obtained the best time-delay neural network with single hidden layer for the 'esquare' series using the neural network toolbox of MATLAB 7.10. Multiple retries were used with different random starting points, in order to avoid local minima and find the global minimum. We varied the number of input nodes from 2 to 6 and the number of hidden nodes from 2 to 10 with an increment of 2 with basic cross validation method. For lags 5 and 6, we varied hidden nodes from 2 to 14 as there was decreasing trend for training error along with selection error. Thus, a total of 29 neural network models were tried for each series before arriving at the final structure of the model. Essentially, the process of exploration and exploitation was carried out to obtain the best model for the given series. There are many variations of the backpropagation algorithm used for training feedforward networks. In this study, the Levenberg-Marquardt algorithm (Haykin 1999) which has been designed to approach second-order training speed without computing the Hessian matrix has been employed. The logistic and identity function have been used as activation function for the hidden nodes and output node respectively. The logistic function is not used at the output stage in time series forecasting unless the data are suitably scaled to lie in the interval (0, 1). We focus primarily on one step ahead forecasting and the multi-step ahead forecasting is done using iterative procedure so only one output node is employed. A typical TDNN structure with one hidden layer is denoted by I:H*s*:O*l*, where I is the number of nodes in the input layer, H the number of nodes in the hidden layer, O the number of nodes in the output layer, *s* denotes the logistic sigmoid transfer function and *l* indicates the linear transfer function.

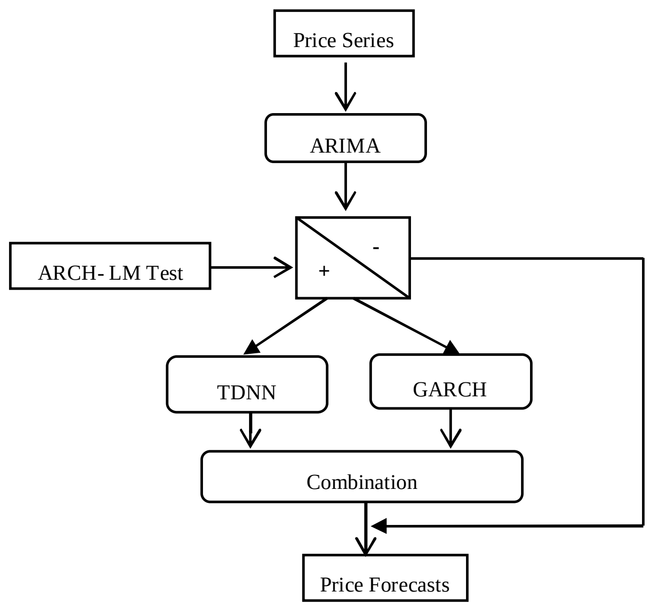

Fig. 3. Hybrid methodology that combines forecasts from TDNN and GARCH models

We also investigated the forecasting performance of combined forecasts obtained from two non-linear models TDNN and GARCH (see Fig. 3). The square of the residuals from the AR process of the AR-GARCH model is comparable to the conditional variance of the AR-GARCH model (Tsangari 2007). The estimates of optimal weights were obtained using non-parametric smoothing techniques. In this study, we employed Gaussian kernel function, which determines the shape of the kernel weights. The size of the weight is parameterized by the bandwidth, *h*. The bandwidth is selected so as to have a trade-off between bias and variance. The smaller bandwidths provide weights more concentrated around the forecast series. We used cross validation and the leaveone-out method to estimate the bandwidth for each forecast time series. After selection of the bandwidth and smoothing of the series, they were combined using Nadaraya-Watson kernel based regression method. In this method the weights given to each series is estimated non-parametrically in the form of mean response function [For details see Section 3]. The forecasting ability of both models for each 'esquare' series was assessed with respect to previously mentioned performance measures viz. mean squared error (MSE), Theil U-statistic and correct directional change (CDC).

5.2 Discussion

As mentioned earlier, we modelled the squared residual (esquare variable) series obtained after fitting AR(2) to both domestic and international edible oils price indices using GARCH and TDNN models. We obtained the best AR-GARCH model for both series based on the lowest AIC and BIC information criteria. We selected AR (2)-GARCH (1, 1) model for two series at hand, parameter estimates along with their standard errors are presented in Table 2. The results indicate a good fit to both the series as the sum of coefficients a and b is less than 1. Due importance is given to the well behaved residuals while selecting the best model. Table 3 summarizes the forecasting performance of various TDNN models for domestic edible oils price series in terms of training and testing RMSE. In case of testing, we calculated RMSE at 4 different time points, 1 month ahead, 3 months ahead, 6 months ahead and 12 months ahead. Table 3 clearly indicated that the forecasting ability of five TDNN structures viz. 5:12s:1*l*, 5:14s:1*l*, 6:10s:1*l*, 6:12s:1*l* and 6:14s:1l were close on the basis of training RMSE. These five models were further assessed on the basis of their out of the sample forecast performance at different time horizons and finally a parsimonious model was selected. Out of a total of 29 neural network structures, a TDNN model with six tapped delay and twelve hidden nodes (*6:12s:1l*), was selected for volatility forecasting of domestic edible oils price series.

