Nonlinear volatility models in economics: smooth transition and neural network augmented GARCH, APGARCH, FIGARCH and FIAPGARCH models

Bildirici, Melike and Ersin, Özgür

Yildiz Technical University, Beykent University

January 2012

Nonlinear Volatility Models in Economics: Smooth Transition and Neural Network Augmented GARCH, APGARCH, FIGARCH and FIAPGARCH Models

Melike Bildirici
Prof.Dr., Yıldız Technical University,Department of Economics, Istanbul, Turkey. Corresponding Author. email:bildiri@yildiz.edu.tr, melikebildirici@gmail.com
Özgür Ömer Ersin
Assist.Prof.Dr., Beykent University, Department of Economics, Istanbul, Turkey.
Email:ozgurersin@beykent.edu.tr

Abstract

Recently, Donaldson and Kamstra (1997) proposed a class of NN-GARCH models which are extended to a class of NN-GARCH family by Bildirici and Ersin (2009). The study aims to analyze the nonlinear behavior and leptokurtic distribution in petrol prices by utilizing a newly developed family of econometric models that deal with these concepts by benefiting from both LSTAR type and ANN based nonlinearity. With this purpose, the study proposed several LSTAR-GARCH-NN family models. It is noted that the multilayer perceptron (MLP) neural network and LSTAR models have significant architectural similarities. Accordingly, linear GARCH, fractionally integrated FI-GARCH, asymmetric power APGARCH and fractionally integrated asymmetric power APGARCH models are augmented with a family of Neural Network models. The study has following contributions: i. STAR-GARCH and LSTAR-GARCH are extended to their fractionally integrated asymmetric power versions and STAR-ST-FIGARCH and STAR-ST-APGARCH, STAR-ST-FIAPGARCH models are developed and evaluated. ii. By extending these models with neural networks, LSTAR-LST-GARCH-MLP family models are developed and investigated. These models benefit from LSTAR type nonlinearity and NN based nonlinear NN-GARCH models to capture time varying volatility and nonlinearity in petrol prices. ANN augmented versions of LSTAR-LST-GARCH models are as follows: LSTAR-LST-GARCH-MLP, LSTAR-LST-FIGARCH-MLP, LSTAR-LST-APGARCH-MLP and LSTAR-LST-FIAPGARCH-MLP.

Empirical findings are collected as follows. *i.* To model petrol prices, fractionally integrated and asymmetric power versions provided improvements among the GARCH family models in terms of forecasting. *ii.* LSTAR-LST-GARCH model family is promising and show significant gains in out-of-sample forecasting. *iii*. MLP-GARCH family provided similar results with the LSTAR-LST-GARCH family models, except for the MLP-FIGARCH and MLP-FIAPGARCH models. *iv.* Volatility clustering, asymmetry and nonlinearity characteristics of petrol prices are captured most efficiently with the LSTAR-LST-GARCH-MLP models benefiting from forecasting capabilities of neural network techniques, whereas, among the newly developed models, LSTAR-LST-APGARCH-MLP model provided the best performance overall.

JEL Classification: G12, C32, C52, C53
Key Words: Volatility, Stock Returns, ARCH, Fractional Integration, MLP, Neural Networks

I. Introduction

Econometric modeling of volatility in financial market returns following the ARCH specification of conditional volatility of Engle (1982) and further extended to Generalized ARCH (GARCH) model in Bollersev (1986) has found many significant applications in light of modeling the distributional aspects such as volatility clustering, heavy tails, non-normal distribution. The Asymmetric GARCH model (AGARCH) developed by Engle (1990) aims modeling asymmetric effects of negative and positive shocks; whereas, negative and positive news have different effects on volatility. Accordingly, the Exponential GARCH (EGARCH) model developed by Nelson (1990) and the GJR-GARCH model developed independently by Glosten, Jaganathan and Runkle (1993) and by Zakoian (1994) are among the main modeling techniques followed in applied econometrics literature. The Asymmetric Power GARCH (APGARCH) model developed by Ding, Granger and Engle (1993) models are based on different power transformations without simple squared shocks and conditional variances as in the traditional GARCH models. Further, by showing that financial macroeconomic time series has long memory characteristics such that volatility show strong persistency, Baillie, Bollerslev and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996) proposed the Fractionally Integrated GARCH (FIGARCH) model that encounters for both the short-run dynamics of the conditional mean process modeled following ARMA process in the standard GARCH model and the long-run persistence that decays following hyperbolic rates and further investigated by Chung (1999) and Conrad and Haag (2006). Alternative specifications of FIGARCH models were further discussed by Giraitis, Robinson, and Surgailis (2004), Karanasos, Psaradakis, and Sola (2004), and Zaffaroni (2004). Further, Tse (1998) combines the FIGARCH model and APGARCH model and obtain the FIAPGARCH model. For a discussion and further analysis of the evolution of GARCH family models, we refer to Bollersev (2009), Zhang and Wei (2010).

Following the Zakoian (1991) Threshold GARCH (TGARCH) model that aims to capture asymmetric effects of negative and positive shocks, the intuition to capture different effects below and above a certain threshold is investigated. The other studies regarding the regime models were important in terms of smooth transition models. Franses and van Dijk (2000) noted the importance of ST-GARCH models. Hagerud (1997) and Gonzalez-Rivera (1998), Lundbergh and Teräsvirta (1998a), Anderson, Nam and Vahid (1999), Dufrénot, Marimoutou and Péguin-Feissolle (2002) developed the STGARCH model. Anè and Rangau (2006) combined the PGARCH model of Ding, Granger and Engle (1993), an extension of the GARCH family models, with RS-GARCH model and thus developed the RS-APGARCH model. Tse and Tsui (1997) determined the APGARCH model. Brooks et.al (2000) showed the leverage effect and the usefulness of including a free power term. Lundbergh and Terasvirta (1998) developed STAR-STGARCH models that allow nonlinearity in both conditional mean and conditional variance. Chan and McAleer (2002, 2003) have determined statistical properties in context of estimation of STAR-STGARCH family models. Busetti and Manera (2003) have used STAR-GARCH models to examine the market interactions in the Pacific Basin Region. Shively (2003) has examined nonlinear dynamics of stock prices for six developed economies using a three-regime threshold random walk model and found that stock prices are consistent with regime reverting process. McMillan (2003) has examined nonlinear predictability of UK Stock Returns. Ostermark et al. (2004) have used STAR type models for modelling Finnish Banking and Finance branch index. Narayan (2005) has examined properties of the stock prices for Australia and New Zealand and found that stock prices for both countries are nonlinear processes with unit root, consistent with the efficient market hypothesis. And most recently Hasanov and Omay (2008) have examined properties of the stock prices for Turkey and Greece and found that stock prices for both countries are nonlinear processes, and found out that nonlinear out of forecasting performance is better than the linear which is inconsistent with the efficient market hypothesis.

Futher, the ANN-GARCH (Artificial Neural Network ARCH) developed by Donaldson and Kamstra (1997) process 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. Further, following increasing advances with respect to asymmetry in volatility (Glosten et al., 1993; Zakoian, 1994; Nelson, 1991), ARCH (GARCH) family models are extended to different nonlinear modeling structures; specifically, regime switching (Cai, 1994; Hamilton and Susmel, 1994; Gray, 1996; Klaassen, 2002; Haas et al. 2004), threshold based regression space division with smooth sigmoid type continuous functions (Hagerud, 1997; Anderson et al., 1999; Gonzalez-Rivera, 1999; Lee and Degennaro, 2000; Lundberg and Terasvirta, 1998) and artificial neural networks (Donaldson and Kamstra, 1997; Bildirici and Ersin, 2009).

Models with STAR type nonlinearity are evaluated in Part II. Models with neural network based architectures are discussed in Part III. Empirical results are given in Part IV. Part V. conludes.

.

II. Models

Time series models may be subject to follow nonlinear processes in different proportions, in the conditional mean and/or in the conditional variance. Accordingly, models investigated in the study are divided into groups by possessing nonlinearity in the conditional mean, variance, or none (or both) in the conditional variance and mean.

In the study, first group of models are linear GARCH, fractionally integrated FI-GARCH, Asymmetric Power APGARCH (Ding, Granger and Engle; 1993) and the fractionally integrated FIAPGARCH models (Baillie, Bollerslev and Mikkelsen; 1996). These models are taken as the baseline family of models.

Models with STAR type nonlinearity in the conditional mean will be investigated under the second group. The STAR-GARCH model (Lundberg and Terasvirta, 1998; Chan and McAleer, 2001) allows the conditional mean to follow STAR type nonlinearity. In the study, STAR-GARCH model is extended to FIGARCH, APGARCH and FIAPGARCH processes and evaluated models under this group are LSTAR-GARCH, LSTAR-FIGARCH, LSTAR-APGARCH and LSTAR-FIAPGARCH models.

In the third group, we allowed models to follow STAR type nonlinearity both in the conditional mean and the 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 is allowed to follow STAR type nonlinearity[1].

Multi-Layer Perceptron type neural networks are commonly applied to economic time series in the literature. MLP-GARCH models are the fourth group of models which follow a similar modeling methodology as given for the STAR-GARCH models. Accordingly, the conditional mean is modeled with MLP with error terms following GARCH process. Models in this group are MLP-GARCH, MLP-FIGARCH, MLP-APGARCH and MLP-FIAPGARCH. One point that cannot be overlooked is that MLP-GARCH models are different than the NN models as discussed by Donaldson and Kamstra (1997). It should be noted that the methodology followed in this group is different in the sense that, MLP-GARCH model allows conditional mean to have MLP as the STAR-GARCH model that has STAR process in the conditional mean; therefore, neural network modeling techniques discussed in

[1] ST-GARCH model shares similarities but have differences with the GJR-GARCH (Glosten, Jagannathan and Runkle; 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.

Bildirici and Ersin (2009); model selection, estimation with ANN learning algorithms and algorithm cooperation and weight decay are not applied.

The fifth group is the neural network augmented versions of the second group, LSTAR-GARCH models to obtain LSTAR-GARCH-MLP models.

The sixth group is LSTAR-LST-GARCH-MLP model group that augments the third group with MLP architecture and modeling techniques to improve the generalization power of LSTAR-LST-GARCH models.

