

# Forecasting volatility in oil prices with a class of nonlinear volatility models: smooth transition RBF and MLP neural networks augmented GARCH approach

Melike Bildirici<sup>1</sup> · Özgür Ersin<sup>2</sup>

Received: 24 January 2014 / Published online: 23 July 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** In this study, the forecasting capabilities of a new class of nonlinear econometric models, namely, the LSTAR-LST-GARCH-RBF and MLP models are evaluated. The models are utilized to model and to forecast the daily returns of crude oil prices. Many financial time series are subjected to leptokurtic distribution, heavy tails, and nonlinear conditional volatility. This characteristic feature leads to deterioration in the forecast capabilities of traditional models such as the ARCH and GARCH models. According to the empirical findings, the oil prices and their daily returns could be classified as possessing nonlinearity in the conditional mean and conditional variance processes. Several model groups are evaluated: (i) the models proposed in the first group are the LSTAR-LST-GARCH models that are augmented with fractional integration and asymmetric power terms (FIGARCH, APGARCH, and FIAPGARCH); (ii) the models proposed in the second group are the LSTAR-LST-GARCH models further augmented with MLP and RBF type neural networks. The models are compared in terms of MSE, RMSE, and MAE criteria for in-sample and out-of-sample forecast capabilities. The results show that the LSTAR based and neural network augmented models provide important gains over the single-regime baseline GARCH models, followed by the LSTAR-LST-GARCH type models in terms of modeling and forecasting volatility in crude oil prices.

**Keywords** Volatility · Petrol prices · ARCH · STAR · Neural networks · LSTAR-LST-GARCH family

## 1 Introduction

The volatility of crude oil prices has received much attention recently because the crude oil is the most strategic and the most traded commodity in the world. Crude oil is traded internationally by many different players such as the oil producing nations, oil companies, individual refineries, oil importing nations, and speculators. Although crude oil price is basically determined by its supply and demand (Hagen 1994; Stevens 1995), it is also under the influence of many irregular events like stock levels, economic growth, political aspects, political instability, the decisions implemented by OPEC, and further psychological expectations of traders (Yu et al. 2008).

The volatility of oil prices is accepted to have important effects on economic activity. The fluctuations of the commodity market prices depend on the rise-and-fall of the oil price so that any sudden increase or decrease in oil prices cause economic slowdown and price fluctuations in other commodity prices. As a result, crude oil price forecasting is a very important field of research, and modeling/forecasting oil prices is hindered by its intrinsic difficulties such as the high volatility (Wang et al. 2005). As the crude oil spot price series are usually considered as a nonlinear and nonstationary time series, which is interactively affected by many factors, predicting crude oil price accurately is rather challenging (Yu et al. 2008).

Oil prices may not always adjust instantaneously to new information, on the other hand, low liquidity and infrequent trading in imperfect markets could lead to a delay in response to new information (McMillan and Speight 2006;

✉ Melike Bildirici  
bildiri@yildiz.edu.tr

<sup>1</sup> Department of Economics, Yıldız Technical University, Istanbul, Turkey

<sup>2</sup> Department of Economics, Beykent University, Istanbul, Turkey

Monoyios and Sarno 2002; Lee et al. 2008). In this perspective, there is a significant literature focusing on improving the capabilities of econometric models to model oil prices. A fraction of the studies investigate the path followed by the oil prices by utilizing various GARCH models. In addition, many studies focus on the economic factors that have had strong impacts on the increasing volatility especially for the periods with regime changes. One point to be taken into consideration is the fact that regime changes caused by many economic factors decrease the forecast capabilities of the single-regime models drastically. Furthermore, economic factors have important effects on the business cycles by disturbing the processes followed by economic time series. As an example, the 1st and 2nd oil shocks in 1974 and 1979 had significant impacts on the performances of the econometric models.

Consequently the traditional volatility models which do not take into consideration the regime switching characteristics of the factors such as oil shocks became obsolete in modeling volatility in petrol prices.

In this paper, the volatility in oil prices is aimed to be investigated. In accordance with this purpose, the oil prices will be modeled by improving both the conditional mean and the conditional variance with nonlinear time series and neural network models to achieve possible gains in forecasting and modeling capabilities. The business cycles in the economies reveal different dynamics under different regimes so that a traditional GARCH model becomes insufficient once volatility is encountered. The motive behind the usage of LSTAR structure (one of the nonlinear time series modeling) is to improve forecasting and modeling power of the GARCH models for policy purposes. In addition to the expectation of augmenting the GARCH models with LSTAR models, augmenting these models with MLP and RBF type neural networks would likely to bring increase in the forecasting capabilities. With this respect, our study aims to incorporate regime switching and neural networks with GARCH models. With this purpose, GARCH structure will be incorporated with LSTAR and multi-layer perceptron (MLP) and radial basis function (RBF) models. These three approaches consider the characteristics of oil prices which exhibit strong regime changes in addition to nonlinear volatility. Accordingly, at the first step, ST-GARCH models will be extended to LSTAR-LST-GARCH. Afterwards, by incorporating fractional integration and asymmetric power properties, four models will be evaluated: the LSTAR-LST-GARCH, LSTAR-LST-FIGARCH, LSTAR-LST-APGARCH, and LSTAR-LST-FIAPGARCH. At the second step, models will be augmented with RBF and MLP type neural networks. Hence, the LSTAR type, nonlinearity is introduced in the conditional mean and conditional variance processes of neural networks models to obtain the LSTAR-LST-GARCH-MLP and LSTAR-LST-GARCH-RBF models. Similarly, models are augmented with asymmetric

power terms and fractional integration. As a result, at the second step, MLP and RBF neural networks are introduced to obtain LSTAR-LST-GARCH-MLP and LSTAR-LST-GARCH-RBF models. Following the asymmetric power and fractional integration augmentations, models are denoted as LSTAR-LST-GARCH-MLP and LSTAR-LST-GARCH-RBF, LSTAR-LST-APGARCH-MLP and LSTAR-LST-APGARCH-RBF, LSTAR-LST-FIGARCH-MLP and LSTAR-LST-FIGARCH-RBF, and LSTAR-LST-FIAPGARCH-MLP and LSTAR-LST-FIAPGARCH-RBF models.

A literature review is given in Part 2. Econometric methodology is given in Part 3 where both the newly proposed LSTAR-LST-GARCH family and neural network-based LSTAR-LST-GARCH-NN family of models are evaluated. Empirical application to oil prices is given in Part 4 and concludes in Part 5.

## 2 Literature review

Hamilton (1983, 1985) are among the early studies that drew attention on the relationship between energy prices and showed that oil prices have important effects on the economy. Barone et al. (1998) suggested a semi-parametric modeling technique for oil price forecasting. Further, Alvarez et al. (2008) showed in their research that the random walk-type behavior in energy futures prices, thus the autocorrelation in oil prices diminishes over time. Adrangi et al. (2001) tested the presence of low-dimensional chaotic structure in crude oil, heating oil, and unleaded gasoline futures prices with their sample starting by the early 1980s. In their study, they pointed at chaotic structure and high persistence in the series investigated and concluded that chaotic structure and high persistence in the data would create strong implications for regulators and short-term trading strategies. Ewing and Maliq (2013) employs univariate and bivariate GARCH models to examine the volatility of gold and oil futures and showed that incorporating structural breaks is important for empirical analysis focusing on oil prices.

Ye et al. (2002, 2005, 2006) defined an econometric model for evaluating WTI prices by using variables such as the OECD petroleum inventory levels, relative inventories, and high and low-inventory variables. Lanza et al. (2005) analyzed crude oil and oil products' prices by using error correction models.

Further, many papers demonstrated that the prediction performance might be very poor if the traditional statistical and econometric models such as linear regressions are employed (Weigend and Gershenfeld 1994). The main reason behind is the phenomenon that the traditional statistical and econometric models are built on linear assumptions, which, as a result, fail to capture the

nonlinear patterns hidden in the crude oil price series (Yu et al. 2008). It is the fact that the oil prices may not always adjust instantaneously to the newly available information. Low liquidity and infrequent trading that occur under imperfect markets could cause to delays in response, following the availability of new information (McMillan and Speight 2006; Monoyios and Sarno 2002; Lee et al. 2008).

To overcome the difficulty in terms of forecast accuracy encountered in forecasting crude oil prices, Abramson and Finizza (1991) study is among the early studies that followed the neural network approach to model the crude oil market. Elsharkawy (1998) showed that radial basis function type neural network model had better forecast accuracy than the conventional methods in terms of predicting the oil formation volume factor, solution gas–oil ratio, oil viscosity, saturated oil density, under-saturated oil compressibility, and evolved gas gravity. Kaboudan (2001) showed that though both neural networks and genetic programming proved better results compared to the random walk to model oil prices, genetic programming provided superior results than the neural networks. Similarly, Mirmirani and Li (2004) showed that genetic algorithms and ANN models provided better results in forecasting oil prices. Tang and Hammoudeh (2002) derived a conclusion that shows the importance of both nonlinearity and heteroscedasticity in oil prices. They showed that by taking the GARCH effects into consideration, a nonlinear regression model with GARCH-type errors provide significant gains in modeling OPEC oil prices. Malliaris and Malliaris (2005) showed that the nonlinear models derived by the neural network provided superior forecasting in the majority of different oil types; namely for crude oil, heating oil, gasoline, and natural gas; with propane, the neural network gave the least accurate prediction. Yu et al. (2007) followed neural network modeling methods to forecast crude oil prices and showed that NN models provide significant gains in terms of different error criteria. Yu et al. (2007, 2008) proposed an efficient EMD-based neural network ensemble learning algorithm that uses feed-forward neural networks for modeling and forecasting world crude oil spot prices. Qunli et al. (2009) used radial basis functions (RBF) and showed that a RBF type neural network that benefits from wavelet transformations provided better results than the linear approaches in modeling monthly crude oil prices. Alizadeh and Mafinezhad (2010) used neural network models that utilized a predefined crisis variable to model and forecast Brent petrol prices and showed that the model is capable in forecasting both in normal and critical conditions. Bildirici and Ersin (2013) modeled the oil prices with newly introduced LST-LST-GARCH-MLP models.

### 3 Econometric methodology

Econometric modeling of volatility in the autoregressive conditional heteroscedasticity (ARCH) specification of conditional volatility gained significance especially following the important paper of Engle (1982). Further, the model is extended to generalized ARCH (GARCH) model by Bollerslev (1986), a model which has found many significant applications to capture the distributional aspects such as volatility clustering, heavy tails or excess kurtosis, and non-normal distribution. Additionally, the asymmetric power GARCH (APGARCH) model developed by Ding et al. (1983) further augments the model with power transformations without simple squared shocks and conditional variances as in the traditional GARCH models. Further, Baillie et al. (1996) and Bollerslev and Mikkelsen (1996) proposed the fractionally integrated GARCH (FIGARCH) model that encounters for the short-run dynamics of the conditional mean process modeled following ARMA process in the standard GARCH model. An important finding shown by Baillie et al. (1996) is that financial macroeconomic time series are subject to long memory characteristics such that volatility shows strong persistency. Chung (1999) and Conrad and Haag (2006) showed that the long-run persistence decays with hyperbolic rates. Tse and Tsui (1997) followed by Tse (1998) propose a combination of the FIGARCH and APGARCH model and the obtained FIAPGARCH model incorporates fractional integration with asymmetric power terms to capture the above-mentioned distributional aspects in financial time series.