Similar exercise was carried out to identify the best TDNN model for volatility forecasting of international edible oil price series. Table 4 summarizes the forecasting performance of various TDNN models for international edible oils price series in terms of training and testing RMSE. Various models were fitted and on basis of their out of the sample volatility forecast performance at different time horizons, the TDNN model with five input nodes and twelve hidden nodes (*5:12s:1l)* was selected for modelling and forecasting of the international edible oils price series. The issue of finding a parsimonious model is taken into account while selecting the best model for each price series. The parsimonious models not only have the recognition ability but also have the more important generalization ability. Table 5 provides in-sample and out of sample performances for the best TDNN, GARCH and combined models with respect to the MSE, Theil-U and CDC (section 4) for both price series. Table 5 revealed that the TDNN outperformed the GARCH model in terms of MSE and Theil-U for both the domestic and international edible oils price series, while, GARCH dominated in terms of CDC for both the price series under consideration. In this study, our interest centres on short-term forecasting and hence we consider forecast horizon of up to a year. Further mixed results were obtained for both models. The TDNN outperformed GARCH model for domestic price series, while GARCH was found better than TDNN for the international price series in terms of MSE and Theil–U statistics. At this juncture, it is worth mentioning that for all cases the best neural network model in terms of test RMSE is obtained for a forecasting horizon of 1 month. The same model was used for other forecast horizons also. In this context, several researchers (Swanson and White 1997, Hervai *et al.* 2004) have recommended that a specific neural network model should be selected for each forecast horizon which implies that p and q may vary over forecast horizon. This will in general improve the performance of TDNN model with respect to each forecast horizon. The multiple model approach is not of much advantage in case of GARCH model.

In this study, forecast accuracy of the competing models is measured by the correct directional change (CDC) besides MSE and Theil-U statistics. Table 5 provides post-sample performance of the competing models and the combined model in terms of CDC at 12 months ahead. The implications of the direction of change results of Table 5 are, however, different from the results based on RMSE. At 12 months horizon, the neural network model always has a larger percentage of correct sign than the GARCH model for both price series. This ability holds immense importance in the markets where the goods are mostly perishable and holding them for a longer period is useless if it does not pay the needed revenue. Thus the market players largely rely on the forecasting of the direction of change. So, on the basis of this criterion it can be said that the TDNN is undoubtedly the best model for the two series under consideration. In this context, Dacco and Satchell (1999) have shown that MSE type measures may be inappropriate for nonlinear models since these measures can imply that the nonlinear model is less accurate than a linear one even when the nonlinear model is the true data generating process. In effect, the nonlinear model may generate more variation in forecast values than a linear model, and hence it may be unduly penalized for errors that are large in magnitude.

Table 2: Parameter estimates of the GARCH model

Series	Model	a_0	a_1	b_1	AIC Value
Domestic edible oils price series	AR(2)-GARCH(1,1)	0.05 (0.02)	0.09 (0.03)	0.88 (0.03)	1191.90
International edible oils price series	AR(2)-GARCH(1,1)	3.41 (0.79)	0.40 (0.07)	0.54 (0.06)	2091.03

Table 3. Forecasting performance of TDNN models for domestic edible oils price series