Following Bildirici and Ersin (2009), estimation of LSTAR-GARCH-MLP and LSTAR-LST-GARCH-MLP models is conducted with conjugant-gradient based back-propagation algorithm (for a review, see: Bishop; 1995). The learning and model selection processes are gathered to improve forecast accuracy as follows. Neural networks are estimated for n number of models with optimization conducted simultaneously in the training and test samples. Optimization is early stopped at the epoch at which MSE in the test sample starts to increase though still continues to decrease in the training sample; the model with the lowest MSE is selected. During learning, weight decay in the output layer and hidden layer is utilized to eliminate the insignificant coefficients (Weigend, Rumelhart and Huberman, 1991; Bartlett, 1997; Krogh and Hertz; 1995). For details regarding weight decay in learning process, an investigation is given by Gupta and Lam (1998). In total, each model is estimated with different architecture variations in terms of number of neurons. Number of estimated models of each architecture type, n is selected as 20 for saving CPU time. Only the best model is reported for each model architecture. Models to be compared are allowed to have their number of neurons to range between 3 to 10 considering the sample size. Neurons are constrained as being logistic activation functions in the hidden layer and linear function in the output layer. Best models with the lowest error criteria such as MSE or RMSE are selected. The selected models are further utilized for out-of-sample forecasting. Therefore, since each model architecture is estimated n=20 times, and since there are 8 different neural network based model architecture to be estimated; namely, LSTAR-GARCH-MLP, LSTAR-APGARCH-MLP, LSTAR-FIGARCH-MLP, LSTAR-FIAPGARCH-MLP, LSTAR-LST-GARCH-MLP, LSTAR-LST-APGARCH-MLP, LSTAR-LST-FIGARCH-MLP and LSTAR-LST-FIAPGARCH-MLP models, total number of estimated models are 160; whereas, the best 8 model is taken into consideration[2].

In the next section, GJR-GARCH, ST-GARCH, STAR-GARCH and STAR-ST-GARCH models will be investigated. The threshold principle of GJR-GARCH will provide basis for STAR type nonlinearity which will be further extended to MLP models.

[2] 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 2nd order derivative based optimization algorithms such as Conjugate Gradient Descent, Quasi-Newton, Quick Propagation, Delta-Bar-Delta and Levenberg-Marquandt, 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. Firstly, all models were trained over a given training sample vis-à-vis checking for generalization accuracy in 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 parsimonic models, 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 are referred to Bildirici and Ersin (2009).

i. ST-GARCH Model

GJR-GARCH model, developed by Glosten, Jagannathan and Runkle (1993), is based on the modeling of conditional variance with varying responses to negative and positive lagged innovations with respect to an indicator function. GJR-GARCH model is represented as

$$\sigma_t^2 = w + \alpha\varepsilon_t^2 + I(\varepsilon_{t-1})\gamma\varepsilon_t^2 + \beta\sigma_{t-1}^2 \tag{1}$$

where, $I(\varepsilon_{t-1})$ is an indicator function being $I(\varepsilon_{t-1})$=0 if $\varepsilon_{t-1} \geq 0$ and $I(\varepsilon_{t-1})$=1 otherwise. The asymmetry introduced with the γ and the indicator function *I*(.) is called as "the leverage effect"; hence, γ is typically estimated to be positive so that the volatility is increasing proportionately more after negative shocks compared to the impact of the positive shocks. The identity function will be augmented with the logistic function and GJR structure will provide a basis for ST-GARCH models.

The Smooth Transition Autoregressive (STAR) model further developed by Luukkonen *et al.* (1988), Granger and Terasvirta (1993) and Terasvirta (1994) aim nonlinear modeling of the conditional mean by introducing smooth transition between regimes of autoregressive processes based on logistic and exponential functions belonging to squashing functions of neural network models. In STAR methodology (Terasvirta, 1994), by taking logistic and exponential functions as transition functions, LSTAR and ESTAR models are obtained. Hagerud (1997) and Gonzalez-Rivera (1998) proposed the ST-GARCH model that allows smooth transition between the α and Υ, coefficients of lagged squared error terms of the GJR-GARCH model. A convenient way to formulate the GJR,

$$\sigma_t^2 = w + (1 - I|\varepsilon_{t-1} > 0|)\alpha_1\varepsilon_{t-1}^2 + (I[\varepsilon_{t-1} > 0])\Upsilon_t\varepsilon_{t-1}^2 + \beta_t\sigma_{t-1}^2 \tag{2}$$

if the *I*(.) indicator function is replaced with the *F*(.) logistic transition function, the Logistic Smooth Transition GARCH (LSTGARCH(1,1)) model is obtained as

$$\sigma_t^2 = w + \left(1 - F(\varepsilon_{t-1})\right)\alpha_1\varepsilon_{t-1}^2 + F(\varepsilon_{t-1})\Upsilon_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \tag{3}$$

where, the transformation function *F* is defined as

$$F(\varepsilon_{t-1}) = \frac{1}{1 + e^{\left(-\theta\varepsilon_{t-1}\right)}}. \tag{4}$$

The logistic function is bounded between [0 , 1] and the transition between the regimes occurs from negative to positive values, $\theta > 0$ has non-negativity constraint and the logistic transition function *F* is a monotonic and increasing function of ε_{t-1}. As ε_{t-1} increases from negative values to positive values the impact of $\varepsilon^2{}_{t-1}$ moves proportionately from α_1 to Υ_1. If θ is positive and large enough, LSTGARCH model transforms into the GJRGARCH model.

By replacing the the logistic transformation function with the exponential function Hagerud (1997) proposedthe Exponential Smooth Transformation GARCH (ESTGARCH) model. ESTGARCH(1,1), differentiated from the LST-GARCH model with the exponential function,

$$F(\varepsilon_{t-1}) = \left(1 - e^{\left(-\theta\varepsilon_{t-1}^2\right)}\right) \tag{5}$$

As a result of formulating (3) with the exponential function given in Eq. (5), the dynamics of the conditional variance is modeled depending on the size of shocks. This type of nonlinear GARCH formulation is symmetric in terms of the sign of the shocks. The most significant reason of using the exponential function instead of logistic function is the allowance of $F(\varepsilon_{t-1})$ to vary between the boundaries of [0, 1] as $\varepsilon^2{}_{t-1}$ varies between the extreme values.

It is noted that, in the ST-GARCH models presented above following the models of Hegerud(1997), Gonzales-Rivera(1998) and Lee and Degennaro (2000), the smooth transition is introduced in ARCH parameters. Following Anderson(1999) and Lundbergh and Terasvirta

(2002), ST-GARCH model may be modeled by allowing the intercept, ARCH and GARCH terms to follow smooth transition between regimes as

$$\sigma_t^2 = \left(1 - F\left(\varepsilon_{t-1}, \theta\right)\right) + \left(w + \beta^* \sigma_{t-1}^2 + \alpha^* \varepsilon_{t-1}^2\right) F\left(\varepsilon_{t-1}, \theta\right) \tag{6}$$

where, the parameters of the second regime is denoted with an asterix. The conditional volatility may depend both on the size and sign of the shocks on ε_{t-1}. Relative effects of negative and positive shocks with equal magnitude depend on the amplitude of the conditional volatility of shocks so that a negative shock may produce a larger shock compared to the one that a positive shock with similar size could have produced. Negative surprises with large amplitudes may show leverage effects and may lead to volatility with comparatively larger size compared to the positive surprises. (Taylor J.W., 2004).

ii. ST-FIGARCH Model

The ARCH and GARCH models, developed by Engle (1982) and Bollerslev (1986) respectively, are short memory processes resulting from the fact that the response of a shock on the conditional variance decreases at an exponential rate. On the other hand, the conditional volatility of financial market returns may change slowly over time as a result of long memory characteristics of financial series. Consequently, the autocorrelation functions may decay at a hyperbolic rate[3].

Fractionally Integrated GARCH (FIGARCH(1, *d,* 1)) model is developed under these findings by Bollersev and Mikkelsen(1996) and Baillie, Bollerslev, and Mikkelsen(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.

Assume that a time series following a random walk process in its conditional mean and its conditional variance, $\sigma_t^2 = \mathrm{Var}\left(\varepsilon_t \,\middle|\, \Omega_{t-1}\right)$, where the information set up to time t-1 is denoted as Ω_{t-1}, follows a FIGARCH(1,d,1) process

$$\left(1 - \beta L\right)\sigma_t^2 = \alpha + \left(\left(1 - \beta L\right) - \left(1 - \phi L\right)\left(1 - L\right)^d\right)\left(\left|\varepsilon_{t-1}\right| - \gamma\varepsilon_{t-1}\right)^2 \tag{7}$$

or alternatively,

$$\sigma_t^2 = \alpha + \beta\sigma_{t-1}^2 + \left(\left(1 - \beta L\right) - \left(1 - \phi L\right)\left(1 - L\right)^d\right)\left(\left|\varepsilon_{t-1}\right| - \gamma\varepsilon_{t-1}\right)^2 \tag{8}$$

where, z_t is assumed to be normally distributed *N(0,1)* white noise process,

$$z_t \sim N\left(0,1\right) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{z_t^2}{2}\right) \tag{9}$$

FIGARCH(1, *d,* 1) model nests the GARCH model if $d = 0$ and the IGARCH model of Engle and Bollerslev (1986) if $d = 1$, the estimated fractional integration parameter. The fractional

[3] Bollerslev and Mikkelsen (1996) develop the necessary conditions for FIGARCH model and note that for a well defined FIGARCH model, all the coefficients in the infinite ARCH representation must be non-negative (see: Bollerslev andMikkelsen (1996, p. 159). FIGARCH models are further discussed in Nelson and Cao (1992) and Conrad and Haag (2006), following these studies, nonnegativity constraints on parameters of FIGARCH processes are relaxed and shown that, for p=2, the second lag of conditional variance can become negative. Futher, Conrad and Haag (2006) allow conditions so that even if all parameters are negative (apart from d), the conditional variance can be nonnegative for FIGARCH models following the inequality constraints of Conrad and Haag (2006).

integration parameter d is 0<d<1 and as $d \rightarrow 0$ $(d \rightarrow 1)$ the model has short memory (long memory) characteristics. For alternative specifications of FIGARCH model, readers are referred to Karanasos, Psaradakis, and Sola (2004), Giraitis, Robinson, and Surgailis (2004) and Zaffaroni (2004).