Further, one of the expectations is that by integrating fractional integration with GARCH models, certain improvements in terms of forecasting accuracy and volatility modeling of oil prices could be achieved. Additionally, augmenting GARCH models with asymmetric power terms to obtain APGARCH models and merging fractional integration with APGARCH models to obtain FIAPGARCH models provide improvements in terms of volatility modeling. Considering the aspects such as intervention in oil prices and sudden changes in prices, FIGARCH and FIAPGARCH models provide important tools to analyze the finite persistence in oil prices though the models maintain assuming single-regime architectures without taking regime switches or regime transitions.

#### 3.1 LSTAR type nonlinearity in the conditional mean and variance

To model nonlinearity in GARCH processes, Franses and van Dijk (2000) evaluated the smooth transition GARCH (ST-GARCH), where regime changes are governed with

transition functions similar to the modeling and evaluation techniques of STAR models for the conditional mean processes developed by Terasvirta (1994). Lundberg and Terasvirta (1998) developed the STAR-ST-GARCH model that allows nonlinearity in both the conditional mean and the conditional variance processes of a time series. Chan and McAleer (2002, 2003) evaluated the statistical properties in context of estimation of STAR-GARCH models. In the study, we will allow models to follow STAR type nonlinearity both in the conditional mean and the conditional variance which are evaluated under LSTAR-LST-GARCH architecture. LSTAR-LST-GARCH models are LSTAR-LST-GARCH, LSTAR-LST-FIGARCH, LSTAR-LST-APGARCH, and LSTAR-LST-FIAPGARCH models and possess both ST-GARCH (Lundberg and Terasvirta 1998) and STAR-GARCH characteristics since both the conditional mean and the conditional variance are allowed to follow STAR type nonlinearity. ST-GARCH model shares similarities but have differences with the GJR-GARCH (Glosten et al. 1993) and TGARCH (Zakoian 1994) models in terms of the transition function since ST-GARCH models allow smooth transition functions instead of threshold function in defining regime changes.

Further, the artificial neural network ARCH process (ANN-GARCH) developed by Donaldson and Kamstra (1997) augments the GJR model with multi-layer perceptron-based neural network architecture with logistic squashing functions to capture nonlinearity by utilizing the universal approximation property (Cybenko 1989) of ANN models. In pursuit of these concepts, many papers developed neural network models. Lai and Wong (2001) contributed to the nonlinear time series modeling methodology by making use of single-layer neural networks; further, modeling of NN models for estimation and prediction for time series has important contributions governed by Weigend et al. (1991), Weigend and Gershenfeld (1994), White (1992), Hutchinson et al. (1994), Gencay and Liu (1997), Gencay and Stengos (1997, 1998) and Refenes et al. (1997) which contributed to financial analysis and stock market returns estimation, to pattern recognition and optimization. Dutta and Shekhar (1998) provided applications of neural networks for bond ratings. NN modeling methodology is applied successfully by Wang et al. (2005) for forecasting the value of a stock index. Bildirici and Ersin (2009) modeled NN-GARCH family models to forecast daily stock returns for short- and long-run horizons and they showed that GARCH models under NN architecture provide significant forecasting performance.

### 3.1.1 LSTAR-LST-GARCH model

Chan and McAleer (2002) discussed the STAR-GARCH model that has STAR type nonlinearity in the conditional

mean process. Franses and van Dijk (2000) discuss the STAR-STGARCH model that allows STAR type nonlinearity in both the conditional mean and the conditional variance and is developed based on the Terasvirta (1994) type LSTAR model with logistic transition function governing the dynamics of the transition between different regimes. In the paper, the LSTAR-LST-GARCH model will be extended to RBF type neural networks in addition to MLP type variants proposed in Bildirici and Ersin (2013).

At the first stage, the conditional mean process,  $y_t$ , is assumed to follow a LSTAR( $p$ ) process with two regimes as,

$$y_t = \left( \theta_{1,0} + \sum_{i=1}^r \theta_{1,i} y_{t-i} \right) \times (1 - H(\varepsilon_{t-1}; \gamma, n)) + \left( \theta_{2,0} + \sum_{i=1}^r \theta_{2,i} y_{t-i} \right) \times H(\varepsilon_{t-1}; \gamma, n) + \varepsilon_t. \tag{1}$$

The conditional variance follows a smooth transition LST-GARCH process,

$$\sigma_t^2 = \left( w_{1,0} + \sum_{p=1}^p \alpha_{1,p} \varepsilon_{t-p}^2 + \sum_{q=1}^q \beta_{1,q} \sigma_{t-q}^2 \right) \times (1 - H(\varepsilon_{t-1}; \gamma, n)) + \left( w_{2,0} + \sum_{p=1}^p \alpha_{2,p} \varepsilon_{t-p}^2 + \sum_{q=1}^q \beta_{2,q} \sigma_{t-q}^2 \right) \times H(\varepsilon_{t-1}; \gamma, n), \tag{2}$$

where, the transition between regimes is defined with a logistic function,

$$H(\varepsilon_{t-1}; \gamma, n) = \frac{1}{1 + e^{-\gamma(\varepsilon_{t-1} - n)}}. \tag{3}$$

In the LSTAR-LST-GARCH model given in Eq.'s (1) and (2) with the transition function Eq. (3), the speed of transition function is determined by the estimate of the  $\gamma$  parameter and the  $n$  is the threshold parameter. The logistic function  $H(\varepsilon_{t-1}; \gamma, n)$  is a twice differentiable continuous function bounded between [0, 1] lower and upper bounds for different values of the transition variable  $\varepsilon_{t-1}$  and its distance to the threshold  $n$  (see Bildirici and Ersin 2013). The transition is observed to be relatively slow for low values of  $\gamma$ , though the transition between regimes speeds up as  $\gamma$  takes larger values. The ARCH and GARCH parameter estimates  $\alpha_{1,p}, \beta_{1,q}$  approach to  $\alpha_{2,p}, \beta_{2,q}$  depending on the transition between regimes defined with  $H(\varepsilon_{t-1}; \gamma, n)$ . The stability condition,  $(\alpha + \beta) < 1$ , could vary for different values taken by the  $H(\varepsilon_{t-1}; \gamma, n)$ : as  $H(\varepsilon_{t-1}; \gamma, n) \rightarrow 1$  for  $\varepsilon_{t-1} > n$  innovations larger than the threshold, based on the regime dynamics the stability structure of the model approaches  $(\alpha_{1,p} + \beta_{1,q}) \rightarrow (\alpha_{2,p} + \beta_{2,q})$ . Further, for positive  $\varepsilon_{t-1} > n$ , as  $\varepsilon_{t-1} \rightarrow +$  values,  $H(\cdot) \rightarrow 1$ ; while, for negative values of the



transition variable and as long as  $\varepsilon_{t-1} < n$ ,  $\varepsilon_{t-1} \rightarrow -$  negative large values, transition function approaches  $H(\cdot) \rightarrow 0$  zero. The speed of transition is determined by the parameter  $\gamma$  and the above-mentioned characteristics of transition between volatility dynamics of the two regimes are highly influenced by the values taken by  $\varepsilon_{t-1}$ ,  $n$  and  $\gamma$ . The inflection point for the transition function occurs at  $H(\varepsilon_{t-1}; \gamma, n) = 1/2$  if the transition function is equal to the threshold parameter,  $\varepsilon_{t-1} = n$  where the stability condition holds if  $\{((\alpha_{1,p} + \alpha_{2,p}) + (\beta_{1,q} + \beta_{2,q}))/2\} < 1$ . Additionally, for very large values of the transition variable  $\gamma \rightarrow \infty$ ; and  $H(\cdot)$  transition function behaves like the identity function that gives sudden shifts between two regimes, i.e., for  $\gamma = \infty$  and  $\varepsilon_{t-1} < n$ ,  $H(\cdot) = 0$  and, for  $\varepsilon_{t-1} > n$ ,  $H(\cdot) = 1$ . As a result, the ST-GARCH process reduces to the TAR-TGARCH model for very large values of  $\gamma$ . Further, if  $\gamma = 0$ , transition function becomes  $H(\cdot) = 0.5$ , and hence the process reduces to a single-regime AR-GARCH( $p, q$ ) process.

### 3.1.2 LSTAR-LST-FIGARCH model

Fractionally Integrated GARCH (FIGARCH(1,  $d$ , 1)) model is developed under these findings by Bollerslev and Mikkelsen (1996) and Baillie et al. (1996) as an extension of the GARCH model to account for long memory. In this section, we will first evaluate fractional integration in a GARCH setting to evaluate long memory in conditional variance. Afterwards, smooth transition type nonlinearity setting will be introduced to the evaluated FIGARCH and FIAPGARCH models. The LSTAR-LST-FIGARCH model generalizes the LSTAR-LST-GARCH type nonlinearity to account for long memory in the conditional variance process,

$$\begin{aligned} (1 - \beta_i L) \sigma_t^2 = & \left( \omega_1 + \left( (1 - \beta_1 L) - (1 - \phi_1 L)(1 - L)^{d,1} \right) \right. \\ & \times \left. (|\varepsilon_{t-1}| - \theta_1 \varepsilon_{t-1})^2 \right) (1 - H(\varepsilon_{t-1}; \gamma, n)) \\ & + \left( \omega_2 + \left( (1 - \beta_2 L) - (1 - \phi_2 L)(1 - L)^{d,2} \right) \right. \\ & \times \left. (|\varepsilon_{t-1}| - \theta_2 \varepsilon_{t-1})^2 \right) (H(\varepsilon_{t-1}; \gamma, n)) \end{aligned} \quad (4)$$

with the transition function defined in Eq. (3). The range of the cluster of the volatility depends on the transition function and changes between  $H(\varepsilon_{t-1}; \gamma, n) = 0$  and  $H(\varepsilon_{t-1}; \gamma, n) = 1$ . Further,  $\gamma$  is the speed of transition parameter and  $\gamma > 0$  ensures that the transition between the regimes follows a nonlinear sigmoid type transition in modeling the dynamics of the conditional volatility. The constant term takes on values between  $\varphi = \omega/(1 - \alpha)$  and  $\varphi = \omega/(1 - \beta)$  based upon whether the conditional volatility is the regime dictated by  $H(\cdot) = 0$  and  $H(\cdot) = 1$ . Similar to the ST-GARCH model, the constant term ranges

between the extreme regimes, the level of conditional volatility will change in different regimes (For ST-FIGARCH models, readers are referred to Kılıç (2011) and Bildirici and Ersin (2013)).

### 3.1.3 LSTAR-LST-APGARCH model

The Asymmetric Power GARCH (APGARCH) model is developed by Ding et al. (1983). The model is based on different power transformations without simple squared shocks and conditional variances as in the traditional GARCH models. The STAR-ST-APGARCH model possesses nonlinear dynamics with smooth transition functions to allow different asymmetric power terms in two regimes with the following conditional variance process,

$$\begin{aligned} \sigma_t^{\delta,i} = & \left( \omega_{0,1} + \sum_{p=1}^p \alpha_{p,1} (|\varepsilon_{t-p,1}| - \gamma_{p,1} \varepsilon_{t-p,1})^{\delta,1} \right. \\ & \left. + \sum_{q=1}^q \beta_{q,1} \sigma_{t-q,1}^{\delta,1} \right) \times (1 - H(\varepsilon_{t-1}; \gamma, n)) \\ & + \left( \omega_{0,2} + \sum_{p=1}^p \alpha_{p,2} (|\varepsilon_{t-p,2}| - \gamma_{p,2} \varepsilon_{t-p,2})^{\delta,2} \right. \\ & \left. + \sum_{q=1}^q \beta_{q,2} \sigma_{t-q,2}^{\delta,2} \right) H(\varepsilon_{t-1}; \gamma, n). \end{aligned} \quad (5)$$

Similarly, the transition function is defined by Eq. (3) to obtain the LSTAR-LST-APGARCH model. Accordingly, the obtained model possesses such dynamics that both the conditional mean and the conditional variance follows nonlinear dynamics restricted to have two regimes between which the transition is defined by a smooth and continuously differentiable logistic function. The obtained model is defined as the *logistic smooth transition asymmetric power GARCH* model in which  $\delta$  represents the asymmetric power parameter to be estimated by nonlinear least squares or maximum likelihood methods as in the APGARCH methodology of Conrad et al. (2010). The estimation of the threshold parameter  $n$  and the  $\gamma$  parameter that defines the speed of transition is conducted through a grid search following the Terasvirta (1994).