Model	Parameters	RMSE Training	RMSE			
			1 Month Ahead	3 Months Ahead	6 Months Ahead	12 Months Ahead
2:2s:1l	9	3.44	1.07	6.96	7.68	15.29
2:4s:1l	17	3.45	1.09	6.88	9.34	14.48
2:6s:1l	25	3.05	0.87	5.92	10.43	11.92
2:8s:1l	33	3.01	0.33	6.37	11.10	22.95
2:10s:1l	41	2.99	0.02	6.49	9.70	19.61
3:2s:1l	11	3.49	0.58	7.18	10.07	15.29
3:4s:1l	21	3.42	0.23	6.32	7.38	14.46
3:6s:1l	31	3.21	0.96	5.38	10.14	17.44
3:8s:1l	41	3.07	0.33	3.77	6.85	20.31
3:10s:1l	51	3.15	0.65	3.60	8.41	16.11
4:2s:1l	13	3.48	0.94	7.09	11.67	17.13
4:4s:1l	25	3.24	0.93	7.33	9.03	17.52
4:6s:1l	37	3.22	1.25	7.30	8.12	16.31
4:8s:1l	49	3.07	0.76	6.85	12.18	15.02
4:10s:1l	61	3.04	1.98	6.10	8.05	14.48
5:2s:1l	15	3.44	0.15	7.28	11.20	16.67
5:4s:1l	29	3.11	0.12	6.47	6.75	17.02
5:6s:1l	43	2.93	0.34	7.67	9.48	20.13
5:8s:1l	57	2.98	0.34	7.26	6.44	16.93
5:10s:1l	71	2.92	0.17	5.89	8.13	15.64
5:12s:1l	85	2.29	0.11	6.68	8.85	22.89
5:14s:1l	99	2.35	0.13	4.14	9.29	17.45
6:2s:1l	17	3.37	0.14	6.53	10.99	18.02
6:4s:1l	33	3.01	0.15	5.36	9.11	16.13
6:6s:1l	49	3.01	0.01	4.97	9.67	10.47
6:8s:1l	65	2.76	0.31	6.48	8.74	13.07
6:10s:1l	81	2.45	0.02	6.27	7.78	15.30
6:12s:1l	97	1.63	0.42	7.85	5.94	14.83
6:14s:1l	113	1.95	0.32	4.75	8.25	25.23

Table 4. Forecasting performance of TDNN models for international edible oils price series

Model	Parameters	RMSE Training	RMSE			
			1 Month Ahead	3 Months Ahead	6 Months Ahead	12 Months Ahead
2:2s:1l	9	86.28	15.49	14.35	13.16	100.26
2:4s:1l	17	62.16	9.92	9.16	9.35	93.27
2:6s:1l	25	58.12	8.14	7.05	7.97	94.44
2:8s:1l	33	57.44	4.98	8.54	8.64	100.80
2:10s:1l	41	55.46	9.48	8.46	8.75	112.05
3:2s:1l	11	82.96	17.52	15.73	13.90	92.01
3:4s:1l	21	54.00	11.29	12.21	11.76	99.72
3:6s:1l	31	36.88	8.35	16.18	14.29	82.75
3:8s:1l	41	35.58	9.87	14.97	13.06	116.55
3:10s:1l	51	34.57	9.74	14.89	12.91	144.26
4:2s:1l	13	82.78	10.64	15.62	14.43	98.24
4:4s:1l	25	57.09	7.30	15.23	15.94	88.60
4:6s:1l	37	44.07	4.28	10.95	11.30	103.57
4:8s:1l	49	33.68	8.87	11.56	10.98	180.29
4:10s:1l	61	32.71	8.06	13.75	12.55	521.89
5:2s:1l	15	67.07	16.75	15.67	14.38	94.93
5:4s:1l	29	50.21	7.45	11.36	11.07	101.02
5:6s:1l	43	30.56	9.55	8.60	9.37	109.47
5:8s:1l	57	30.73	2.79	1.90	8.35	308.07
5:10s:1l	71	23.59	5.29	11.89	11.05	176.48
5:12s:1l	85	22.22	1.05	6.58	7.12	294.38
5:14s:1l	99	21.58	3.39	9.37	10.06	187.72
6:2s:1l	17	68.99	13.83	12.76	11.91	93.85
6:4s:1l	33	38.95	0.57	8.55	9.58	87.82
6:6s:1l	49	38.68	2.51	4.65	7.64	86.46
6:8s:1l	65	28.11	5.78	7.54	8.05	273.28
6:10s:1l	81	22.44	3.64	4.72	7.35	238.79
6:12s:1l	97	20.63	1.78	6.22	9.69	497.47
6:14s:1l	113	20.72	9.52	10.19	9.81	492.50

Table 5. Forecasting performance of GARCH, TDNN and combined models

	GARCH	TDNN	Combined
Domestic edible oils price series			
MSE (In sample)	14.33	9.61	12.72
MSE (Out of sample)[a]	55.29	33.04	39.37
CDC (In sample)	85.96	50.58	50.00
CDC (Out of sample)	18.18	35.86	43.74
THEIL–U (In sample)	0.58	0.42	0.55
THEIL–U (Out of sample)	0.94	0.51	0.67
International edible oils price series			
MSE (In sample)[b]	12.65	2.49	9.22
MSE (Out of sample)[c]	0.25	17.78	19.35
CDC (In sample)	97.65	49.26	53.95
CDC (Out of sample)	27.27	53.64	44.00
THEIL–U (In sample)	0.70	0.22	0.53
THEIL–U (Out of sample)	0.70	0.86	0.73