The ST-FIGARCH model which generalizes the ST-GARCH type nonlinearity to account for fractional integration is represented as follows,

$$\sigma_t^2 = \omega + \left(1 - F\left(\varepsilon_{t-s}, \gamma\right)\right)\alpha\sigma_{t-1}^2 + \beta F\left(\varepsilon_{t-s}, \gamma\right)\sigma_{t-1}^2$$
$$+ \left[\left[1 - \alpha L\left(1 - F\left(\varepsilon_{t-s}, \gamma\right)\right) - \beta L F\left(\varepsilon_{t-s}, \gamma\right)\right] - \left(1 - \phi L\right)\left(1 - \phi L\right)^d\right] u_t^2 \quad (10)$$

for $\gamma \neq 0$ the width of the volatility clusters and, α and β characterizes the dynamics of the conditional volatility. The range of the cluster of the volatility changes between $F(.) = 0$ and $F(.) = 1$. The constant term takes on values between $\varphi = \omega / (1 - \alpha)$ and $\varphi = \omega / (1 - \beta)$ based upon whether the conditional volatility is is the regime dictated by $F(.) = 0$ and $F(.) = 1$. Accordingly, since, in the ST-GARCH model, the constant term ranges between the extreme regimes, the level of conditional volatility will change in different regimes (Kılıç, 2010). If the transition function $F(.)$ is logistic function

$$F(\varepsilon_{t-1}) = \frac{1}{1 + e^{\left(-\theta\varepsilon_{t-1}\right)}} \quad (11)$$

the model becomes logistic smooth transition FIGARCH (LST-FIGARCH) model.

iii. ST-FIAPGARCH Model

Tse (1998) introduced the FIAPGARCH model which combines long memory property of Baillie, Bollerslev, and Mikkelsen (1996) FIGARCH model with Asymmetric Power GARCH (APGARCH) model of Ding, Engle, and Granger (1993) by extending the FIGARCH model to account for different asymmetric dynamics. Accordingly, the fractionally integrated APGARCH model is represented as,

$$\left(1 - \beta L\right)\sigma_n^\delta = \omega + \left(\left(1 - \beta L\right) - \left(1 - \phi L\right)\left(1 - L\right)^d\right)\left(\left|\varepsilon_{n-1}\right| - \gamma\varepsilon_{n-1}\right)^\delta \quad (12)$$

where; L denotes the lag operator, d is the $0 \leq d \leq 1$ functional differencing parameter, β denotes the autoregressive parameters, ϕ represents the moving average parameters of the conditional variance equation. δ represents the optimal power transformation. γ represents the asymmetry parameter and $|\gamma| < 1$ ensures that positive and negative innovations of the same size can have asymmetric effects on the conditional variance (Conrad, Rittler and Rotfuss; 2010). Further, after imposing the restrictions δ=2 and d=0, the FIAPGARCH model reduces to AGARCH model; whereas, if the restriction δ=2 is applied, the model reduces to FIAGARCH, and if d=0 the model reduces to APGARCH model.

The ST-ARCH modeling methodology developed by Hegerud (1997), Gonzales-Rivera (1998), Lee and Degennaro (2000) allows smooth transition type nonlinearity in ARCH parameters and the ST-GARCH models of Anderson (1999) and Lundbergh and Terasvirta (2002) accept a modeling structure so that in addition to the ARCH terms, the intercept and the GARCH terms are extended to be modeled with smooth transition type nonlinearity in different regimes. Accordingly, following the ST-FIGARCH model structure, smooth transition fractionally integrated asymmetric power GARCH model denoted as ST-FIAPGARCH is obtained by allowing the smooth transition type nonlinearity between two FIAPGARCH models in two different regimes defined as

$$\sigma_t^\delta = \omega + \left(1 - F\left(\varepsilon_{t-s}, \gamma\right)\right)\alpha\sigma_{t-1}^\delta + \beta F\left(\varepsilon_{t-s}, \gamma\right)\sigma_{t-1}^\delta$$

$$+\left[\left[1-\alpha L\left(1-F\left(\varepsilon_{t-s},\gamma\right)\right)-\beta LF\left(\varepsilon_{t-s},\gamma\right)\right]-\left(1-\phi L\right)\left(1-\phi L\right)^{d}\right]\left(\left|\varepsilon_{n-1}\right|-\gamma\varepsilon_{n-1}\right)^{\delta} \quad (13)$$

if the transition function $F(.)$ is defined as a logistic function bounded between 0 and 1,

$$F(\varepsilon_{t-1})=\frac{1}{1+\mathbf{exp}\left(-\gamma\varepsilon_{t-1}\right)} \quad (14)$$

the obtained model is defined as the *logistic smooth transition fractionally integrated asymmetric power GARCH* (LST-FIAPGARCH) model.

v. STAR-GARCH Models

STAR-GARCH models, evaluated by Lundberg and Terasvirta (1999, 2000) and Franses Neele and van Dijk (1998) and further examined by Chan and McAleer (2001) are time series models with STAR type nonlinear processes in the conditional mean with heteroscedasticy given as GARCH errors. Consider the following STAR model (Terasvirta, 1994) with two regimes,

$$y_t=\phi_1+\sum_{i=1}^{r}\phi_{1i}y_{t-i}\left(1-F\left(s_t;\gamma,c\right)\right)+\phi_2+\sum_{i=1}^{r}\phi_{2i}y_{t-i}F\left(s_t;\gamma,c\right)+\varepsilon_t \quad (15)$$

where,

$$F\left(s_t;\gamma,c\right)=\frac{1}{1+e^{-\gamma(s_t-c)}} \quad (16)$$

defined with the logistic function. By allowing GARCH errors,

$$\sigma_t^2=\omega+\sum_{i=1}^{p}\alpha_i\varepsilon_{t-1}^2+\sum_{i=1}^{r}\beta_i\sigma_{t-1}^2 \quad (17)$$

the model is called Logistic Smooth Transition Autoregressive GARCH (LSTAR-GARCH) model. As the information matrix of the log-likelihood function of STAR-GARCH is block diagonal, the parameters in the conditional mean and conditional variance equations can be estimated separately, as in the case of ARMA-GARCH. The general GARCH properties are expected to hold (Chan and McAleer, 1999).

iv. STAR-ST-GARCH Model

The Smooth Transition Autoregressive (STAR) model further developed by Luukkonen *et al.* (1988), Granger and Terasvirta (1993) and Terasvirta (1994) aim nonlinear modeling of the conditional mean by introducing smooth transition between regimes of autoregressive processes based on logistic and exponential functions belonging to squashing functions. In STAR models, commonly applied transition functions are logistic and exponential functions and the relevant models are called LSTAR and ESTAR models. STAR–STGARCH model is a model that allows STAR type nonlinearity in both the conditional mean and the conditional variance and is developed based on the following STAR model. The error terms follow smooth transition in the GARCH process,

$$\begin{aligned}\sigma_t^2=&\left(w_1+\sum_{i=1}^{p}\alpha_{1i}\varepsilon_{t-1}^2+\sum_{i=1}^{r}\beta_{1i}\sigma_{t-i}^2\right)\left(1-H\left(\varepsilon_{t-1};\varsigma,n\right)\right)\\&+\left(w_2+\sum_{i=1}^{p}\alpha_{2i}\varepsilon_{t-1}^2+\sum_{i=1}^{r}\beta_{2i}\sigma_{t-i}^2\right)H\left(\varepsilon_{t-1};\varsigma,n\right)\end{aligned} \quad (18)$$

with the transition function,

$$H\left(\varepsilon_t;\varsigma,n\right)=\frac{1}{1+e^{-\varsigma(\varepsilon_{t-1}-n)}}\quad . \quad (19)$$

ς_t is the parameter defining the speed of transition and n is the threshold coefficient. Model will be extended to STAR-ST-FIGARCH model.

vii. STAR-ST-FIGARCH Model

The ARCH and GARCH models, developed by Engle (1982) and Bollerslev (1986) respectively, are short memory processes resulting from the fact that the response of a shock on the conditional variance decreases at an exponential rate. On the other hand, the conditional volatility of financial market returns may change slowly over time as a result of long memory characteristics of financial series. Consequently, the autocorrelation functions may decay at a hyperbolic rate[4].

Fractionally Integrated GARCH (FIGARCH(1, *d,* 1)) model is developed under these findings by Bollersev and Mikkelsen (1996) and Baillie, Bollerslev, and Mikkelsen (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.
Assume that a time series following a random walk process in its conditional mean and its conditional variance, $h_t = \mathrm{var}\left(\varepsilon_t \,\middle|\, \Omega_{t-1}\right)$, where the information set up to time *t*-1 is denoted as Ω_{t-1}, follows a FIGARCH(1,d,1) process,

$$(1-\beta L)\sigma_t^2 = \omega + \left((1-\beta L)-(1-\phi L)(1-L)^d\right)\left(|\varepsilon_{t-1}|-\gamma\varepsilon_{t-1}\right)^2 \qquad (20)$$

or alternatively,

$$\sigma_t^2 = \omega + \beta h_t + \left((1-\beta L)-(1-\phi L)(1-L)^d\right)\left(|\varepsilon_{t-1}|-\gamma\varepsilon_{t-1}\right)^2 \qquad (20)$$

where, z_t is assumed to be normally distributed $N(0,1)$ white noise process

$$z_t \sim N(0,1) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{z_t^2}{2}\right) \qquad (21)$$

FIGARCH(1, *d,* 1) model nests the GARCH model if $d = 0$ and the IGARCH model of Engle and Bollerslev (1986) for estimated fractional integration parameter of d = 1. Consequently, the fractional integration parameter d is 0<d<1 and as $d \to 0$ $(d \to 1)$ the model has short memory (long memory) characteristics. For alternative specifications of FIGARCH model, readers are referred to Karanasos, Psaradakis, and Sola (2004), Giraitis, Robinson, and Surgailis (2004) and Zaffaroni (2004).

The STAR-STFIGARCH model which generalizes the ST-GARCH type nonlinearity to account for long memory is represented as,

$$\begin{aligned}\sigma_t^\delta &= w_1 + \alpha_{1i}\sigma_{t-1}^\delta + \left(1-H(\varepsilon_i;\varsigma,n)\right)\beta_{1i}h_{t-1}^\delta + \beta_{2i}\sigma_{t-1}^\delta H(\varepsilon_i;\varsigma,n) + w_2 \\ &+\alpha_{2i}\varepsilon_{t-1}^2\left[\left[1-\beta L\left(1-\left(H(\varepsilon_i;\varsigma,n)\right)\right)-\beta L\left(H(\varepsilon_i;\varsigma,n)\right)\right]-(1-\phi L)(1-\phi L)^d\right]\left(|\varepsilon_{n-1}|-\gamma\varepsilon_{n-1}\right)^\delta\end{aligned} \qquad (22)$$

[4] Bollerslev and Mikkelsen (1996) develop the necessary conditions for FIGARCH model and note that for a well defined FIGARCH model, all the coefficients in the infinite ARCH representation must be non-negative (see: Bollerslev andMikkelsen (1996, p. 159). FIGARCH models are further discussed in Nelson and Cao (1992) and Conrad and Haag (2006), following these studies, nonnegativity constraints on parameters of FIGARCH processes are relaxed and shown that, for p=2, the second lag of conditional variance can become negative. Futher, Conrad and Haag (2006) allow conditions so that even if all parameters are negative (apart from d), the conditional variance can be nonnegative for FIGARCH models following the inequality constraints of Conrad and Haag (2006).

for $\gamma \neq 0$ the width of the volatility clusters and, α and β characterizes the dynamics of the conditional volatility. The range of the cluster of the volatility changes between $F(.)=0$ and $F(.)=1$. The constant term takes on values between $\varphi=\omega/(1-\alpha)$ and $\varphi=\omega/(1-\beta)$ based upon whether the conditional volatility is is the regime dictated by $F(.)=0$ and $F(.)=1$. Accordingly, since, in the ST-GARCH model, the constant term ranges between the extreme regimes, the level of conditional volatility will change in different regimes (Kılıç, 2010). If the transition function $F(.)$ is logistic function

$$F(\varepsilon_{t-1})=\frac{1}{1+\mathbf{exp}\left(-\theta\varepsilon_{t-1}\right)} \tag{23}$$

the model becomes logistic smooth transition FIGARCH (LST-FIGARCH) model.

v. STAR-ST-FIAPGARCH Model

Tse (1998) introduced the FIAPGARCH model which combines long memory property of Baillie, Bollerslev, and Mikkelsen (1996) FIGARCH model with Asymmetric Power GARCH (APGARCH) model of Ding, Engle, and Granger (1993) by extending the FIGARCH model to account for different asymmetric dynamics. Accordingly, the fractionally integrated APGARCH model is represented as

$$(1-\beta L)\sigma_n^{\delta}=w+\left((1-\beta L)-(1-\phi L)(1-L)^{d}\right)\left(|\varepsilon_{n-1}|-\gamma\varepsilon_{n-1}\right)^{\delta} \tag{24}$$

where; L denotes the lag operator, d is the $0 \leq d \leq 1$ functional differencing parameter, β denotes the autoregressive parameters, ϕ represents the moving average parameters of the conditional variance equation. δ represents the optimal power transformation. γ represents the asymmetry parameter and $|\gamma| < 1$ ensures that positive and negative innovations of the same size can have asymmetric effects on the conditional variance (Conrad, Rittler and Rotfuss; 2010). Further, after imposing the restrictions δ=2 and d=0, the FIAPGARCH model reduces to AGARCH model; whereas, if the restriction δ=2 is applied, the model reduces to FIAGARCH, and if d=0 the model reduces to APGARCH model.