### 3.1.4 LSTAR-LST-FIAPGARCH model

Following the LST-FIGARCH model structure, smooth transition fractionally integrated asymmetric power GARCH model denoted as LST-FIAPGARCH which is obtained by allowing the smooth transition type nonlinearity between two FIAPGARCH processes in two regimes as,

$$\begin{aligned}
 (1 - \beta_i L) \sigma_i^{\delta,i} = & \left( \omega_{0,1} + \left( (1 - \beta_1 L) - (1 - \phi_1 L)(1 - L)^{d,1} \right) \right. \\
 & \times \left. (|\varepsilon_{t-1}| - \theta_1 \varepsilon_{t-1})^{\delta,1} \right) \times (1 - (H(\varepsilon_{t-1}; \gamma, n))) \\
 & + \left( \omega_{0,2} + \left( (1 - \beta_2 L) - (1 - \phi_2 L) \right) \right. \\
 & \times \left. \text{left}(1 - L)^{d,2} \right) (|\varepsilon_{t-1}| - \theta_2 \varepsilon_{t-1})^{\delta,2} \\
 & \times (H(\varepsilon_{t-1}; \gamma, n)). \tag{6}
 \end{aligned}$$

As previously, the transition function  $H(\cdot)$  is defined as a logistic function bounded between 0 and 1. The obtained model is defined as the *logistic smooth transition fractionally integrated asymmetric power GARCH(LST-FIAPGARCH)* model in which  $\delta$  represents the asymmetric power parameter to be estimated (Conrad et al. 2010; Bildirici and Ersin 2013).

### 3.2 Neural network models, nonlinear GARCH models, and ANN augmented GARCH models

#### 3.2.1 Neural networks: an overview

Artificial neural networks models (ANN) are functional models that provide well-known approximation properties applied in many fields such as finance, medicine, and engineering. In economics and business literature, the early studies could be given as Tam and Kiang (1992), Do and Grudnitski (1992) which used neural networks for banking failure detection and residential property appraisals. Freisleben (1992) and Refenes et al. (1997) utilized ANN models for stock prediction. Hutchinson et al. (1994) showed that the learning networks could be used efficiently for pricing and hedging in securities markets. Studies such as Gencay and Stengos (1997, 1998), Gencay and Liu (1997), Kanas (2003), Kanas and Yannopoulos (2001), Shively (2003) and Bildirici and Ersin (2009) applied ANN models to stock market return forecasting and financial analysis.

The MLP model is evaluated as an important class of neural network models. MLP consists of a set of sensory units based on three layers, while a common application of such ANN model possesses mostly a single hidden layer. Hence, a MLP consists of the input layer, one or more hidden layers and an output layer.

Estimation in the LSTAR-LST-GARCH-MLP and LSTAR-LST-GARCH-RBF models is conducted with conjugant gradient-based backpropagation algorithm. The learning and model selection processes are gathered to improve forecast accuracy. During the learning process, *weight decay* is conducted to further improve the model eliminating the insignificant coefficients (Weigend et al. 1991; Bartlett 1997; Krogh and Hertz 1995). For details regarding weight decay in learning process, an

investigation is given by Gupta and Lam (1998). The algorithm cooperation and early stopping for NN-GARCH processes are given in Bildirici and Ersin (2009). The algorithm used in the study could be taken as estimating neural network models with LSTAR type nonlinear structures with different number of neurons in the hidden layer, this means estimating models with different architecture variations. Once the optimum architecture is selected, the model is re-estimated with early stopping and weight decay  $k$  times. To save CPU time,  $k$  is preferred as 20 in the study. Further, the number of neurons is allowed to vary ranging from 2 to 20 considering the sample size. Neurons are constrained as being logistic activation functions, which have similar transition properties as the LSTAR models in the hidden layer. The output layer is restricted to have linear functions. The models estimated are utilized for out-of-sample forecasting. Each model architecture is estimated  $k = 20$  times, and because there are eight different neural network-based model architecture to be estimated in the study, the total number of estimated models is 160; whereas, the selected 8 models (based on the lowest MSE error criteria) will be reported in the study. The methodology is as follows. Model estimation is gathered through utilizing backpropagation algorithm and the parameters are updated with respect to a quadratic loss function; whereas, the weights are iteratively calculated with *weight decay* method to achieve the lowest error. Alternative methods include Genetic Algorithms (Goldberg 1989) and second-order derivative-based optimization algorithms such as Conjugate Gradient Descent, Quasi-Newton, Quick Propagation, Delta-Bar-Delta, and Levenberg–Marquardt, which are fast and effective algorithms but may be subject to over-fitting (see Patterson 1996; Haykin 1994; Fausett 1994). In the study, we followed a two-step methodology. At the first step, all models were trained over a given training sample vis-à-vis checking for generalization accuracy in the light of MSE criteria in test sample. The approach is repeated for estimating each model for 100 times with different number of sigmoid activation functions in the hidden layer. To obtain parsimony, best model is further selected with respect to the AIC information criterion (see Faraway and Chatfield 1998). For estimating NN-GARCH models with early stopping combined with algorithm corporation, readers can refer to Bildirici and Ersin (2009).

The models below represent the architectures to be evaluated in the analysis for estimating the daily petrol prices in the application section. Each model possesses LSTAR type nonlinear structure in the conditional mean and conditional variance processes. The models are optimized to forecast the conditional variance of the petrol prices. Therefore, the LSTAR-LST-GARCH-MLP and

LSTAR-LST-GARCH-RBF structures of these models will be introduced as follows.

### 3.2.2 LSTAR-LST-GARCH-MLP model

Following Bildirici and Ersin (2013), a MLP neural network model hints a similar architecture with a logistic smooth transition GARCH (LST-GARCH) process. A *logistic smooth transition GARCH neural network model* is a LSTAR-LST-GARCH model with MLP type neural networks in each regime,

$$\begin{aligned} \sigma_t^2 = & \left( \omega_{0,1} + \sum_{p=1}^p \alpha_{p,1} \varepsilon_{t-p}^2 + \sum_{q=1}^q \beta_{q,1} \sigma_{t-q}^2 \right. \\ & \left. + \sum_{h=1}^h \xi_{1,h} \psi_1(z_{1,h} \lambda_{1,h}) \right) (1 - H(\varepsilon_{t-1}; \gamma, n)) \\ & + \left( \omega_{0,2} + \sum_{p=1}^p \alpha_{p,2} \varepsilon_{t-p}^2 + \sum_{q=1}^q \beta_{q,2} \sigma_{t-q}^2 \right. \\ & \left. + \sum_{h=1}^h \xi_{2,h} \psi_2(z_{2,h} \lambda_{2,h}) \right) H(\varepsilon_{t-1}; \gamma, n), \end{aligned} \tag{7}$$

where,  $n$  is the threshold,  $\gamma$  is the parameter that defines the speed of transition in the logistic transition function,

$$H(\varepsilon_{t-1}; \gamma, n) = \frac{1}{1 + e^{-\gamma(\varepsilon_{t-1} - n)}}. \tag{8}$$

MLP neural networks that possesses  $h$  hidden neurons in each  $i = 1, 2$  regimes are,

$$\psi_i(z_{i,h} \lambda_{i,h}) = \frac{1}{1 + \exp\left(-\left(\lambda_{i,h} + \sum_{d=1}^d \left[ \sum_{h=1}^h \lambda_{h,d,i} z_{t-d,i,h} \right] \right)\right)}, \tag{9}$$

where  $\psi_i(z_{i,h}, \lambda_{i,h})$  is a log-sigmoid activation function. The inputs are normalized and are defined as follows:

$$z_{t-d} = [\varepsilon_{t-d} - E(\varepsilon)] / \sqrt{E(\varepsilon^2)}, \tag{10}$$

$$\frac{1}{2} \lambda_{h,d,i} \sim \text{uniform} [-1, +1]. \tag{11}$$

The LSTAR-LST-GARCH-MLP model given in Eqs. (7)–(11) is a neural network augmented version of the LSTAR-LST-GARCH model given in Eq. (2).

The model proposed above will be augmented with asymmetric power term in the conditional variance to obtain LSTAR-LST-APGARCH-MLP model.

### 3.2.3 LSTAR-LST-APGARCH-MLP model

By augmenting the LSTAR-LST-GARCH-MLP model with asymmetric power terms, the LSTAR-LST-

APGARCH-MLP model is obtained. This model is a two-regime nonlinear model where both the conditional mean and the conditional variance follow a nonlinear process in the fashion of Terasvirta (1994) LSTAR model. The model also benefits from the well-known generalization properties of neural networks in the fashion of Cybenko (1989). The LSTAR-LST-APGARCH-MLP model allows smooth transition between two regimes of the conditional variance defined as a LST-APGARCH process with neural network augmentations as follows:

$$\begin{aligned} \sigma_t^{\delta,i} = & \left( \omega_{0,1} + \sum_{p=1}^p \alpha_{p,1} (|\varepsilon_{t-p,1}| - \theta_{p,1} \varepsilon_{t-p})^{\delta,1} \right. \\ & \left. + \sum_{q=1}^q \beta_{q,1} \sigma_{t-q,1}^{\delta,1} + \sum_{h=1}^h \xi_{h,1} \psi_1(z_{1,h} \lambda_{1,h}) \right) (1 - H(\varepsilon_{t-1}; \gamma, n)) \\ & + \left( \omega_{0,2} + \sum_{p=1}^p \alpha_{p,2} (|\varepsilon_{t-p,2}| - \theta_{p,2} \varepsilon_{t-p})^{\delta,2} \right. \\ & \left. + \sum_{q=1}^q \beta_{q,2} \sigma_{t-q,2}^{\delta,2} + \sum_{h=1}^h \xi_{h,2} \psi_2(z_{2,h} \lambda_{2,h}) \right) H(\varepsilon_{t-1}; \gamma, n). \end{aligned} \tag{12}$$

Accordingly, Eq. (12) is a hybrid model consisting of two regime LSTAR process in the conditional mean with residuals following a nonlinear neural network model for the conditional variance with multi-layer perceptrons in each regime of the conditional variance process. For the estimation and the statistical properties of the model, readers are referred to Bildirici and Ersin (2009, 2013) and to the neural network section of this paper.

### 3.2.4 LSTAR-LST-FIGARCH-MLP model

LSTAR-LST-FIGARCH-MLP model is a fractionally integrated volatility model augmented with two regime MLP neural networks in the conditional variance,

$$\begin{aligned} (1 - \beta_i L) \sigma_t^2 = & \left( \omega_{0,1} + \left( (1 - \beta_1 L) - (1 - \phi_1 L)(1 - L)^{d,1} \right) \right. \\ & \left. \times (|\varepsilon_{t-1}| - \theta_1 \varepsilon_{t-1})^2 + \sum_{h=1}^h \xi_{h,1} \psi_1(z_{1,h} \lambda_{1,h}) \right) \\ & \times (1 - H(\varepsilon_{t-1}; \gamma, n)) + \left( \omega_{0,2} + \left( (1 - \beta_2 L) \right. \right. \\ & \left. \left. - (1 - \phi_2 L)(1 - L)^{d,2} \right) (|\varepsilon_{t-1}| - \theta_2 \varepsilon_{t-1})^2 \right. \\ & \left. + \sum_{h=1}^h \xi_{h,2} \psi_2(z_{2,h} \lambda_{2,h}) \right) (H(\varepsilon_{t-1}; \gamma, n)). \end{aligned} \tag{13}$$

It should be noted that the LSTAR-LST-FIGARCH-MLP model reduces to LSTAR-LST-GARCH-MLP following the application of the restriction on the fractional integration parameter  $d = 0$ .