Note: a= values are to be multiplied by 10^2, b = values are to be multiplied by 10^3,

c = values are to be multiplied by 10^5

To find out whether the combined model performed better than single model or not, the out of the sample performance in terms of MSE and Theil-U statistics at different forecast horizons is presented in Table 5 for domestic and international edible oils price series. Analysis revealed that the in-sample performance of combined model was better as compared to GARCH model in terms of MSE and Theil-U statistics, but was not as efficient as the TDNN model. Results relating to combined model provided mixed inferences. Of course, its performance was in between the competing models. The post-sample forecast accuracy of the combined model was found to be better than the GARCH model but not as efficient as TDNN in terms of MSE and Theil-U statistics for the domestic price series. In case of international edible oils price series, GARCH model was found superior to the other two models in terms of MSE and Theil-U statistics. In terms of CDC, TDNN model outperformed other two models in case of international price series while combined was superior than the other two models in case of domestic edible oils price series.

6. CONCLUSION

In this study, we found the best time lagged neural network with single hidden layer for each series by conducting an experiment with the basic cross validation method. A total of 29 TDNN models were tried to capture the variability more efficiently. The study provided mixed results for both competing models in terms of MSE and Theil-U statistics. The neural network model always provided larger percentage of correct sign than the GARCH model for both price series at 12 months horizon. The results relating to direction of change imply that the relative forecasting performance of TDNN and GARCH models crucially depends on how performance is measured. We also investigated the forecasting accuracy of non-parametric combination of GARCH and TDNN models forecasts which too provided mixed results. In case of domestic oil series the combined model was uniformly better, unlike in the international oil series, where its forecast performance was in between the two competing models. The performance of combined model depends largely on size of bandwidth and the form of kernel. One of the limitations of this study is that we have evaluated combined model with Gaussian kernel and one estimated bandwidth instead of varying bandwidth.

REFERENCES

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econ.*, **31**, 307-327.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). Time series analysis. In: *Forecasting and Control*, 3rd edition, Prentice Hall, New Delhi.

Clemen, R. (1989). Combining forecasts: A review and annotated bibliography. *Inter. J. Forecast.*, **5,** 559-583.

Dacco, R. and Satchell, S. (1999). Why do regime-switching models forecast so badly. *J. Forecast.*, **18,** 1-16.

Dunis, C. and Huang, X. (2002). Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *J. Forecast.*, **21,** 317-354.

Elman, J. (1990). Finding structure in time. *Longtive Sci.*, **14,** 179-211.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50,** 987-1008.

Fine, T.L. (1999). *Feed Forward Neural Network Methodology*. Springer, New York.

Granger, C.W.J. and Anderson, A.P. (1978). *An Introduction to Bilinear Models*. Van derhoeck and Ruprecht, Gottingen.

Granger, C.W.J. and Ramanathan, R. (1984). Improved methods for forecasting. *J. Forecast.*, **3,** 197-204.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. 2nd edition, Prentice Hall.

Hervai, S., Osborn, D.R. and Birchenhall. C.R. (2004). Linear versus neural network forecast for European industrial production series. *Inter. J. Forecast.*, **20,** 435-446.

Jha, Girish K. and Sinha, K. (2014). Time delay neural networks for time series prediction: An application to the monthly wholesale price of oilseeds in India. *Neural Comput. Appl.*, **24,** 563-571.

Lama, A., Jha, G.K., Paul, R.K. and Gurung, B. (2015). Modelling and forecasting of price volatility: An application of GARCH and EGARCH models. *Agric. Econ. Res. Rev.*, **28,** 73-82.

Lupoletti, W.M. and Webb, R.H. (1986). Defining and improving the accuracy of macroeconomic forecasts: contributions from a VAR model. *J. Business*, **59,** 263-285.

Sundaramoorthy, C., Jha, Girish. K., Pal, Suresh and Mathur, V.C. (2014). Market integration and volatility in edible oil sector in India. *J. Ind. Soc. Agric. Statist.*, **68,** 67-76.

Swanson, N.R. and White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Rev. Econ. Statist.*, **79,** 540-550.

Terui, N. and Van Dijk, H. (2002). Combined forecasts from linear and nonlinear time series models. *Inter. J. Forecast.*, **18,** 421-438.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer-Verlag, New York.

Tsangari, H. (2007). An alternative methodology for combining different forecast models. *J. Appl. Statist.*, **34,** 403-421.

Tsay, R.S. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.*, **84,** 231-240.

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

Zhang, G., Patuwo, B.E. and Hu, M.Y. (1998). Forecasting with artificial neural networks: The state of the art. *Inter. J. Forecast.*, **14,** 35-62.