The ST-ARCH modeling methodology developed by Hegerud (1997), Gonzales-Rivera (1998), Lee and Degennaro (2000) allows smooth transition type nonlinearity in ARCH parameters and the ST-GARCH models of Anderson (1999) and Lundbergh and Terasvirta (2002) accept a modeling structure so that in addition to the ARCH terms, the intercept and the GARCH terms are extended to be modeled with smooth transition type nonlinearity in different regimes. Accordingly, following the ST-FIGARCH model structure, smooth transition fractionally integrated asymmetric power GARCH model denoted as ST-FIAPGARCH is obtained by allowing the smooth transition type nonlinearity between two FIAPGARCH models in two different regimes defined as,

$$\begin{aligned}\sigma_t^{\delta} &= \omega+\left(1-\left(H\left(\varepsilon_i;\varsigma,n\right)\right)\right)\alpha\sigma_{t-1}^{\delta}+\beta\left(H\left(\varepsilon_i;\varsigma,n\right)\right)\sigma_{t-1}^{\delta}\\ &+\left[\left[1-\alpha L\left(1-\left(H\left(\varepsilon_i;\varsigma,n\right)\right)\right)-\beta L\left(H\left(\varepsilon_i;\varsigma,n\right)\right)\right]-(1-\phi L)(1-\phi L)^{d}\right]\left(|\varepsilon_{n-1}|-\gamma\varepsilon_{n-1}\right)^{\delta}\end{aligned} \tag{25}$$

if the transition function $F(.)$ is defined as a logistic function bounded between 0 and 1,

$$F(\varepsilon_{t-1})=\frac{1}{1+\mathbf{exp}\left(-\gamma\varepsilon_{t-1}\right)} \tag{26}$$

the obtained model is defined as the *logistic smooth transition fractionally integrated asymmetric power GARCH* (LST-FIAPGARCH) model.

III. Neural Network Augmentations of the Nonlinear GARCH Models

Artificial Neural Network (ANN) models have significant applications in modeling economic variables and time series. Kanas (2001), Kanas and Yannopoulos (2001), Shively (2003) applied ANN models to stock return forecasting, whereas, Donaldson and Kamstra (1996) proposed hybrid architecture of commonly applied GARCH family models, GARCH, GJR and EGARCH, with ANN architecture. Further analysis is conducted with Bildirici and Ersin (2009) to obtain a large class of GARCH family models with benefits from ANN modeling. Multi Layer Perceptron (MLP), an important class of neural networks consists of a set of sensory units defined with an input layer, one or more hidden layers and an output layer with estimation algorithms that include back-propagation and gradient descent type algorithms (See, Rumelhart et al., 1986; Bishop, 1994). Following Donaldson and Kamstra (1996) GJR-GARCH-NN, EGARCH-NN and GARCH-NN models, Bildirici ve Ersin (2009) proposed a family of NN-GARCH models including the NN-APGARCH model.

i. NN-GARCH Model

Start with the basic model, NN-GARCH *(p,q,m)* model is an augmented GARCH*(p,q)* process with single hidden layer ANN consisting sigmoid type neuron functions,

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 + \sum_{h=1}^{m} \xi_h \psi(z_t \lambda_h) \tag{27}$$

$$\psi(z_t \lambda_h) = \left[1 + \exp\left(\lambda_{h,d,w} + \sum_{d=1}^{1} \left[\sum_{w=1}^{m} \lambda_{h,d,w} z_{t-d}^{w} \right] \right) \right]^{-1} \tag{28}$$

$$z_{t-d} = [\varepsilon_{t-d} - E(\varepsilon)] \big/ \sqrt{E(\varepsilon^2)} \tag{29}$$

$$\left(1/2\right) \lambda_{h,d,w} \sim uniform\ [-1,+1] \tag{30}$$

$\psi(z_t \lambda_h)$ is the sigmoid type activation function of the form 1/(1+exp(-x)); $\xi = w$ is the weight vector; define $z_t \lambda_h = x_i$ as the input variables in the activation function with λ_h as given in equation (30).

ii. NN-APGARCH Model

Asymmetric power GARCH (APGARCH) structure of Ding et.al. (1993) has interesting features in volatility modeling. The NN-APGARCH model belongs to the NN-GARCH models discussed in Bildirici and Ersin (2009) and is an extention of Donaldson and Kamstra (1997) NN-GARCH models. The NN-APGARCH model is obtained by augmenting APGARCH model with artificial neural network architecture and modeling techniques,

$$\sigma_t^{\delta} = \omega + \sum_{k=1}^{r} \varphi_k \left(|\varepsilon_{t-k}| - \gamma_k \varepsilon_{t-k} \right)^{\delta} + \sum_{t=1}^{q} \beta_i \sigma_{t-i}^{\delta} + \sum_{h=1}^{s} \xi_h \psi(z_t \lambda_h) \tag{31}$$

$$\psi(z_t \lambda_h) = \left[1 + \exp\left(\lambda_{h,d,w} + \sum_{d=1}^{1} \left[\sum_{w=1}^{m} \lambda_{h,d,w} z_{t-d}^{w} \right] \right) \right]^{-1} \tag{32}$$

$$z_{t-d} = [\varepsilon_{t-d} - E(\varepsilon)] \big/ \sqrt{E(\varepsilon^2)} \tag{33}$$

$$\frac{1}{2} \lambda_{h,d,w} \sim uniform\ [-1,+1] \tag{34}$$

where, $\psi(z_t\lambda_h)$ is the logistic function. The NN-APGARCH nests several models. The model reduces to the standard NN-GARCH model for δ =2 and γ_k =0, the NN-NGARCH model for γ_k =0, and the NN-GJR-GARCH model for δ =2 and $0 \le \gamma_k \le 1$; the NN-TGARCH model for for δ =1 and $0 \le \gamma_k \le 1$. For estimation of NN-APGARCH models, readers are referred to Bildirici and Ersin (2009).

iii. NN-FIAPGARCH Model

In this study, NN-FIAPGARCH model is an augmented version of NN-APGARCH model proposed by Bildirici and Ersin (2009). NN-FIAPGARCH model is also an augmented version of fractionally integrated asymmetric power GARCH model with neural network architecture. The model is defined as,

$$(1-\beta L)\sigma_n^{\delta} = \omega + \left((1-\beta L)-(1-\phi L)(1-L)^d\right)\left(|\varepsilon_{n-1}|-\gamma_k\varepsilon_{n-1}\right)^{\delta} + \sum_{h=1}^{s}\xi_{h_t}\psi(z_t\lambda_h) \tag{35}$$

$$\psi(z_t\lambda_h) = \left[1+\exp\left(\lambda_{h,d,w}+\sum_{d=1}^{1}\left[\sum_{w=1}^{m}\lambda_{h,d,w}z_{t-d}^{w}\right]\right)\right]^{-1} \tag{36}$$

$$z_{t-d} = \left[\varepsilon_{t-d}-E(\varepsilon)\right]/\sqrt{E(\varepsilon^2)} \tag{37}$$

$$\frac{1}{2}\lambda_{h,d,w} \sim uniform\ [-1,+1] \tag{38}$$

where, $\psi(z_t\lambda_h)$ is the logistic function and *h* number of neurons. Logistic function belongs to the sigmoid type function family applied in neural network literature. The NN-FIAPGARCH nests several models. The model given in (35)-(38) reduces to the NN-FIGARCH model for restrictions on the power term δ =2 and γ_k =0; the model reduces to NN-FINGARCH model for γ_k =0; and to the NN-FIGJRGARCH model if δ =2 and γ_k is so that it varies between $0 \le \gamma_k \le 1$. Further, the model may be shown as NN-GARCH model if δ =1 in addition to the $0 \le \gamma_k \le 1$ restriction. For traditional representations of GARCH models readers may refer to Bollersev, 2007). Furthermore, the model could be represented with short memory characteristics under restrictions on fractional integration parameters. By imposing $d=0$ to the fractional differentiation parameter the model in Eq. (35) reduces to NN-APGARCH model, the short memory model variant. In this study, only FIGARCH and FIAPGARCH versions will be evaluated.

iv. LSTAR-GARCH-MLP Model

In this section of the study, the Multi Layer Perceptron Neural Network models that belong to the ANN family will be combined with LSTAR-GARCH models to benefit from the forecast capabilities of ANN models[5]. The LSTAR-GARCH-MLP model is a neural network model that consists of a set of sensory units with an input layer passed to two or more locally linear conditional mean processes with smooth transition logistic transition function, namely a LSTAR process with errors following GARCH type conditional volatility modeled as a NN-GARCH process.

[5] Donaldson and Kamstra (1996) proposed hybrid modeling to combine GARCH, GJR and EGARCH models with ANN architecture; whereas, NN-GARCH models are further extended to NN-GARCH, NN-EGARCH, NN-TGARCH, NN-GJR-GARCH, NN-SAGARCH, NN-PGARCH, NN-NPGARCH, and NN-APGARCH models by Bildirici and Ersin (2009).