### 3.2.5 LSTAR-LST-FIAPGARCH-MLP model

The model is defined as follows:

$$\begin{aligned}
 (1 - \beta L)\sigma_t^{\delta,i} = & \left( \omega_{0,1} + \left( (1 - \beta_1 L) - (1 - \phi_1 L)(1 - L)^{d,1} \right) \right. \\
 & \times \left( |\varepsilon_{t-1}| - \theta_1 \varepsilon_{t-1} \right)^{\delta,1} + \sum_{h=1}^h \xi_{h,1} \psi_1(z_{1,h} \lambda_{1,h}) \Big) \\
 & \times \left( 1 - H(\varepsilon_{t-1}; \gamma, n) \right) + \left( \omega_{0,2} + \left( (1 - \beta_2 L) \right. \right. \\
 & \left. \left. - (1 - \phi_2 L)(1 - L)^{d,2} \right) \left( |\varepsilon_{t-1}| - \theta_2 \varepsilon_{t-1} \right)^{\delta,2} \right. \\
 & \left. + \sum_{h=1}^h \xi_{h,2} \psi_2(z_{2,h} \lambda_{2,h}) \right) H(\varepsilon_{t-1}; \gamma, n).
 \end{aligned}
 \tag{14}$$

The model assumes regime-dependent asymmetry based on  $\delta(i)$ , where  $i = 1, 2$  for the two regime structure. By applying  $d = 0$  restriction, Eq. (14) reduces to LSTAR-LST-APGARCH-MLP and additional restriction on the asymmetry parameters  $\delta(i) = 2$ , leads the model further reduce to LSTAR-LST-GARCH-MLP.

In fact, asymmetry is also introduced in the model through the ARCH terms. As a typical, the ARCH term is written as,  $(|\varepsilon_{t-1}| - \theta_i \varepsilon_{t-1})^{\delta,i}$  in each  $i = 1, 2$  regimes. To reduce the APGARCH process to baseline GARCH, one need also specify the ARCH term in a way to eliminate the deviations from the absolute innovations  $|\varepsilon_{t-1}|=0$  to obtain  $(-\theta_i \varepsilon_{t-1})^{\delta,i}$ . Further by restricting the model as  $\theta_i = (-1)$  so that  $\theta_i (\varepsilon_{t-1})^{\delta,i}$  followed by  $\delta = 2$ , the APGARCH model reduces to a GARCH process. One possibility is that, for modeling time series, it is possible to obtain different types of GARCH processes in each regime, i.e., a time series could follow a GARCH process in one regime, while following an APGARCH or FIAPGARCH process in the second regime. The study restricts the models to follow the same type of GARCH processes for simplicity. On the other hand, the fractional integration parameters are regime specific and allow different dynamics to be modeled simultaneously. The parameter  $d$  could be estimated as less than 0.5 in regime 1 and more than 0.5 in regime 2, suggesting different long memory dynamics and stationarity processes in each regime occurring below and above the threshold,  $n$ . As a result, assuming same GARCH structure in two regimes produce interesting findings and an approach that allows regime-wise comparative analysis (See Bildirici and Ersin 2013).

### 3.3 RBF neural network augmentations of the nonlinear GARCH models

The RBF neural network is represented as a composition of three layers of nodes; first, the input layer that feeds the input data to each of the nodes in the second or hidden

layer; the second layer that differs from other neural networks in that each node represents a data cluster which is centered at a particular point and has a given radius and in the third layer, consisting of one node (Bishop 1995).

Wright (2003) discusses the radial basis function interpolation and shows the developments in RBF networks. Liu and Zhang (2010) combined RBF neural network models with the Markov switching model to merge Markov switching Neural Network model based on RBF models. RBF neural network in their models are trained to generate both time series forecasts and certainty factors. Santos et al. (2010) developed a RBF-GARCH model that possesses a modeling structure that assumes a RBF type neural network in the conditional mean, where the residuals follow a GARCH process. Further, Coelho and Santos (2011) extended their RBF-GARCH approach and provided an application to Spanish energy pool prices and showed that RBF-GARCH approach provided significant improvement in future forecasts. It should be noted that, their approach is similar in one way to the STAR-GARCH approach of Chan and McAleer (2003) that assumes STAR type nonlinearity in the conditional mean process only. However, one important fact is that RBF-GARCH approaches of Santos et al. (2010) and Coelho and Santos (2011) benefit from different NN learning algorithms.

Our approach is differentiated than the above-mentioned studies in three ways. First, similar to the MLP-based models given in this paper, the proposed LSTAR-LST-GARCH-RBF models utilize neural network architectures in the conditional variance processes. It should be noted that heteroscedasticity is a strong factor that diminishes the forecast capabilities of the model. Second, models are estimated with neural network learning algorithms and the estimation of the models benefits from algorithm cooperation weight decay and early stopping. Third, our models follow STAR type division of the regression space both in the conditional mean and in the conditional variance with an expectation that this approach provides improvement in the modeling and forecasting capabilities as will be evaluated in Sect. 4.

#### 3.3.1 LSTAR-LST-GARCH-RBF model

LSTAR-LST-GARCH-RBF model is defined as follows:

$$\begin{aligned}
 \sigma_t^2 = & \left( \omega_{0,1} + \sum_{p=1}^p \alpha_{p,1} \varepsilon_{t-p}^2 + \sum_{q=1}^q \beta_{q,1} \sigma_{t-q}^2 \right. \\
 & \left. + \sum_{h=1}^h \xi_{1,h} \phi_1(\|z_t - \mu_1\|) \right) \left( 1 - H(\varepsilon_{t-1}; \gamma, n) \right) \\
 & + \left( \omega_{0,2} + \sum_{p=1}^p \alpha_{p,2} \varepsilon_{t-p}^2 + \sum_{q=1}^q \beta_{q,2} \sigma_{t-q}^2 \right. \\
 & \left. + \sum_{h=1}^h \xi_{2,h} \phi_1(\|z_t - \mu_2\|) \right) H(\varepsilon_{t-1}; \gamma, n).
 \end{aligned}
 \tag{15}$$



A Gaussian basis function for the hidden units given as  $\phi(x)$  for  $x = 1, 2, \dots, X$ , where the activation function is defined as Gaussian function,

$$\phi(x) = \exp\left(\frac{-\|x - \mu_j\|^2}{2\rho^2}\right) \tag{16}$$

with  $\rho$  defining the width of each function.  $z_t$  is a vector of lagged explanatory variables,  $\alpha + \beta < 1$  is essential to ensure stationarity. Networks of this type can generate any real-valued output, but in their applications where they have a priori knowledge of the range of the desired outputs, it is computationally more efficient to apply some nonlinear transfer function to the outputs to reflect that knowledge. The conditional variance is subject to smooth transition based on the logistic function,  $H(\varepsilon_{t-1}; \gamma, n) = 1 / (1 + e^{-\gamma(\varepsilon_{t-1} - n)})$ , where the speed of transition is given by  $\gamma$ . For the two regime model,  $i = 1, 2$ , the inputs are subject to,

$$z_{t-d} = [\varepsilon_{t-d} - E(\varepsilon)] / \sqrt{E(\varepsilon^2)} \tag{17}$$

$$\frac{1}{2} \lambda_{h,d,i} \sim \text{Uniform} [-1, +1]. \tag{18}$$

### 3.3.2 LSTAR-LST-APGARCH-RBF model

Radial basis functions are three layer neural network models with linear output functions and nonlinear activation functions defined as Gaussian functions in hidden layer utilized to the inputs in light of modeling a radial function of the distance between the inputs and calculated value in the hidden unit. The output unit produces a linear combination of the basis functions to provide a mapping between the input and output vectors.

$$\begin{aligned} \sigma_t^{\delta,i} = & \left( \omega_{0,1} + \sum_{p=1}^p \alpha_{p,1} (|\varepsilon_{t-p,1}| - \theta_{p,1} \varepsilon_{t-p})^{\delta,1} + \sum_{q=1}^q \beta_{q,1} \sigma_{t-q,1}^{\delta,1} \right. \\ & \left. + \sum_{h=1}^h \xi_{h,1} \phi_1(\|z_t - \mu_1\|) \right) (1 - H(\varepsilon_{t-1}; \gamma, n)) \\ & + \left( \omega_{0,2} + \sum_{p=1}^p \alpha_{p,2} (|\varepsilon_{t-p,2}| - \theta_{p,2} \varepsilon_{t-p})^{\delta,2} \right. \\ & \left. + \sum_{q=1}^q \beta_{q,2} \sigma_{t-q,2}^{\delta,2} + \sum_{h=1}^h \xi_{h,2} \phi_2(\|z_t - \mu_2\|) \right) H(\varepsilon_{t-1}; \gamma, n), \end{aligned} \tag{19}$$

where,  $i = 1, 2$  is the number of regimes. Similar to the LSTAR-LST-APGARCH-MLP model, the LSTAR-LST-APGARCH-RBF model nests several models. Equation (19) reduces to the LSTAR-LST-GARCH-RBF model if the power term  $\delta = 2$  and  $\theta_{p,i} = 0$ , to the LSTAR-GARCH-RBF model for  $\theta_{p,i} = 0$ , and to the LSTAR-GJR-

RBF model if  $\delta = 2$  and  $0 \leq \theta_{p,i} \leq 1$  restrictions are allowed. The model may be shown as LSTAR-TGARCH-RBF model if  $\delta = 1$  and  $0 \leq \theta_{p,i} \leq 1$ .

### 3.3.3 LSTAR-LST-FIAPGARCH-RBF model

LSTAR-LST-FIAPGARCH-RBF model is defined as follows:

$$\begin{aligned} (1 - \beta L) \sigma_t^{\delta,i} = & \left( \omega_{0,1} + \left( (1 - \beta_1 L) - (1 - \phi_1 L)(1 - L)^{d,1} \right) \right. \\ & \times (|\varepsilon_{t-1}| - \theta_1 \varepsilon_{t-1})^{\delta,1} + \sum_{h=1}^h \xi_{h,1} \phi_1(\|z_t - \mu_1\|) \left. \right) \\ & \times (1 - H(\varepsilon_{t-1}; \gamma, n)) + \left( \omega_{0,2} + \left( (1 - \beta_2 L) \right. \right. \\ & \left. \left. - (1 - \phi_2 L)(1 - L)^{d,2} \right) (|\varepsilon_{t-1}| - \theta_2 \varepsilon_{t-1})^{\delta,2} \right. \\ & \left. + \sum_{h=1}^h \xi_{h,2} \phi_2(\|z_t - \mu_2\|) \right) H(\varepsilon_{t-1}; \gamma, n), \end{aligned} \tag{20}$$

where,  $h$  is neurons defined with Gaussian function as in Eq. (16). The LSTAR-LST-FIAPGARCH-RBF model is a variant of the LSTAR-LST-APGARCH-RBF model with fractional integration. To obtain the model with short memory characteristics,  $d = 0$  restriction on the fractional integration parameter should be imposed. As a result, the model reduces to LSTAR-APGARCH-RBF model. Additionally, by applying  $d = 0$  with the restrictions discussed above, models with no fractional integration discussed above could be easily achieved. In addition to  $d = 0$  restriction, Eq. (20) reduces to LSTAR-LST-GARCH-RBF with the restriction on the asymmetry parameters  $\delta(i) = 2$  after eliminating the deviations from the absolute innovations with  $|\varepsilon_{t-1}| = 0$  and  $\theta_i = (-1)$ .