The LSTAR-GARCH-MLP model is defined as a two regime LSTAR process in the conditional mean of which errors follows a single regime GARCH process augmented with neural networks with multi-layer perceptron structure,

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-1}^2 + \sum_{i=1}^{r} \beta_i \sigma_{t-i}^2 + \sum_{h=1}^{s} \xi_h \psi(z_t \lambda_h) \tag{39}$$

$$\psi(z_t \lambda_h) = \frac{1}{1+\exp\left(-\left(\lambda_{h,d,w} + \sum_{d=1}^{1}\left[\sum_{w=1}^{m} \lambda_{h,d,w} z_{t-d}^{w}\right]\right)\right)} \tag{40}$$

$$z_{t-d} = [\varepsilon_{t-d} - E(\varepsilon)] \big/ \sqrt{E(\varepsilon^2)} \tag{41}$$

$$\frac{1}{2}\lambda_{h,d,w} \sim uniform\ [-1,+1] \tag{42}$$

where, $F(s_t;\gamma,c)$ is the logistic transition function restricted to allow the transition to be a function of a single variable *s* and its respective distance to the threshold *c*. $\psi(z_t \lambda_h)$ is the logistic activation function, h is the number of neurons. The weight vector $\xi = w$; $\psi = g$ logistic activation function and input variables are defined as $z_t \lambda_h = x_i$ where λ_h is defined as in Eq.(42).

v. LSTAR-APGARCH-MLP Model

LSTAR-APGARCH-MLP model is a model with conditional mean following a LSTAR process with APGARCH type heteroscedasticity modeling of the conditional variance extended to NN-APGARCH model of Bildirici and Ersin (2009) following Donaldson and Kamstra (1997) NN-GARCH models. The LSTAR-APGARCH-MLP model is a two regime LSTAR model with the conditional variance following APGARCH process augmented with MLP neural network structure,

$$\sigma_t^{\delta} = \omega + \sum_{k=1}^{r} \alpha_k \left(|\varepsilon_{t-k}| - \gamma_k \varepsilon_{t-k}\right)^{\delta} + \sum_{t=1}^{q} \beta_i \sigma_{t-i}^{\delta} + \sum_{h=1}^{s} \xi_h \psi(z_t \lambda_h) \tag{43}$$

$$\psi(z_t \lambda_h) = \left[1+\exp\left(\lambda_{h,d,w} + \sum_{d=1}^{1}\left[\sum_{w=1}^{m} \lambda_{h,d,w} z_{t-d}^{w}\right]\right)\right]^{-1} \tag{44}$$

$$z_{t-d} = [\varepsilon_{t-d} - E(\varepsilon)] \big/ \sqrt{E(\varepsilon^2)} \tag{45}$$

$$\frac{1}{2}\lambda_{h,d,w} \sim uniform\ [-1,+1] \tag{46}$$

where, $\psi(z_t \lambda_h)$ is the logistic function. The model given in Eq.'s (41)-(46) nest several models. If the restrictions are applied as δ =2 and γ_k =0 in **Error! Reference source not found.**, the LSTAR-APGARCH-MLP model reduces to the LSTAR-GARCH-MLP model; for ξ_h=0, the model reduces to LSTAR-APGARCH model; and further, if γ=0, δ =2 and γ_k =0 in addition to ξ_h=0, the model reduces to a single regime GARCH model.

vi. LSTAR-FIGARCH-MLP Model

LSTAR-FIGARCH-MLP model is a LSTAR in the conditional mean process with errors following time varying conditional variance. The model is an augmented version of NN-

FIGARCH model which allows for fractional integrated time-varying conditional variance with neural networks.
The model is defined as,

$$(1-\beta L)\sigma_n^2 = \omega + \left((1-\beta L)-(1-\phi L)(1-L)^d\right)\left(|\varepsilon_{n-1}|-\gamma_k\varepsilon_{n-1}\right)^2 + \sum_{h=1}^{s}\xi_{h_t}\psi(z_t\lambda_h) \tag{47}$$

$$\psi(z_t\lambda_h) = \left[1+\exp\left(\lambda_{h,d,w} + \sum_{d=1}^{1}\left[\sum_{w=1}^{m}\lambda_{h,d,w} z_{t-d}^{w}\right]\right)\right]^{-1} \tag{48}$$

$$z_{t-d} = \left[\varepsilon_{t-d} - E(\varepsilon)\right] \Big/ \sqrt{E(\varepsilon^2)} \tag{49}$$

$$\frac{1}{2}\lambda_{h,d,w} \sim uniform\ [-1,+1] \tag{50}$$

Similar to the previous models, $\psi(z_t\lambda_h)$ is the logistic function with h number of neurons.

vii. LSTAR-FIAPGARCH-MLP Model

LSTAR-FIAPGARCH-MLP model is an augmented version LSTAR-FIGARCH-MLP with asymmetric power structure in the conditional variance. The model is stated as,
The model is defined as,

$$(1-\beta L)\sigma_n^\delta = \omega + \left((1-\beta L)-(1-\phi L)(1-L)^d\right)\left(|\varepsilon_{n-1}|-\gamma_k\varepsilon_{n-1}\right)^\delta + \sum_{h=1}^{s}\xi_{h_t}\psi(z_t\lambda_h) \tag{51}$$

$$\psi(z_t\lambda_h) = \left[1+\exp\left(\lambda_{h,d,w} + \sum_{d=1}^{1}\left[\sum_{w=1}^{m}\lambda_{h,d,w} z_{t-d}^{w}\right]\right)\right]^{-1} \tag{52}$$

$$z_{t-d} = \left[\varepsilon_{t-d} - E(\varepsilon)\right] \Big/ \sqrt{E(\varepsilon^2)} \tag{53}$$

$$\frac{1}{2}\lambda_{h,d,w} \sim uniform\ [-1,+1] \tag{54}$$

$\psi(z_t\lambda_h)$ is taken as logistic activation function. The model given in (51) nests the LSTAR-FIGARCH-MLP model for restrictions of δ =2 and γ_k =0. Further, reduces to LSTAR-GARCH-MLP model if δ=1 and by allowing $0 \le \gamma_k \le 1$. If the fractional differentiation parameter $d = 0$, the model reduces to LSTAR-APGARCH-MLP.

viii. LSTAR-LST-GARCH-MLP Model

By augmenting the LSTAR-LST-GARCH model defined with neural networks, following LSTAR-LST-GARCH-MLP model is obtained with LSTAR process in the conditional mean and the conditional variance modeled with LST-GARCH-MLP.
MLP augmented LST-GARCH process,

$$\begin{aligned}\sigma_t^2 = &\left(w_1 + \sum_{i=1}^{p}\alpha_{1i}\varepsilon_{t-1}^2 + \sum_{i=1}^{r}\beta_{1i}\sigma_{t-i}^2 + \sum_{h=1}^{s}\xi_{h_t,1}\psi_1(z_{t,1}\lambda_{h,1})\right)\left(1-H(\varepsilon_{t-1};\varsigma,n)\right)\\ &+\left(w_2 + \sum_{i=1}^{p}\alpha_{2i}\varepsilon_{t-1}^2 + \sum_{i=1}^{r}\beta_{2i}\sigma_{t-i}^2 + \sum_{h=1}^{s}\xi_{h_t,2}\psi_2(z_{t,2}\lambda_{h,2})\right)H(\varepsilon_{t-1};\varsigma,n)\end{aligned} \tag{55}$$

where,

$$H(\varepsilon_t;\varsigma,n) = \frac{1}{1+e^{-\varsigma(\varepsilon_{t-1}-n)}} \tag{56}$$

and,

$$\psi_i\left(z_{t,i}\lambda_{h,i}\right)=\frac{1}{1+\exp\left(-\left(\lambda_{h,d,w,i}+\sum_{d=1}^{1}\left[\sum_{w=1}^{m}\lambda_{h,d,w,i}z_{t-d,i}^{w,i}\right]\right)\right)} \tag{57}$$

i=1,2 and inputs defined as,

$$z_{t-d}=\left[\varepsilon_{t-d}-E(\varepsilon)\right]\Big/\sqrt{E\left(\varepsilon^2\right)} \tag{58}$$

$$\frac{1}{2}\lambda_{h,d,w}\sim uniform\ \left[-1,+1\right] \tag{59}$$

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

ix. LSTAR-LST-APGARCH-MLP Model

LSTAR-LST-APGARCH-MLP model is a LSTAR-LST-APGARCH model augmented with neural networks in each regime of the conditional variance process. The model is defined as,

$$\begin{aligned}\sigma_t^\delta=&\left(\omega_{0,1}+\sum_{k=1}^{r}\alpha_{k,1}\left(\left|\varepsilon_{t-k,1}\right|-\gamma_{k,1}\varepsilon_{t-k,1}\right)^\delta+\sum_{t=1}^{q}\beta_{i,1}\sigma_{t-i,1}^\delta+\sum_{h=1}^{s}\xi_{h,1}\psi_1\left(z_{t,1}\lambda_{h,1}\right)\right)\left(1\text{-}H\left(\varepsilon_{t-1};\varsigma,n\right)\right)\\&+\left(\omega_{0,2}+\sum_{k=1}^{r}\alpha_{k,2}\left(\left|\varepsilon_{t-k,2}\right|-\gamma_{k,2}\varepsilon_{t-k,2}\right)^\delta+\sum_{t=1}^{q}\beta_{i,2}\sigma_{t-i,2}^\delta+\sum_{h=1}^{s}\xi_{h,2}\psi_2\left(z_{t,2}\lambda_{h,2}\right)\right)\left(H\left(\varepsilon_{t-1};\varsigma,n\right)\right)\end{aligned} \tag{7}$$

$$\psi_i\left(z_{t,i}\lambda_{h,i}\right)=\left[1+\exp\left(-\left(\lambda_{h,d,w,i}+\sum_{d=1}^{1}\left[\sum_{w=1}^{m}\lambda_{h,d,w,i}z_{t-d,i}^{w,i}\right]\right)\right)\right]^{-1} \tag{60}$$

$$z_{t-d,i}=\left[\varepsilon_{t-d,i}-E\left(\varepsilon\right)\right]\Big/\sqrt{E\left(\varepsilon^2\right)} \tag{61}$$

$$\frac{1}{2}\lambda_{h,d,w,i}\sim uniform\ \left[-1,+1\right] \tag{62}$$

where i=1,2, the number of regimes. Accordingly, LSTAR-LST-APGARCH-MLP model is an hybrid model consisting of two regime LSTAR process in the conditional mean as in Eq.'s **Error! Reference source not found.-Error! Reference source not found.**, with residuals following a nonlinear neural network model for the conditional variance as in Eq.'s (7)-**Error! Reference source not found.** with multi-layer perceptrons in each regime of LST-APGARCH process. Therefore, the model is an augmented version of LSTAR-LST-APGARCH model to benefit from generalization capabilities of neural networks.

x. LSTAR-LST-FIGARCH-MLP Model

LSTAR-LST-FIGARCH-MLP model is a LSTAR-LST-GARCH-MLP model with fractional integration in the conditional variance process. LSTAR-LST-FIGARCH model is defined as,

$$(1-\beta L)\sigma_n^2=\Big(\omega_{0,1}+\big((1-\beta L)-(1-\phi_1 L)(1-L)^{d,1}\big)\big(|\varepsilon_{n-1}|-\gamma_{k,1}\varepsilon_{n-1}\big)^2\Big)\big(1-H(\varepsilon_{t-1};\varsigma,n)\big)$$
$$+\Big(\omega_{0,2}+\big((1-\beta L)-(1-\phi_2 L)(1-L)^{d,2}\big)\big(|\varepsilon_{n-1}|-\gamma_{k,2}\varepsilon_{n-1}\big)^2\Big)\big(H(\varepsilon_{t-1};\varsigma,n)\big) \quad (63)$$

$$\psi_i\left(z_{t,i}\lambda_{h,i}\right)=\left[1+\exp\left(-\left(\lambda_{h,d,w,i}+\sum_{d=1}^{1}\left[\sum_{w=1}^{m}\lambda_{h,d,w,i}z_{t-d,i}^{w,i}\right]\right)\right)\right]^{-1} \quad (64)$$

$$z_{t-d,i}=\left[\varepsilon_{t-d,i}-E(\varepsilon)\right]\Big/\sqrt{E\left(\varepsilon^2\right)} \quad (65)$$

$$\frac{1}{2}\lambda_{h,d,w,i}\sim uniform\ [-1,+1] \quad (66)$$

where i=1,2. The LSTAR-LST-FIGARCH-MLP model reduces to LSTAR-LST-GARCH-MLP if the fractional integration parameter d=0.

xi. LSTAR-LST-FIAPGARCH-MLP Model

LSTAR-LST-FIAPGARCH-MLP model is a model based on the LSTAR-LST-FIAPGARCH model augmented with MLP neural networks. The model is defined as,

$$(1-\beta L)\sigma_n^{\delta,i}=\Big(\omega_{0,1}+\big((1-\beta L)-(1-\phi_1 L)(1-L)^{d,1}\big)\big(|\varepsilon_{n-1}|-\gamma_{k,1}\varepsilon_{n-1}\big)^{\delta,1}\Big)\big(1-H(\varepsilon_{t-1};\varsigma,n)\big)$$
$$+\Big(\omega_{0,2}+\big((1-\beta L)-(1-\phi_2 L)(1-L)^{d,2}\big)\big(|\varepsilon_{n-1}|-\gamma_{k,2}\varepsilon_{n-1}\big)^{\delta,2}\Big)\big(H(\varepsilon_{t-1};\varsigma,n)\big) \quad (67)$$

$$\psi_i\left(z_{t,i}\lambda_{h,i}\right)=\left[1+\exp\left(-\left(\lambda_{h,d,w,i}+\sum_{d=1}^{1}\left[\sum_{w=1}^{m}\lambda_{h,d,w,i}z_{t-d,i}^{w,i}\right]\right)\right)\right]^{-1} \quad (68)$$

$$z_{t-d,i}=\left[\varepsilon_{t-d,i}-E(\varepsilon)\right]\Big/\sqrt{E\left(\varepsilon^2\right)} \quad (69)$$

$$\frac{1}{2}\lambda_{h,d,w,i}\sim uniform\ [-1,+1] \quad (70)$$

where i=1,2. LSTAR-LST-FIAPGARCH-MLP nests the models analyzed in the study. The model reduces to LSTAR-LST-GARCH-MLP if the fractional integration parameter d=0. Further, if the asymmetric power term is equal to 2, the model becomes LSTAR-LST-FIGARCH-MLP model.

IV. Econometric Results

4.1. Data

According to Lundbergh and Terasvirta (1998) and Hagerud (1996), Chan and McAleer (2002) and Chan and McAleer (2003), the empirical specification procedure for STAR-GARCH and STAR-STGARCH models consist of the six steps.

In order to test forecasting performance of the above-mentioned models, stock return in Turkey is calculated by using the daily closing prices of Istanbul Stock Index ISE 100 covering the 07.12.1986-13.12.2010 period corresponding to 5852 observations. To obtain return series, the stock returns data is calculated as follows: $y=ln(P_t/P_{t-1})$ where $ln(.)$ is the natural logarithms. In the process of model estimation, the sample is divided between training, test and out-of-sample samples with the percentages of 80%, 10%, 10%.

4.2. Econometric Results

At the first stage, among the GARCH family models, we selected basic GARCH model, and APGARCH models FIGARCH, taken as baseline models are estimated for evaluation purposes. Results are given in Table 1. Included models have different characteristics to be evaluated; namely, fractional integration, asymmetric power and fractionally integrated asymmetric power models, namely, GARCH, APGARCH, FIGARCH and FIAPGARCH models.

It is observed that, all volatility models perform better than the FIAPGARCH model in light of Log Likelihood criteria. If AIC and SIC criteria are evaluated, the lowest AIC (-4.5612) is calculated for the FIAPGARCH model; whereas, the lowest SIC is calculated as -4.5548 for the FIGARCH model. The sum of ARCH and GARCH parameters is calculated as 0.9857 for the GARCH model and similarly is less than 1 for the APGARCH, FIGARCH and FIAPGARCH model. For the fractionally integrated models, the differentiation parameters are estimated as 0.40 (FIGARCH) and 0.38 (FIAPGARCH).

Table 1. Baseline Models

	Baseline GARCH Models		***Baseline Fractionally Integrated GARCH Models***	
	1. GARCH	***2. APGARCH***	***3. FIGARCH***	***4. FIAPGARCH***
Cst(M)	0.0016**	0.0014**	0.0016**	0.00145**
	(5.36)	4.917	(5.488)	(5.16)
Cst(V)	0.1998**	0.7554	0.2842**	0.15264
	(3.21)	(1.20)	(3.15)	(1.08)
d-Figarch			0.4064**	0.3797**
			(8.21)	(6.66)
ARCH	0.1572**	0.1643**	0.2317**	0.2067*
	(6.36)	(7.07)	(2.24)	(1.76)
GARCH	0.8285**	0.8326**	0.4338**	0.3793**
	(29.99)	(31.71)	(3.73)	(2.72)
APARCH(Gamma1)		0.0518*		0.0592*
		(1.60)		(1.78)
APARCH(Delta)		1.6594**		2.0994**
		(8.18)		(17.12)
LogL	13361.16	13366.35	13405.95	13410.05
AIC:	-4.5455	-4.5466	-4.5605	-4.5612
SIC:	-4.5410	-4.5398	-4.5548	-4.5532
Q(5)	14.1874 [0.00]	17.6420 [0.00]	5.8026 [0.12]	5.29105 [0.15]
Q(10)	26.9742 [0.00]	30.9775 [0.00]	14.085 [0.07]	12.6800 [0.12]
SB:	0.38924 [0.69]	0.4958 [0.61]	0.6236 [0.53]	0.8391 [0.40]
ARCH (1-2):	4.4142 [0.012]	6.1645 [0.00]	0.8560 [0.42]	0.55257 [0.57]
ARCH (1-5):	2.7684 [0.02]	3.4571 [0.00]	1.1540 [0.32]	1.0520 [0.38]

LogL: Loglikelihood, Q(p): pth order autocorrelation test, SB: Sign bias test, ARCH(p): pth order ARCH-LM test.

Power terms obtained for returns calculated for stock indices in many developing economies are calculated comparatively higher than those obtained for the various indices in developed countries in various studies. The calculated power term is 1.65 in the APGARCH model and is estimated as 2.09 in the FIAPGARCH model showing high levels of asymmetry. It is noteworthy to evaluate several studies. Haas (2008) calculated three state RS-GARCH, RS-PGARCH and RS-APGARCH models for the daily returns in NYSE and estimated the power terms are calculated as 1.25, 1.09 and 1.08. For Turkey, Ural (2009) estimated a RS-APGARCH model for returns in ISE100 index in Turkey in addition to United Kingdom FTSE100, CAC40 in France and NIKKEI 225 indices in Japan and reported highest power estimates (1.84) compared to the power terms calculated as 1.26, 1.31 and 1.24 for FTSE100, NIKKEI 225 and CAC40. Telatar and Binay (2001) estimated APARCH models for Turkey and 10 national stock indices and noted that power terms reported for developing countries tend to be high and varying though those reported for the developed countries are estimated with low and close values. Ané and Ureche-Rangau (2006), estimated single regime GARCH and APGARCH models in addition to RS-GARCH and RS-APGARCH models following Gray (1996) model. Power terms in single regime APGARCH models were calculated for daily returns as 1.57 in Nikkei 225 Index, as 1.81 in Hang Seng Index, as 1.69 in Kuala Lumpur Composite Index and as 2.41 in Singapore SES-ALL Index. We will further evaluate LSTAR-GARCH and LSTAR-LST-GARCH models. LSTAR-GARCH models are tested by assuming that the error terms follow student-t distribution with the help of BFGS algorithm. Statistical inference regarding the empirical validity of two-regime switching process was carried out by using nonstandard LR tests (Davies, 1987). The non-standard LR test is statistically significant and this suggests that linearity is strongly rejected. Further, Lukkonnen et al. (1988) LM type nonlinearity tests are evaluated and concluded that the nonlinearity is accepted and linearity is rejected for the transition variable of one lagged daily returns.

STAR-GARCH models allow STAR type nonlinearity in the conditional mean with GARCH type heteroscedasticity in the conditional variance, where, the GARCH process is a single regime process. Chan and McAleer (2003) discuss that the results obtained with modeling time series inherently heteroscedastic in STAR-GARCH models and draws attention on the following three possibilities: (a) the variance is not constant, so that STAR-GARCH should be used; (ii) the use of alternative optimization algorithms is required for gains in modeling, and (iii) the use of alternative initial values in optimization. In this study, different types of nonlinearity either in mean or variance are evaluated with augmenting the STAR-ST-GARCH models with neural networks and support vector machines to encounter the following problems noted by Chan and McAleer (2003), regarding the likelihood functions: a) the log-likelihood functions of Exponential STAR-GARCH (ESTAR-GARCH) models tend to be flat around the global optimum near the true values of the transition rates. There are difficulties in estimating the transition rates by maximizing the log-likelihood functions using conventional gradient-based optimization algorithms. b) The planes of the log-likelihood functions of the Logistic STAR-GARCH models are prone to be lumpy in addition to being flat around the local optimums. These situations explain the sensitivity of QMLE to initial values. As noted by Lundbergh and Terasvirta (1999) and van Dijk, Terasvirta and Franses (2002), the convergence of QMLE is sensitive to the initial values. There are two result of these findings: (i) the shapes of the log-likelihood functions are determined mostly by the choice of transition functions and (ii) it may be possible to transform the shapes of the log-likelihood functions by transforming the parameters in the models. According to Chan and Theoharakis (2011), Although there are the popularity in applying regimes switching models, the statistical and structural properties for STAR-GARCH models are limited and the results are generally restricted to the two-regimes state. As their opinion, the lack of general structural and statistical properties makes valid inferences difficult to conduct for multi-regimes switching models. The transition rates in the STAR models are particularly difficult to estimate with the Quasi-Maximum Likelihood Estimator (QMLE).