## 4 Econometric results

### 4.1 Data

In order to test forecasting performance of the above-mentioned models, Brent crude oil spot prices were used for oil price volatility. We take the daily data from January 20, 1986 to January 30, 2013, excluding public holidays, data are converted into daily returns by taking first differenced logarithms as  $y = \ln(P_t/P_{t-1})$ . In the process of model estimation, the sample is divided between training, test, and out-of-sample (forecasting) samples with the percentages of 80 %, 10 %, and 10 %, respectively. The descriptive statistics are reported in Table 1 below. Accordingly, the daily return series are subject to

**Table 1** Descriptives of Brent crude oil daily returns, January 20th, 1986 to January 30th, 2013

Mean	Median	Max	Min	SD	Skewness	Kurtosis	JB	SW	ARCH
8.13e−05	0.000320	0.08317	−0.176495	0.011105	−0.759616	17.742	63300.56 [0.0000]	0.91201 [0.000]	77.86521 [0.0000]

JB and the SW are the Jarque–Berra and Shapiro–Wilk normality tests

ARCH test is the ARCH-type heteroscedasticity test in the residuals of the AR(1) model selected by SIC information criterium. The probability values for the reported tests are given in brackets

**Table 2** GARCH family results

	1. GARCH	2. APGARCH	3. FIGARCH	4. FIAPGARCH
Cst( <i>M</i> )	0.000120 (0.00010935)	0.0002159** (0.000061)	0.0002579*** (0.00008093)	0.0002549*** (0.00008215)
Cst( <i>V</i> )	0.012561*** (0.0039790)	0.619456* (0.33203)	1.097080* (0.56766)	5.319243* (3.0849)
<i>d</i> -Figarch	–	–	0.474443*** (0.053529)	0.438994*** (0.042718)
ARCH	0.096773*** (0.018932)	0.077441*** (0.0087413)	0.219596*** (0.067838)	0.238516*** (0.071217)
GARCH	0.898413*** (0.017768)	0.929974*** (0.0082689)	0.590660*** (0.095018)	0.577437*** (0.089258)
APARCH (Gamma1)	–	0.166567*** (0.060379)	–	0.089858* (0.058840)
APARCH (Delta)	–	1.125234*** (0.10841)	–	1.764091*** (0.081130)
Log <i>L</i>	22,372.768	22,625.828	22,583.532	22,589.793
AIC	−6.466831	−6.539123	−6.527185	−6.528416
SIC	−6.462875	−6.532200	−6.521251	−6.620504
JB	3203.4	5381.8	2987.5	3821.6
Kurtosis	6.2693	7.2522	6.1596	6.5782

Standard errors are given in parentheses

\*, \*\*, and \*\*\* denote 10 %, 5 %, and 1 % significance levels

Log*L* is the Log-likelihood statistic. AIC and SIC denote the Akaike and Schwarz information criteria

leptokurtic distribution with the kurtosis statistic being 17.74 and skewness statistic calculated as −0.76. Jarque–Berra and Shapiro–Wilk tests suggest that the null hypothesis of normal distribution for daily returns can be rejected at the 5 % significance level. Further, the ARCH-type heteroscedasticity cannot be rejected for the daily returns series.

#### 4.2 Econometric results: model evaluation

At the first stage, the GARCH family models were taken as baseline models and are estimated for evaluation purposes. Results are given in Table 2. The models given in the table have different characteristics to be evaluated: namely, fractional integration, asymmetric power, and fractionally integrated asymmetric power models, namely, GARCH, APGARCH, FIGARCH, and FIAPGARCH models. By hybridization of two groups of nonlinear models, we obtained STAR-ST-GARCH models that allow for STAR type nonlinearity in both the conditional mean and variance.

The LSTAR-LST-GARCH models are reported in Table 3. The results show significant improvements of LSTAR-LST-GARCH models over their single-regime

variants reported in Table 2. The log-likelihood statistics are also high as AIC and SIC criteria report similar conclusions for the in-sample results.<sup>1</sup> Models have similar performances in the in-sample modeling. Further, models will be evaluated for out-of-sample forecasting capabilities with MSE, MAE, and RMSE statistics.

After allowing the GARCH processes to follow LST type nonlinearity, the dynamics are strikingly different in the light of the estimated parameters. In the LSTAR-LST-FIGARCH model, *d* parameters are estimated as 0.437 and 0.822 for regime 1 and 2, respectively, suggesting strong persistence in the second regime. For the LSTAR-LST-FIAPGARCH, after the inclusion of the asymmetric power terms, the *d* parameters are estimated as 0.44 and 0.45. The results also suggest that different conclusions could be derived due to the parametric specification of the analyzed GARCH models in addition to possible neglected nonlinearity.

The RBF and MLP type neural network augmented versions of the models will be analyzed. The model architectures of the proposed ANN models and their

<sup>1</sup> Models have similar performances in the in-sample modeling. For model performances, the models will be evaluated in terms of out-of-sample forecasting statistics.

**Table 3** The LSTAR-LST-GARCH family models: estimates

	LSTAR-LST-GARCH		LSTAR-LST-APGARCH		LSTAR-LST-FIGARCH		LSTAR-LST-FIAPGARCH	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
Cst(M)	-0.000038*** (0.2684e-017)	0.00056*** (0.5350e-005)	-0.0000276 *** (0.4935e-006)	0.00009*** (0.3389e-006)	0.0000757*** (0.3974e-006)	0.0000962*** (.4932e-005)	0.00050*** (0.0011653)	0.000099*** (7.6142e-006)
Cst(V)	-0.00002*** (.1627e-005)	0.00011*** (.1627e-005)	0.0036114*** (0.1555e-005)	0.001824*** (0.8884e-005)	0.031261*** (0.00014335)	0.031463*** (0.010639)	0.030331 (0.032575)	0.040000*** (0.0035798)
d-Figarch	-	-	-	-	0.437453*** (0.0067429)	0.821958*** (0.14722)	0.443713*** (0.2050)	0.450001*** (0.097121)
ARCH	0.070244*** (0.004856)	0.05804*** (0.02902)	0.157681 *** (0.019837)	0.154319* (0.09616)	0.098575*** (0.0048703)	0.160380*** (0.04430)	0.099359*** (0.009024)	0.100000*** (0.043737)
GARCH	0.955639*** (0.001756)	0.934448*** (0.20991)	0.914455*** (0.000928)	0.818426 *** (0.0061486)	0.402214*** (0.00286)	0.326375*** (0.1646)	0.401029*** (0.0662)	0.400000*** (0.20677)
APARCH (Gamma1)	-	-	0.576147*** (0.02203)	0.107065*** (0.041587)	-	-	0.006459*** (0.0003546)	0.010004*** (0.002871)
APARCH (Delta)	-	-	2.036870*** (0.022905)	2.006556*** (0.023098)	-	-	2.006473*** (0.29007)	2.000000*** (0.040708)
ARCH in mean	-	-	-0.001073 (0.00131)	-0.000818 (0.0389)	-0.000769*** (0.0001)	-0.00019** (0.0001)	-0.00008*** (0.000012)	-0.00006*** (0.00002)
Diagnostic tests								
LogL	23,310.192		16,613.846		13,393.367		13,189.971	
AIC	-12.653782		-15.734425		-16.621534		-16.934091	
SIC	-12.556380		-15.716555		-16.60163		-16.88337	
Kurtosis	8.9001		8.8815		7.864		6.2124	
ARCH (1-2)	1158.7 [0.00]		367.67 [0.00]		5.6985 [0.00]		3.3956 [0.00]	
ARCH (1-5)	470.78 [0.00]		164.13 [0.00]		2.6327 [0.00]		1.6217 [0.00]	

Standard errors are given in parentheses

\*, \*\*, and \*\*\* denote 10 %, 5 %, and 1 % significance levels

LogL is the Log-likelihood statistic. AIC and SIC denote the Akaike and Schwarz information criteria. For the ARCH type, remaining heteroscedasticity tests, calculated *p* values are reported in brackets

**Table 4** Neural networks augmented LSTAR-LST-GARCH-MLP and RBF models: architecture and training results

MLP-based ANN models and architectures				
Training statistics	1. LSTAR-LST-GARCH-MLP(5:2:2:6:1)	2. LSTAR-LST-APGARCH-MLP(5:2:2:11:1)	3. LSTAR-LST-FIGARCH-MLP(5:2:2:8:1)	4. LSTAR-LST-FIAPGARCH-MLP(5:2:2:4:1)
Training $\rho$	0.928664	0.951848	0.908285	0.905530
Test $\rho$	0.946714	0.952677	0.917540	0.912975
Training MSE	0.001284	0.0007911	0.001035	0.001139
TEST MSE	0.000989	0.0007660	0.000946	0.001122
Training algorithm (convergence)	BFGS (10)	BFGS (12)	BFGS (22)	BFGS (24)
RBF-based ANN models and architectures				
Learning results	1. LSTAR-LST-GARCH-RBF (5:2:2:28:1)	2. LSTAR-LST-APGARCH-RBF (5:2:2:26:1)	3. LSTAR-LST-FIGARCH-RBF (5:2:2:30:1)	4. LSTAR-LST-FIAPGARCH-RBF (5:2:2:27:1)
Training $\rho$	0.877056	0.930455	0.862266	0.879122
Test $\rho$	0.902086	0.939193	0.870629	0.887971
Training MSE	0.002152	0.001129	0.001516	0.001437
TEST MSE	0.001789	0.000974	0.001446	0.001423
Training algorithm (convergence)	RBFT	RBFT	RBFT	RBFT

$\rho$  denotes Pearson’s correlation statistic calculated for the targets and forecasts. MSE represents training and test sample mean squared errors. BFGS is the Broyden–Fletcher–Goldfarb–Shanno nonlinear optimization algorithm. The epoch shows the step number the algorithm converged. All models are restricted to have logistic activation functions in the hidden layer and identity activation functions in the output layers. Model architectures are given in parenthesis. As a typical, LSTAR-GARCH-MLP(5:2:2:6:1) model is a nonlinear model with 5 input variables in the input layer modeled as a 2 regime LSTAR process in the conditional mean with 2 regimes following LST-GARCH conditional variance processes passing through 6 neurons to the output layer connected to the output layer to produce 1 output

training results are reported in Table 4 in terms of MSE and  $\rho$  correlation statistics for the training and test samples.

Among the LSTAR-GARCH-NN models, the training and test MSE errors are calculated comparatively lower for the LSTAR-LST-GARCH-MLP models. Training MSE statistics for the LSTAR-LST-GARCH-RBF, LSTAR-LST-APGARCH-RBF, LSTAR-LST-FIGARCH-RBF, and LSTAR-LST-FIAPGARCH-RBF models are 0.001789, 0.000974, 0.001446, and 0.001423, respectively. On the other hand, the MSE statistics calculated for their MLP variants are 0.000989, 0.0007660, 0.000946, and 0.001122, respectively.