Furthermore, GARCH models are extended to model nonlinearity in both the conditional mean and the conditional variance. ST-GARCH model has a linear process such as the random walk for the majority of studies, whereas, the conditional mean follows a two regime GARCH process in which the transition between the regimes are governed by a continuous, twice differentiable function such as the exponential or the logistic function to smooth transition. By hybridization of two groups of nonlinear models; we obtain STAR-ST-GARCH model that allows for STAR type nonlinearity in both the conditional mean and variance. By allowing the transitions to be governed by logistic function, LSTAR-LST-GARCH model is obtained. By comparing three groups of models, single regime GARCH, STAR-GARCH and STAR-ST-GARCH, we obtained several results. To encounter the problem of forecast accuracy, the study extends GARCH models to nonlinearity both in mean and variance with neural networks. Firstly, following Donaldson and Kamstra (1997) and Bildirici and Ersin (2009) we estimated NN-GARCH models. Similar to the methodology followed to obtain NN-GARCH models. Secondly, we suggested the hybrid modeling methodology as proposed in the paper to augment STAR-ST-GARCH models with neural networks and generalize to LSTAR-LST-GARCH-MLP, Logistic Smooth Transition Autoregressive in conditional mean, logistic smooth transition in conditional variance augmented with multi layer perceptron neural network model.

For comparative purposes, LSTAR-GARCH and LSTAR-LST-GARCH models are reported in Table 2. In the first part of Table 2, the stability condition is achieved for all LSTAR-GARCH family models, LSTAR-GARCH, LSTAR-APGARCH, LSTAR-FIGARCH and LSTAR-FIAPGARCH. The fractional integration parameters are estimated as 0.44 and 0.43 for the LSTAR-FIGARCH and LSTAR-FIAPGARCH models showing that the degree of fractional integration is calculated close but higher than those reported for the single regime in conditional variance models; FIGARCH and FIAPGARCH. The asymmetric power parameter is estimated at high levels as 1.73 and 1.95 for the LSTAR-APGARCH and LSTAR-FIAPGARCH models. The loglikelihood values are also high as was for the single regime models. AIC and SIC criteria report similar conclusions for the in-sample results [6]. On the other hand, the results show significant improvements after LSTAR-LST-GARCH models which allow nonlinearity in the conditional variance as well as in the conditional mean. Loglikelihood values are significantly reduced and AIC and SIC information criteria are significantly lower. Further differences include, after allowing the GARCH processes to follow LST type nonlinearity, the dynamics are strikingly different in light of the estimated parameters. In the LSTAR-LST-FIGARCH model, d parameters are estimated as 0.69 and 0.16 for regime 1 and 2. Fort he LSTAR-LST-FIAPGARCH model, d parameters are estimated as 0.17 and 0.24 with comparatively low values.

[6] On the other hand, though models have similar performances in the in-sample modeling, the relevant gains are achieved for the out-of-sample forecasting. The results will be reported in the following section.

Table 2. Models with STAR Type Nonlinearity in the Conditional Mean and Conditional Variance

	LSTAR-GARCH	***LSTAR-APGARCH***	***LSTAR-FIGARCH***	***LSTAR-FIAPGARCH***	***LSTAR-LST-GARCH***		***LSTAR-LST-APGARCH***		***LSTAR-LST-FIGARCH***		***LSTAR-LST-FIAPGARCH***	
	Single regime in GARCH process				***Regime 1***	***Regime 2***	***Regime 1***	***Regime 2***	***Regime 1***	***Regime 2***	***Regime 1***	***Regime 2***
Cst(M)	0.0002	0.00003	0.0002	0.00003	0.0043**	0.0115**	-0.0294**		0.0139**	-0.0141*	-0.0199**	0.0014
	(0.8)	(0.11)	(0.89)	(0.12)	(10.46)	(16.48)	(-45.20)		(11.69)	(-1.88)	(-18.26)	(1.49)
Cst(V)	0.1877**	0.5568	14.4879	17.7108	0.0586*	0.4759*	25.6249		0.0103**	2.7721**	0.010	0.0449*
	(3.15)	(1.06)	(-1.24)	(0.71)	(1.701)	(1.624)	(0.6704)		(2.36)	(3.11)	(0.89)	(1.65)
d-Figarch			0.4397**	0.4295**					0.6906**	0.1584*	0.1701**	0.2364**
			(-5.79)	(7.28)					(11.78)	(1.74)	(1.90)	(3.70)
ARCH	0.1489**	0.1576**	0.1811*	0.1715**	0.0822**	0.1685**	0.0783*	0.2475**	0.2925**	0.8486**	0.3056**	0.4420**
	(6.11)	(6.59)	(-1.85)	(1.73)	(4.504)	(2.149)	(1.71)	(2.62)	(4.48)	(6.50)	(2.48)	(7.068)
GARCH	0.8367**	0.8361**	0.4322	0.4136**	0.9137**	0.7239**	0.6920**	0.8816**	0.7966**	0.8435**	0.6786**	0.6212**
	(30.71)	(31.27)	(-3.24)	(3.15)	(46.85)	(5.673)	(2.38)	(28.67)	(14.47)	(10.39)	(2.746)	(7.935)
APARCH (Gamma1)		0.0664*		0.0741**			0.9995**	0.2677**			0.7518	0.5761**
		(1.84)		(2.02)			(93.66)	(2.73)			(1.08)	(1.99)
APARCH (Delta)		1.7318**		1.9518**			1.1437**	1.2286**			1.2027**	1.3730**
		(7.67)		(11.87)			(2.784)	(6.38)			(4.36)	(10.69)
ARCH in mean									-0.5072**	1.4267**		2.9349**
									(-15.63)	(3.41)		(8.95)
LogL	13385.219	13390.251	13430.24	13434.8	6761.74		3945.71		6898.49		6792.50	
AIC:	-4.5568	-4.5578	-4.5718	-4.5727	-4.7725		-4.9625		-4.8654		-4.5537	
SIC:	-4.5523	-4.5511	-4.5662	-4.5647	-4.7628		-4.9438		-4.8508		-4.5432	
Q(5)	8.5582 [0.03]	9.76862 [0.02]	4.9787 [0.17]	4.8868 [0.18]	3.747 [0.29]		3.3781 [0.34]		1.9448 [0.58]		1.9632 [0.58]	
Q(10)	20.4037 [0.00]	21.6716 [0.01]	14.0515 [0.08]	13.4890 [0.09]	6.1430 [0.63]		5.9495 [0.65]		3.8414 [0.87]		4.7870 [0.78]	
SB:	0.6801 [0.49]	0.7548 [0.45]	0.467 [0.62]	0.9517 [0.34]	2.339 [0.02]		0.2332 [0.82]		1.5086 [0.13]		2.1264 [0.03]	
ARCH (1-2):	2.4227 [0.08]	2.9871 [0.05]	0.9872 [0.42]	0.4458 [0.64]	0.99732 [0.37]		1.0956 [0.33]		0.0214 [0.97]		0.0752 [0.92]	
ARCH (1-5):	1.6886 [0.13]	1.9327 [0.08]	1.4537 [0.15]	0.9722 [0.43]	0.73513 [0.60]		0.6585 [0.65]		0.3885 [0.86]		0.3895 [0.85]	

Note: P-values are given in brackets. t-statistics are given in parentheses. * (**) denotes %10 (%5) significance level.

Table 3. Models with NN Type Nonlinearity in the Conditional Mean

	MLP-GARCH	***MLP-APGARCH***	***MLP – FIGARCH***	***MLP – FIAPGARCH***
Cst(M)	-0.0002	-0.00001	-0.0002	-0.00003
	(-0.68)	(-0.06)	(-0.78)	(-0.12)
Cst(V)	0.2003**	0.5808	0.0133	0.0153
	(3.25)	(1.19)	(1.29)	(0.70)
d-Figarch			0.4332**	0.4259**
			(6.20)	(7.44)
ARCH	0.1560**	0.1634**	0.1930**	0.1848*
	(6.37)	(6.91)	(1.99)	(1.89)
GARCH	0.8290**	0.8305**	0.4290**	0.4146**
	(30.31)	(31.08)	(3.39)	(3.26)
APARCH (Gamma1)		-0.060*		-0.0675**
		(-1.82)		(-1.96)
APARCH (Delta)		1.7318**		1.9684**
		(8.52)		(12.07)
LogL	13372.56	13377.5	13417.23	13421.37
AIC:	-4.5525	-4.5535	-4.5674	-4.5681
SIC:	-4.5480	-4.5467	-4.5617	-4.5602
Q(5)	12.4986 [0.01]	14.3909 [0.00]	6.6036 [0.08]	6.2299 [0.10]
Q(10)	25.1365 [0.00]	26.9583 [0.00]	15.5264 [0.05]	14.8638 [0.06]
SB:	0.8317 [0.41]	0.8713 [0.3835]	1.0132 [0.31]	1.2050 [0.22]
ARCH (1-2):	3.9319 [0.01]	4.8583 [0.01]	1.1320 [0.32]	0.9353 [0.39]
ARCH (1-5):	2.4587 [0.03]	2.8385 [0.01]	1.3099 [0.26]	1.2395 [0.28]

Note: P-values are given in brackets. t-statistics are given in parentheses. * (**) denotes %10 (%5) significance level.

Though differentiation parameters suggest stationarity except for the 1st regime of LSTAR-LST-FIGARCH model, all LST-GARCH type models suggest that stability condition of addition of ARCH and GARCH parameters is not achieved. Overall, it is noteworthy that following the LST-GARCH specification LogL, AIC and SIC calculations show significant improvement in light of in sample estimation. On the other hand, out-of-sample performances will be given in the next section.

MLP-GARCH models are estimated to evaluate possible augmentation of GARCH models to overcome the known out of sample forecasting capability. We estimated 4 models with multi-layer perceptron (MLP) architecture in the conditional mean. The methodology aims to cope with the random walk in the mean and is similar to the approach followed by Chan and McAleer (2003) LSTAR-GARCH approach. Estimated models are given in Table 3, whereas, model selection and learning results are given in Table 4. The estimated models show improvement over simple GARCH models in the in-sample and out-of sample performances as reported for the LSTAR-GARCH family models.

Table 4. Model Architecture and Learning Results

	Models and Their Architectures:			
Learning Results:	***1. MLP-GARCH (4:6:1:1)***	***2. MLP-APGARCH (4:4:1:1)***	***3. MLP-FIGARCH (4:8:1:1)***	***4. MLP-FIAPGARCH (4:4:1:1)***
Training rho**	0.103297	0.112415	0.117989	0.109957
Test rho	0.110626	0.100718	0.110351	0.097282
Training MSE	0.002248	0.002243	0.00224	0.002244
TEST MSE	0.002246	0.002249	0.002245	0.002251
Training algorithm (Convergenge)	BFGS 4	BFGS 25	BFGS 40	BFGS 6

*All models possess logistic and identity activation functions in the hidden and output layers, respectively. Models are read as follows: a MLP-GARCH (4:6:1:1) model has 3 variables in the input layer (independent variables), passed to the hidden layer with 6 neurons connected to the output layer with single output with errors specified with single regime GARCH process.
** Rho and MSE represent training and test sample correlation coefficient and mean squared error, respectively. BFGS is the Broyden–Fletcher–Goldfarb–Shanno nonlinear optimization algorithm. The algorithm and the epoch at which the algorithm is converged are reported in parentheses.