Radial Basis Function augmented versions of the LSTAR-LST-GARCH family models provided small deviation from the results obtained for their MLP variants in terms of training performances. As a typical, the highest training  $\rho$  is obtained as 0.93 for the LSTAR-LST-APGARCH-RBF and is higher than 3 out of 4 MLP-based models. Among the MLP-based models,  $\rho$  statistic is calculated as 0.95 for the LSTAR-LST-APGARCH-MLP. Overall, the MLP- and RBF-based models provide improvement over the GARCH and LSTAR-LST-GARCH

two regime variants. The results at this stage showed a general improvement of the RBF- and MLP-based models over the LSTAR-LST-GARCH models. To obtain conclusions, the out-of-sample forecasting capabilities of the models should be evaluated. One-step-ahead forecast results are given in Table 5.

The one-step-ahead forecast RMSE is the lowest for the LSTAR-LST-APGARCH-MLP (RMSE = 0.00000091651514) followed by the LSTAR-LST-FIGARCH-MLP (RMSE = 0.00000118743421), LSTAR-LST-FIAPGARCH-MLP (RMSE = 0.00000399624824), and LSTAR-LST-GARCH-MLP (RMSE = 0.00000474763099) models. Compared to the LSTAR-LST-GARCH models, the models provided significant improvement. Overall result is that, LSTAR-LST-GARCH-MLP models provided the highest one-step-ahead forecast accuracy followed by the LSTAR-LST-GARCH-RBF models.

Models are evaluated for their generalization capabilities in the larger out-of-sample horizons in terms of the MSE, RMSE, and MAE criteria. Results are given in Table 6 for a total of 16 models. The forecast horizon is selected as 2, 10, and 40 days ahead to evaluate the models’ performances in longer horizons.



**Table 5** One-step-ahead forecast results

	1. LSTAR-LST-GARCH-MLP(5:2:2:6:1)	2. LSTAR-LST-APGARCH-MLP(5:2:2:11:1)	3. LSTAR-LST-FIGARCH-MLP(5:2:2:8:1)	4. LSTAR- LST-FIAPGARCH-MLP(5:2:2:4:1)
MSE	0.0000000002254	0.0000000000084	0.0000000000141	0.0000000001597
MAE	0.00000332503469	0.00000069276605	0.00000093842087	0.00000301391256
MRSE	0.01202169456515	0.00035311744060	0.00074167725712	0.00942662726642
MRAE	0.07334809470262	0.01433928143676	0.02207744574166	0.07055172460999
$\rho$	0.93236330386609	0.99869776882616	0.99667226091915	0.90707791817108
RMSE	0.00000474763099	0.00000091651514	0.00000118743421	0.00000399624824
	1. LSTAR-LST-GARCH-RBF(5:2:2:28:1)	2. LSTAR-LST-APGARCH-RBF (5:2:2:26:1)	3. LSTAR-LST-FIGARCH-RBF (5:2:2:30:1)	4. LSTAR-LST-FIAPGARCH-RBF (5:2:2:27:1)
MSE	0.0000000003826	0.0000000002264	0.0000000002363	0.0000000002017
MAE	0.00000467340330	0.00000364627551	0.00000367533719	0.00000346294812
MRSE	0.01807061475042	0.01117553124815	0.0152727725031	0.01142473379382
MRAE	0.09949911206105	0.07941845168250	0.08899030858628	0.07972759121250
$\rho$	0.88209622896126	0.93220341293641	0.86380964347867	0.88099209829884
RMSE	0.00000618546684	0.00000475815090	0.00000486106984	0.00000449110231

*MSE* mean squared error, *MAE*, mean absolute error, *MRSE* mean relative absolute error, *MRAE* mean relative absolute error, *RMSE* root mean squared error.  $\rho$  shows the Pearson's correlation coefficient

In Table 6, models with the lowest RMSE, is denoted in bold within each model group for the above-mentioned out-of-sample forecast horizons. As a typical, for two days ahead, within the single regime GARCH type models, the lowest RMSE error is achieved with the FIGARCH model. At the second part, where the LSTAR-LST augmented two regime variants are evaluated, the lowest RMSE is achieved with the LSTAR-LST-GARCH model. Further, among the MLP neural networks augmented LSTAR-LST-GARCH-MLP models, the lowest RMSE is obtained with the LSTAR-LST-GARCH-MLP model. Additionally, among the RBF neural networks augmented variants, the lowest RMSE is achieved by the LSTAR-LST-GARCH-RBF model. Therefore, the models denoted with a RMSE value in bold represent the lowest RMSE achieved “within” the relevant model group only that consists of 4 different models only. Furthermore, a total of 16 different models, the baseline GARCH, their LSTAR-LST augmented two regime variants, the neural networks architecture and learning algorithm augmented models (i.e. RBF and MLP based 8 models) are ranked starting from the 1st towards the 4th model in terms of RMSE again. The models that take the 1st, 2nd, 3rd and 4th places are denoted accordingly. Following this procedure, “the best 4” are reported separately for 3 different forecast horizons i.e. for 2, 10, and 40 days to evaluate the estimated models for their forecast capabilities and to determine if the improvements exist not only in short horizons such as the 2 days ahead forecasts, but also in longer horizons.

The models in the first column have the GARCH architecture in common followed by its nonlinear LSTAR

and MLP, RBF augmentations. A significant decrease in RMSE, MAE, and MSE criteria is achieved as we move from single-regime GARCH model to LSTAR-LST-GARCH, LSTAR-LST-GARCH-MLP, and LSTAR-LST-GARCH-RBF. The RMSE reported for GARCH model is 0.000027 and 0.0000047 for the LSTAR-LST-GARCH, showing a 82.6 % decrease in the RMSE compared to the baseline GARCH for 2 days ahead forecasts. For the MLP augmented LSTAR-LST-GARCH-MLP, the RMSE is calculated as 0.00000134 which shows a 95 % decrease compared to the single-regime GARCH model. Hence, the LSTAR-GARCH model without neural networks provides improvement over the baseline GARCH for 2 days ahead. The LSTAR-GARCH-RBF model has a RMSE = 0.000005003, and performs almost equal to the LSTAR-LST-GARCH model.

For the baseline GARCH, RMSE = 0.0000519 and 0.000051 for 10 and 40 days ahead forecasts, while for the LSTAR-LST-GARCH, the RMSE statistics are calculated as 0.0000504 and 0.0000497. Accordingly, LSTAR-LST-GARCH performs better compared to the baseline GARCH, however, the improvement is limited. The results show that the predictive gains from the LSTAR-LST-GARCH suffer for longer horizons in the out-of-sample forecasts.

The LSTAR-LST-GARCH-RBF and MLP models aim at augmenting the forecasting capabilities of the LSTAR-LST-GARCH in long forecast horizons. For 10 days ahead, the RMSE is calculated as 0.0000016 for the LSTAR-LST-GARCH-MLP and is 0.0000047 for the LSTAR-LST-GARCH-RBF models. For 40 days ahead, the LSTAR-

**Table 6** Out-of-sample forecast statistics

	GARCH	APGARCH	FIGARCH	FIAPGARCH
2 days				
RMSE	0.00002665333000	0.00002516346558	<b>0.00002442130218</b>	0.00002661578479
MAE	0.00002603000000	0.00002454000000	0.00002377000000	0.00002602000000
MSE	0.0000000071040	0.0000000063320	0.0000000059640	0.0000000070840
	LSTAR-LST-GARCH	LSTAR-LST-APGARCH	LSTAR-LST-FIGARCH	LSTAR-LST-FIAPGARCH
2 days				
RMSE	<b>0.00000470106371 (3rd)</b>	0.00000752329715	0.00002251888097	0.00002191506331
MAE	0.00001306620000	0.00001233097000	0.00004970591220	0.00006855640160
MSE	0.0000000002210	0.0000000005660	0.00000000050710	0.00000000048027
	LSTAR-LST-GARCH-MLP	LSTAR-LST-APGARCH-MLP	LSTAR-LST-FIGARCH-MLP	LSTAR-LST-FIAPGARCH-MLP
2 days				
RMSE	<b>0.00000134164079 (1st)</b>	0.00000237907545 ( <b>2nd</b> )	0.00000622334315	0.00000671043963
MAE	0.00000130964370	0.00000226002346	0.00000557109994	0.00000618080882
MSE	0.0000000000180	0.00000000000566	0.00000000003873	0.00000000004503
	LSTAR-LST-GARCH-RBF	LSTAR-LST-APGARCH-RBF	LSTAR-LST-FIGARCH-RBF	LSTAR-LST-FIAPGARCH-RBF
2 days				
RMSE	<b>0.00000500299910 (4th)</b>	0.00000966488489	0.00000600999168	0.00000666933280
MAE	0.00000480348034	0.00000951497399	0.00000600977507	0.00000666288849
MSE	0.0000000002503	0.00000000009341	0.00000000003612	0.00000000004448
	GARCH	APGARCH	FIGARCH	FIAPGARCH
10 days				
RMSE	<b>0.00005195190083</b>	0.00005236410985	0.00005220153254	0.00005235456045
MAE	0.00004212000000	0.00004099000000	0.00004127000000	0.00004116000000
MSE	0.00000000269900	0.00000000274200	0.00000000272500	0.00000000274100
	LSTAR-LST-GARCH	LSTAR-LST-APGARCH	LSTAR-LST-FIGARCH	LSTAR-LST-FIAPGARCH
10 days				
RMSE	0.00005047771786	0.00011618950039	<b>0.00002413917977</b>	0.00005954829972
MAE	0.00003265000000	0.00011570000000	0.00002169000000	0.00004441000000
MSE	0.00000000254800	0.00000001350000	0.00000000058270	0.00000000354600
	LSTAR-LST-GARCH-MLP	LSTAR-LST-APGARCH-MLP	LSTAR-LST-FIGARCH-MLP	LSTAR-LST-FIAPGARCH-MLP
10 days				
RMSE	<b>0.00000162788206 (1st)</b>	0.00000544242593 ( <b>4th</b> )	0.00000561515806	0.00000611800621
MAE	0.00000157805413	0.00000442900285	0.00000462606822	0.00000508919208
MSE	0.00000000000265	0.00000000002962	0.00000000003153	0.00000000003743
	LSTAR-LST-GARCH-RBF	LSTAR-LST-APGARCH-RBF	LSTAR-LST-FIGARCH-RBF	LSTAR-LST-FIAPGARCH-RBF
10 days				
RMSE	<b>0.00000470850295 (2nd)</b>	0.00000558121850	0.00000477388730 ( <b>3rd</b> )	0.00000682129020
MAE	0.00000335796420	0.00000448305125	0.00000391396811	0.00000610994458
MSE	0.00000000002217	0.00000000003115	0.00000000002279	0.00000000004653

**Table 6** continued

	GARCH	APGARCH	FIGARCH	FIAPGARCH
40 days				
RMSE	0.00005116639522	0.00004876474136	0.00005105878964	<b>0.00004737087713</b>
MAE	0.00004242000000	0.00003854000000	0.00004182000000	0.00003692000000
MSE	0.00000000261800	0.00000000237800	0.00000000260700	0.00000000224400
	LSTAR-LST-GARCH	LSTAR-LST-APGARCH	LSTAR-LST-FIGARCH	LSTAR-LST-FIAPGARCH
40 days				
RMSE	0.00004972926704	0.00014869431731	<b>0.00013345411196</b>	0.00013667479651
MAE	0.00003266120910	0.00014800000000	0.00008539000000	0.00010590000000
MSE	0.00000000247300	0.00000002211000	0.00000001781000	0.00000001868000
	LSTAR-LST-GARCH-MLP	LSTAR-LST-APGARCH-MLP	LSTAR-LST-FIGARCH-MLP	LSTAR-LST-FIAPGARCH-MLP
40 days				
RMSE	<b>0.00000243926218 (1st)</b>	0.00000425440948 (2nd)	0.00000442605920 (3rd)	0.00000475078941
MAE	0.00000190278855	0.00000340258977	0.00000370453230	0.00000387005690
MSE	0.00000000000595	0.00000000001810	0.00000000001959	0.00000000002257
	LSTAR-LST-GARCH-RBF	LSTAR-LST-APGARCH-RBF	LSTAR-LST-FIGARCH-RBF	LSTAR-LST-FIAPGARCH-RBF
40 days				
RMSE	0.00000511370707	<b>0.00000449332839 (4th)</b>	0.00000565420198	0.00000606217783
MAE	0.00000394412289	0.00000375403891	0.00000460309756	0.00000507094872
MSE	0.00000000002615	0.00000000002019	0.00000000003197	0.00000000003675

Statistics are defined as follows. *RMSE* root mean squared error, *MAE* mean absolute error. Models are ordered from the lowest error criteria (for both RMSE and MAE) to the highest

The rank of each model is given in () brackets. Models are evaluated in terms of their capability in forecasting the conditional mean and variance separately

LSTAR-GARCH-RBF has a RMSE of 0.0000051, while the RMSE for the LSTAR-LST-GARCH-MLP is 0.0000024 and is almost halve of that obtained for LSTAR-LST-GARCH and its RBF variant. The results show that RBF-based model failed to provide significant improvements over the LSTAR-LST-GARCH, though MLP-based variant had the lowest RMSE, MSE, and MAE statistics. However, at this stage, the conclusions only show that by keeping the GARCH architecture constant, the MLP model showed significant forecast accuracy gains over the models in the first column. Note that, the models with FIGARCH, APGARCH, and FIAPGARCH architectures provided different results.

If an overlook is to be presented, as the forecast horizon is enlarged to 10 and 40 days ahead, the results provide a drastic improvement in longer horizons for the MLP-based models followed by the RBF-based variants. For comparative purposes, the models in each row model group are evaluated among themselves and the model with the lowest RMSE and MSE statistics is denoted in bold for 2, 10, and 40 days ahead forecasts. Additionally, the models are

ranked according to the RMSE statistics from lowest to highest to simplify the evaluation. For 2 days ahead, the FIGARCH model has the best forecast accuracy among the single-regime GARCH models (RMSE = 0.0000244). Among the two regime models, the LSTAR-LST-GARCH model has the best forecast capability (RMSE = 0.0000047). Among the MLP-based models, the LSTAR-LST-GARCH-MLP model has the highest forecast accuracy (RMSE = 0.00000134). Among the RBF-based variants, the LSTAR-LST-GARCH-RBF model has the best forecast accuracy (RMSE = 0.0000050). For 10 days ahead, the GARCH, the LSTAR-LST-FIGARCH, the LSTAR-LST-GARCH-MLP, and the LSTAR-LST-GARCH-RBF are the models with the lowest RMSE (and lowest MSE) among their own model group. For 40 days ahead, the FIAPGARCH, the LSTAR-LST-FIGARCH, the LSTAR-LST-GARCH-MLP, and the LSTAR-LST-APGARCH-RBF have the lowest RMSE statistics among their own model group. For different horizons, as the horizon moves from 2 to 40 days ahead MLP-based models showed significant improvement followed by the RBF-

based GARCH models. For simplicity, the first 4 models are to be reported. For 2 days ahead, the LSTAR-LST-GARCH-MLP is the 1st, while the LSTAR-LST-APGARCH-MLP is the 2nd, the LSTAR-LST-GARCH is the 3rd, and the LSTAR-LST-GARCH-RBF is the 4th. For 10 days ahead forecasts, the LSTAR-LST-GARCH-MLP is the 1st, the LSTAR-LST-GARCH-RBF is the 2nd, the LSTAR-LST-GARCH-RBF is the 3rd and the LSTAR-LST-APGARCH-MLP model is the 4th. For 40 days ahead, among 16 models estimated, the LSTAR-LST-GARCH-MLP takes the 1st place with the lowest RMSE (0.0000024), followed by the LSTAR-LST-APGARCH-MLP taking the 2nd place (RMSE = 0.00000425). LSTAR-LST-FIARCH-MLP takes the 3rd place (RMSE = 0.000004426) and the LSTAR-LST-APGARCH-RBF model is the 4th model (RMSE = 0.00000449).

Results supported the following conclusions for modeling and forecasting volatility in crude oil prices: (i) The nonlinear volatility models with STAR type nonlinearity namely, LSTAR-LST-GARCH family provided significant gains in terms of in-sample (one-step-ahead) forecasting accuracy and these models provided significant improvement over their single-regime GARCH variants. Further, for short horizons, the LSTAR-LST-GARCH family provided significant forecasting gains over their single-regime variants. (ii) RBF- and MLP-based neural networks augmentations of the LSTAR-LST-GARCH family models provided improved modeling capabilities for the crude oil prices that are subject to nonlinearity, asymmetry, and leptokurtic distribution. (iii) LSTAR-LST-GARCH-MLP and LSTAR-LST-GARCH-RBF showed gains in forecast capabilities which concentrate especially on the out-of-sample forecasting. Among the RBF and MLP augmented LSTAR-LST-GARCH family, the MLP-based models augmented the forecast accuracy of the LSTAR-LST-GARCH models followed by the RBF models. Additionally, the fractional integration and asymmetric power terms increased the forecast accuracy separately, though still the GARCH- and APGARCH-based MLP and RBF models provide satisfactory results, while the FIAPGARCH specification provided comparatively low gains in terms of forecast capabilities. It should be noted that the results are gathered for the daily Brent oil data set and cannot be generalized to all financial time series. Since certain financial time series such as the stock index returns possess strong asymmetric power effects and fractional integration, the LSTAR-LST-FIAPGARCH and its MLP/RBF variants may provide improved forecasting capabilities and therefore to obtain generalized results, the models should be evaluated for different financial time series in the developed and developing markets. The overall result of the empirical analysis suggests that nonlinear augmentations of

GARCH models for forecasting crude oil prices with the neural network architectures and nonlinear econometric techniques provide gains for the researchers and policy makers that aim at evaluating the paths followed by oil price time series.

### 4.3 Policy implications

The petrol price is an important variable for explaining business cycles and economic growth. As a result, petrol prices exhibits a large volatility not only through the channels of supply and demand, but also through political factors in addition to OPEC decisions. Volatility in petrol prices has strong impacts on economic variables such as economic growth, industrial production, and employment decisions in labor markets, not to mention its effect on the current account deficits and financial markets through various channels, since crude oil is also a financial commodity traded in spot and future markets. The results obtained in the study through the LSTAR-LST-GARCH-RBF and MLP models showed that, the adjustment process in oil prices do not occur instantaneously to new information. As shown by McMillan and Speight (2006) and Monoyios and Sarno (2002), low liquidity and infrequent trading in imperfect markets cause delays in the adjustment process after new information in financial markets. The results coincided with the fact that, increases in volatility are generally short lived; however, due to the persistence in oil prices, these effects may lead to long-lived effects in terms of persistence. The positive and large fractional coefficient estimates, in addition to large estimates of asymmetric power terms in both regimes justify the fact that shocks have relatively persistence effects; hence, within a political perspective, the governments should evaluate the oil prices and global factors very cautiously and simultaneously, policy makers should keep the interventions at the modest levels to avoid large fluctuations in petrol prices. Further, the long memory characteristics accelerate the expected temporary effects of these shocks, thus the persistency might increase the impacts of the oil shocks. Therefore, policy interventions should be kept at very modest levels to avoid large fluctuations.

If the results are to be summarized, the oil price possesses important characteristics such as nonlinearity, asymmetry, and transition effects, in addition to its fractionally integrated persistence effects. The policy maker and the researcher should evaluate the policies to be applied with great care. However, the nonlinear volatility models that incorporate fractional integration and power terms capture the data generating process more effectively; therefore, might be utilized important tools for policies. On the other hand, within a political perspective, the policies focusing on stabilization of volatility of this crucial



commodity may have destabilizing effects on the production and on the financial markets. Since crude oil prices are interlinked to various financial assets, this result translates itself to different derivatives and the economy, and this destabilization effect is largely under the influence of persistence in oil prices and also in the external shocks that oil prices are subject to. As a result, policies possible destabilizing effects without taking persistence into account result in additional effects in various markets. Secondly, the estimation sample in the study corresponded to a period with large oil shocks and economic crises periods. On the other hand, following the general methodology, the out-of-sample results are obtained for a period corresponding to year 2014, a relatively stable period. The forecasting practice in the paper showed significant gains in terms of forecast accuracy. On the other hand, through incorporating nonlinearity in the GARCH processes, the performance of these models would improve under unexpected changes in oil prices. Further, the utilization of the nonlinear models helps the policy maker by evaluating threshold characteristics of these models. As a result, the nonlinear models that incorporate neural networks' forecast capabilities with nonlinear econometric techniques are to be considered as tools for the investors and policy makers. However, the evaluation of the estimates provided by nonlinear models should always be evaluated with caution not to mention many external factors that lead to fluctuations and sudden/sharp changes in oil prices.

## 5 Conclusion

The study aimed at evaluating a new group of nonlinear models that combine the forecasting capabilities of MLP and RBF type neural networks with GARCH type volatility models and that augment LSTAR type nonlinear econometric time series models proposed by Luukkonen et al. (1988) and Terasvirta (1994). The proposed LSTAR-LST-GARCH-MLP and LSTAR-LST-GARCH-RBF family models aim at modeling not only the conditional mean processes but also the conditional variance simultaneously with STAR type nonlinearity, allowing the transition between the regimes to be captured with logistic transition functions. Accordingly, at the first stage, crude oil prices were modeled with baseline GARCH models with fractional integration and asymmetric power terms. At the second stage, LSTAR type nonlinear architecture was introduced to the baseline models to obtain the LSTAR-LST-GARCH models. At the third stage, LSTAR-LST-GARCH models were augmented with RBF and MLP neural networks to improve the modeling and forecasting capabilities of the researcher aiming at forecasting crude

oil prices. Accordingly, the models were compared in terms of MSE, RMSE, and MAE error criteria for in-sample and out-of-sample forecasts. The results showed that the LSTAR based and neural network augmented models provided significant gains in terms of modeling the daily returns of oil prices when compared with the results of the baseline GARCH family models. The results also showed that the LSTAR-LST-GARCH-RBF and LSTAR-LST-GARCH-MLP models provided significant gains in modeling petrol prices and in forecasting out-of-sample oil price returns. Following the findings of the paper, the future studies should aim at modeling different financial series that are subject to nonlinearity and volatility to test the forecasting capabilities of neural network algorithms and architectures.

The crude oil daily returns are evaluated as a result of their characteristics which could be classified as possessing strong nonlinearity, volatility defined with excess kurtosis. The nonlinearity inherited in crude oil prices has strong implications for regulators and short-term trading strategies. The oil prices possess important characteristics such as nonlinearity, asymmetry, transition effects, fractionally integrated, and persistence effects that should lead the policy maker and the researcher to evaluate the policies to be applied with great care; hence, the nonlinear volatility models that incorporate fractional integration and power terms capture the data generating process more effectively, therefore, provide important tools for policy makers.

**Acknowledgments** We thank anonymous referees for their useful comments.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Abramson B, Finizza A. Using belief networks to forecast oil prices. *Int J Forecast.* 1991;7(3):299–315.
- Adrangi B, Chatrath A, Dhanda KK, et al. Chaos in oil prices? Evidence from futures markets. *Energy Econ.* 2001;23(4):405–25.
- Alizadeh A, Mafinezhad K. Monthly Brent oil price forecasting using artificial neural networks and a crisis index. In: *Electronics and information engineering (ICEIE), international conference.* 2010;2:465–8.
- Alvarez-Ramirez J, Alvarez J, Rodriguez E. Short-term predictability of crude oil markets: a detrended fluctuation analysis approach. *Energy Econ.* 2008;30:2645–56.
- Baillie R, Bollerslev T, Mikkelsen HO. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J Econom.* 1996;74:3–30.

- Barone-Adesi G, Bourgoin F, Giannopoulos K. Don't look back. *Risk*. 1998;11:100–3.
- Bartlett PL. For valid generalization, the size of the weights is more important than the size of the network. In: Mozer MC, Jordan MI, Petsche T, editors. *Advances in neural information processing systems*. Cambridge: MIT Press; 1997. p. 134–40.
- Bildirici M, Ersin Ö. Forecasting oil prices: smooth transition and neural network augmented GARCH family models. *J Pet Sci Eng*. 2013;109:230–40.
- Bildirici M, Ersin Ö. Improving forecasts of GARCH family models with the artificial neural networks: an application to the daily returns in Istanbul Stock Exchange. *Expert Syst Appl*. 2009;36:7355–62.
- Bishop CM. *Neural networks for pattern recognition*. Oxford: Oxford University Press; 1995.
- Bollerslev T. Generalized autoregressive conditional heteroscedasticity. *J Econom*. 1986;31:307–27.
- Bollerslev T, Mikkelsen HO. Modeling and pricing long memory in stock market volatility. *J Econom*. 1996;73(1):151–84.
- Chan F, McAleer M. Estimating smooth transition autoregressive models with GARCH errors in the presence of extreme observations and outliers. *Appl Fin Econ*. 2003;13(8):581–92.
- Chan F, McAleer M. Maximum likelihood estimation of STAR and STAR-GARCH models: theory and Monte Carlo evidence. *J Appl Econom*. 2002;17(5):509–34.
- Chung C. Estimating the fractionally integrated GARCH model, Working paper. Taiwan: National Taiwan University; 1999.
- Coelho LS, Santos A. RBF neural network model with GARCH errors: application to energy price forecasting. *Elect Power Syst Res*. 2011;81(1):74–83.
- Conrad C, Haag BR. Inequality constraints in the fractionally integrated GARCH model. *J Fin Econom*. 2006;4(3):413–49.
- Conrad C, Rittler D, Rotfuss W. Modeling and explaining the dynamics of European Union allowance prices at high-frequency, Working Paper 0497. Heidelberg: University of Heidelberg, Department of Economics; 2010.
- Cybenko G. Approximations by superpositions of sigmoidal functions. *Math Control Signals Syst*. 1989;2:303–14.
- Ding Z, Granger CWJ, Engle RF. A long memory property of stock market returns and a new model. *J Emp Fin*. 1983;1:83–106.
- Do AQ, Grudnitski GA. Neural network approach to residential property appraisal. *Real Estate Apprais*. 1992;Dec:38–45.
- Donaldson RG, Kamstra M. An artificial neural network-GARCH model for international stock return volatility. *J Emp Fin*. 1997;4:17–46.
- Dutta S, Shekhar S. Bond rating: a non-conservative application of neural networks. *IEEE international conference on neural networks*. 1998;443–58.
- Elsharkawy A. Modeling the properties of crude oil and gas systems using RBF network. In: *SPE Asia Pacific oil and gas conference and exhibition*, 12–14 Oct 1998, Perth, Australia. 1998. doi:10.2118/49961-MS:1-12.
- Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*. 1982;50:987–1007.
- Ewing BT, Maliq F. Volatility transmission between gold and oil futures under structural breaks. *Int Rev Econ Financ*. 2013;25:113–21.
- Faraway J, Chatfield C. Time series forecasting with neural networks: a comparative study using airline data. *Appl Stat*. 1998;47(2):231–50.
- Faussett L. *Fundamentals of neural networks*. Englewood Cliffs: Prentice Hall; 1994.
- Franses PH, van Dijk D. *Non-linear time series models in empirical finance*. Cambridge: Cambridge University Press; 2000.
- Freisleben B. Stock market prediction with backpropagation networks. In: *Proceeding of the 5th international conference on industrial and engineering. Applications of applied intelligent and expert system*, 9–12 June 1992. Berlin: Springer; 1992. p. 451–60.
- Gencay R, Liu T. Nonlinear modelling and prediction with feedforward and recurrent networks. *Physica D*. 1997;108:119–34.
- Gencay R, Stengos T. Moving average rules, volume and the predictability of security returns with feed-forward networks. *J Forecast*. 1998;17:401–14.
- Gencay R, Stengos T. Technical trading rules and the size of the risk premium in security returns. *Stud Nonlin Dyn Econom*. 1997;2:23–34.
- Glosten LR, Jagannathan R, Runkle D. On the relation between the expected value and the volatility of the nominal excess return on stocks. *J Fin*. 1993;48:1779–801.
- Goldberg DE. Genetic algorithms and Walsh functions: part II, deception and its analysis. *Complex Syst*. 1989;3:153–71.
- Gupta A, Lam M. The weight decay backpropagation for generalizations with missing values. *Ann Oper Res*. 1998;78:165–87.
- Hagen R. How is the international price of a particular crude determined? *OPEC Rev*. 1994;18(1):145–58.
- Hamilton JD. Historical causes of postwar oil shocks and recessions. *Energy J*. 1985;6:97–116.
- Hamilton JD. Oil and the macroeconomy since world war II. *J Polit Econ*. 1983;91:228–48.
- Haykin S. *Neural networks*. Macmillan: A Comprehensive Foundation; 1994.
- Hutchinson JM, Lo AJ, Poggio T. A nonparametric approach to pricing and hedging derivative securities via learning networks. *J Fin*. 1994;49:851–89.
- Kaboudan MA. Compumetric forecasting of crude oil prices. *IEEE Congr on Evol Comp*. 2001;1:283–7.
- Kanas A. Non-linear forecasts of stock returns. *J Forecast*. 2003;22(4):299–315.
- Kanas A, Yannopoulos A. Comparing linear and nonlinear forecasts for stock returns. *Int Rev Econ Fin*. 2001;10(4):383–98.
- Kılıç R. Long memory and nonlinearity in conditional variances: a smooth transition FIGARCH model. *J Emp Fin*. 2011;18(2):368–78.
- Krogh A, Hertz J. A simple weight decay can improve generalization. In: Moody J, Hanson S, Lippmann R, editors. *Advances in neural information processing systems 4*. San Mateo: Morgan Kaufmann Publishers; 1995.
- Lai TL, Wong SP. Stochastic neural networks with applications to nonlinear time series. *J Am Stat Assoc*. 2001;96:968–81.
- Lanza A, Manera M, Giovannini M. Modeling and forecasting cointegrated relationships among heavy oil and product prices. *Energy Econ*. 2005;27(6):831–48.
- Lee YH, Liu HC, Chiu CL. Nonlinear basis dynamics for the Brent crude oil markets and behavioral interpretation: a STAR-GARCH approach. *Int Res J Fin Econ*. 2008;14:51–60.
- Liu D, Zhang L. China stock market regimes prediction with artificial neural network and Markov regime switching. *World Congr Eng (WCE)*. 2010;1:378–83.
- Lundbergh S, Terasvirta T. Modeling economic high-frequency time series with STAR-STGARCH models. *Stockholm School of Economics: Dep Econ*; 1998.
- Luukkonen R, Saikkonen P, Terasvirta T. Testing linearity against smooth transition autoregressive models. *Biometrika*. 1988;75:491–9.
- Malliaris M, Malliaris S. Forecasting energy product prices. In: *Proceedings of international joint conference on neural networks*. Montreal, Canada. 2005;5:3284–9.
- McMillan DG, Speight AEH. Nonlinear dynamics and competing behavioral interpretations: evidence from intra-day FTSE-100 index and futures data. *J Futur Mark*. 2006;26(4):343–68.

- Mirmirani S, Li HC. A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. *Adv Econom.* 2004;19:203–23.
- Monoyios M, Sarno L. Mean reversion in stock index futures markets: a nonlinear analysis. *J Futur Mark.* 2002;22:285–314.
- Patterson D. *Artificial neural networks*. Singapore: Prentice Hall; 1996.
- Qunli W, Ge H, Xiaodong C. Crude oil price forecasting with an improved model based on wavelet transform and RBF neural network. *Proc Int Forum Inf Technol Appl.* 2009;1:231–4.
- Refenes AP, Burgess AN, Bentz Y. Neural networks in financial engineering: a study in methodology. *IEEE Trans Neural Netw.* 1997;8:1222–67.
- Santos AAP, Coelho LS, Klein CE. Forecasting electricity prices using a RBF neural network with GARCH errors. In: 2010 international joint conference on neural networks (IJCNN), 18–23 July 2010. p. 1–8.
- Shively PA. The nonlinear dynamics of stock prices. *Q Rev Econ Fin.* 2003;43(3):505–17.
- Stevens P. The determination of oil prices 1945–1995. *Energy Policy.* 1995;23(10):861–70.
- Tam KY, Kiang MY. Managerial applications of neural networks: the case of bank failure predictions. *J Manag Sci.* 1992;38(7):926–47.
- Tang L, Hammoudeh S. An empirical exploration of the world oil price under the target zone model. *Energy Econ.* 2002;24:577–96.
- Terasvirta T. Specification, estimation, and evaluation of smooth transition autoregressive models. *J Am Stat Assoc.* 1994;89:208–18.
- Tse Y. The conditional heteroscedasticity of the yen-dollar exchange rate. *J Appl Econom.* 1998;13:49–55.
- Tse YK, Tsui AKC. Conditional volatility in foreign exchange rates: evidence from the Malaysian ringgit and Singapore dollar. *Pacific-Basin Fin J.* 1997;5:345–56.
- Wang SY, Yu L, Lai KK. Crude oil price forecasting with TEI@I methodology. *J Syst Sci Complex.* 2005;18(2):145–66.
- Weigend AS, Rumelhart DE, Huberman BA. Generalization by weight-elimination with application to forecasting. In: Lippmann RP, Moody J, Touretzky DS, editors. *Advances in neural information processing systems 3*. San Mateo: Morgan Kaufmann; 1991.
- Weigend AS, Gershenfeld NA. *Time series prediction: forecasting the future and understanding the past*. Reading: Addison-Wesley; 1994.
- White H. *Artificial neural networks: approximation and learning theory*. Oxford: Blackwell; 1992.
- Wright GB. *Radial basis function interpolation: numerical and analytical developments*, Thesis. University of Colorado. 2003.
- Ye M, Zyren J, Shore J. Forecasting short-run crude oil price using high and low-inventory variables. *Energy Policy.* 2006;34:2736–43.
- Ye M, Zyren J, Shore J. A monthly crude oil spot price forecasting model using relative inventories. *Int J Forecast.* 2005;21:491–501.
- Ye M, Zyren J, Shore J. Forecasting crude oil spot price using OECD petroleum inventory levels. *Int Adv Econ Res.* 2002;8:324–34.
- Yu L, Lai KK, Wang SY, et al. Oil price forecasting with an EMD-based multiscale neural network learning paradigm, vol. 4489., *Lecture Notes in Computer Science Berlin*: Springer; 2007. p. 925–32.
- Yu L, Wang S, Lai KK. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Econ.* 2008;30:2623–35.
- Zakoian JM. Threshold heteroscedastic models. *J Econ Dyn Control.* 1994;18:931–55.