The above mentioned MLP-GARCH model assume neural network type nonlinearity in the conditional mean only; therefore, different than the approach based on neural networks methods. With following the methodology based on neural networks, NN-GARCH models are developed by Donaldson and Kamstra (1997) and further extended to a family of NN-GARCH models in Bildirici and Ersin (2009).

In the study, by following Bildirici and Ersin (2009) NN-GARCH modeling approach, the study extends to LSTAR- GARCH-NN and LSTAR-LST-GARCH-NN models to improve forecasting accuracy. Accordingly, estimation is conducted with conjugant-gradient based back-propagation algorithm; neural networks are estimated for a large amount of models with optimization conducted simultaneously in the training and test samples; optimization is early stopped at the epoch at which MSE in the test sample starts to increase though still continues to decrease in the training sample. During the optimization, weight decay in the output layer and hidden layer is utilized to eliminate the insignificant coefficients. In total, each model is estimated with 20 different NN architectures for each NN model amounting to 80 models with different numbers of hidden neurons constrained to range between 3 to 10 with logistic activation functions in the hidden layer. Best models with the lowest error criteria such as MSE or RMSE are selected. The selected models are further utilized for out-of-sample forecasting[7]. Estimated LSTAR-GARCH-MLP models are given in Table 6 and their relevant one-step-ahead forecast results are given in Table 7.

Table 6. Model Architecture and Learning Results

	Models and Their Architectures:			
Learning Results:	**1. LSTAR-GARCH-MLP(3:2:1:7:1)**	**2. LSTAR-APGARCH-MLP(3:2:1:5:1)**	**3. LSTAR-FIGARCH-MLP(3:2:1:6:1)**	**4. LSTAR-FIAPGARCH-MLP(3:2:1:5:1)**
Training rho**	0.951245000	0.950394000	0.909832000	0.907327000
Test rho	0.917399000	0.921656000	0.859220000	0.855687000
Training MSE	0.000280000	0.000272000	0.000370000	0.000303000
TEST MSE	0.000605000	0.000564000	0.000797000	0.000666000
Training algorithm (Convergenge)	BFGS(21)	BFGS(13)	BFGS(63)	BFGS(7)

*All models possess logistic and identity activation functions in the hidden and output layers, respectively. Models are read as follows: a LSTAR-GARCH-MLP(3:2:1:7:1) model is a model with 3 input variables (independent variables), 2 regime LSTAR model with single regime GARCH conditional variance process passing through 7 neurons to the output layer with 1 output (dependent) variable.
** Rho and MSE represent training and test sample correlation coefficient and mean squared error, respectively. BFGS is the Broyden–Fletcher–Goldfarb–Shanno nonlinear optimization algorithm. The epoch the algorithm converged is reported in parentheses.

Table 7. One-Step-Ahead Forecast Results

	1. LSTAR-GARCH-MLP(3:2:1:7:1)	**2. LSTAR-APGARCH-MLP(3:2:1:5:1)**	**3. LSTAR-FIGARCH-MLP(3:2:1:6:1)**	**4. LSTAR-FIAPGARCH-MLP(3:2:1:5:1)**
Mean square error	0.000000072	0.000000068	0.000000116	0.000000114
Mean absolute error	0.000119285	0.000124875	0.000151344	0.000150456
Mean relative squared error	0.094764102	0.085696768	0.138781470	0.143717708
Mean relative absolute error	0.154683209	0.157094878	0.192220780	0.194691964
Correlation coefficient	0.942973734	0.943076423	0.896710097	0.895582401
RMSE	0.000268701	0.000260770	0.000340147	0.000337639

Among the LSTAR-GARCH-MLP models, the lowest RMSE value for the test sample is 0.000026 and is obtained for the LSTAR-APGARCH-MLP model. LSTAR-GARCH-MLP model is the 2nd with RMSE=0.000268. LSTAR-FIGARCH-MLP and LSTAR-FIAPGARCH-MLP models took the 3th and 4th places with RMSE values calculated as 0.000337 and 0.000340. Compared to the GARCH, LSTAR-GARCH, LSTAR-LST-GARCH, NN-GARCH model given below, LSTAR-GARCH-MLP models show significant improvement in terms of in-sample analysis.

Further, the models are extended to LSTAR-LST-GARCH-MLP models. The model architectures and learning results are reported in Table 8. Compared to the results obtained for LSTAR-GARCH-MLP models, training and test MSE errors are calculated comparatively lower for the LSTAR-LST-GARCH-MLP models. Training MSE errors are 0.000145, 0.00015, 0.00026 and 0.00033 for LSTAR-LST-GARCH, LSTAR-LST-APGARCH, LSTAR-LST-FIGARCH and LSTAR-LST-FIAPGARCH models, respectively, which shows

[7] 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 2nd order derivative based optimization algorithms such as Conjugate Gradient Descent, Quasi-Newton, Quick Propagation, Delta-Bar-Delta and Levenberg-Marquandt, 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. Firstly, all models were trained over a given training sample vis-à-vis checking for generalization accuracy in 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 parsimonic models, 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 are referred to Bildirici and Ersin (2009).

that MSE's are almost half of those reported for LSTAR-GARCH family models. A similar result holds for both one-step ahead and out-of-sample forecasts. For a typical, though MLP-GARCH (Training MSE=0.0022) provides improved in-sample fit compared to the simple GARCH model, LSTAR-GARCH-MLP model provides significant improvement (MSE=0.00028) over MLP-GARCH model; thus LSTAR-LST-GARCH model has the modest in-sample fit (MSE=0.000145).

Table 8. Model Architecture and Learning Results

	Models and Architectures:			
Learning Results:	**1. LSTAR- LST-GARCH- MLP(3:2:2:6:1)**	**2. LSTAR- LST-APGARCH- MLP(3:2:2:5:1)**	**3. LSTAR- LST-FIGARCH- MLP(3:2:2:7:1)**	**4. LSTAR- LST-FIAPGARCH- MLP(3:2:2:8:1)**
Training rho**	0.891528500	0.913539000	0.788931500	0.905314000
Test rho	0.953208500	0.969500000	0.894839000	0.924713000
Training MSE	0.000145500	0.000150000	0.000262500	0.000327500
TEST MSE	0.000099500	0.000059500	0.000090000	0.000237000
Training algorithm (Convergenge)	BFGS 17	BFGS 25	BFGS 19	BFGS 11

*All models possess logistic and identity activation functions in the hidden and output layers, respectively. Models are read as follows: a LSTAR-GARCH-MLP(3:2:2:7:1) model is a model with 3 input variables (independent variables), 2 regime LSTAR model with two regime LST-GARCH conditional variance process passing through 7 neurons to the output layer with 1 output (dependent) variable.

** Rho and MSE represent training and test sample correlation coefficient and mean squared error, respectively. BFGS is the Broyden–Fletcher–Goldfarb–Shanno nonlinear optimization algorithm. The epoch the algorithm converged is reported in parentheses.

One-step-ahead forecast results are given in Table 9. According to the one-step ahead forecast RMSE's, LSTAR-LST-APGARCH-MLP (RMSE=0.000179) model has the lowest RMSE followed by LSTAR-LST-GARCH-MLP (RMSE=0.00019), LSTAR-LST-FIAPGARCH-MLP (RMSE=0.000209) and LSTAR-LST-FIGARCH-MLP (RMSE=0.000306) models. If compared to the in-sample statistics obtained for the previous models, LSTAR-LST-GARCH-MLP models provide the highest in-sample forecast accuracy following MLP specifications. On the other hand, evaluation of the models in terms of their relevant out-of-sample performances will be provided for comparative purposes.

Table 9. One-Step-Ahead Forecast Results

	1. LSTAR- LST-GARCH- MLP(3:2:2:6:1)	**2. LSTAR- LST-APGARCH- MLP(3:2:2:5:1)**	**3. LSTAR- LST-FIGARCH- MLP(3:2:2:7:1)**	**4. LSTAR- LST-FIAPGARCH- MLP(3:2:2:8:1)**
MSE*	0.000000039	0.000000034	0.000000095	0.000000045
MAE	0.000074217	0.000072350	0.000124469	0.000106440
MRSE	0.158236075	0.067518093	0.282407490	0.110035573
MRAE	0.134280974	0.118900139	0.204427622	0.169819328
Rho	0.900019470	0.922051824	0.798404384	0.907570660
RMSE	0.000194822	0.000179186	0.000306982	0.000209570

*MSE: Mean Squared Error, MAE: Mean Absolute Error, MRSE: Mean Relative Absolute Error, MRAE: Mean Relative Absolute Error, Rho: Correlation, RMSE: Root Mean Square Error.

Table 10. Out of Sample Forecast Statistics, 80 Days Ahead

	GARCH	APGARCH	FIGARCH	FIAPGARCH
RMSE	0.000819000	0.00083000	0.00079600	0.00078900
MAE	0.000321000	0.00032000	0.00032800	0.00034200
	LSTAR-GARCH	**LSTAR-APGARCH**	**LSTAR-FIGARCH**	**LSTAR-FIAPGARCH**
RMSE	0.000736000	0.00074700	0.000699000	0.00070100
MAE	0.000326000	0.00032500	0.000338000	0.00033700
	LSTAR-LSTGARCH	**LSTAR-LSTAPGARCH**	**LSTAR-LSTFIGARCH**	**LSTAR-LSTFIAPGARCH**
RMSE	0.000935000	0.07046200	0.00090800	0.00094700
MAE	0.000374000	0.03681100	0.00035500	0.00029700
	MLP-GARCH	**MLP-APGARCH**	**MLP-FIGARCH**	**MLP-FIAPGARCH**
RMSE	0.000777800	0.000787800	0.000740200	0.000741200
MAE	0.000323100	0.000323100	0.000343600	0.000342700
	LSTAR-GARCH-MLP	**LSTAR-APGARCH-MLP**	**LSTAR-FIGARCH-MLP**	**LSTAR-FIAPGARCH-MLP**
RMSE	0.000054772	0.000060000	0.000068557	0.000080623
MAE	0.000045418	0.000045972	0.000056786	0.000069807
	LSTAR-LST-GARCH-MLP	**LSTAR-LST-APGARCH-MLP**	**LSTAR-LST-FIGARCH-MLP**	**LSTAR-LST-FIAPGARCH-MLP**
RMSE	0.000036142	0.000025243	0.000057878	0.000036227
MAE	0.000031714	0.000020036	0.000051148	0.000028500

*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.

Models are evaluated for their generalization capabilities in the out-of-sample with RMSE and MAE criteria. Results are given in Table 10 in which GARCH, LSTAR-GARCH, LSTAR-LST-GARCH, MLP-GARCH, LSTAR-GARCH-MLP and LSTAR-LST-GARCH-MLP models generalized to APGARCH, FIGARCH and FIAPGARCH architectures totaling to 28 different conditional volatility models are compared to investigate their forecast accuracy for 80 work days (4 month period) ahead.

First of all, the models having GARCH structure in common corresponding to the 1st column will be evaluated. If an overlook is to be provided, though there is improvement as we move from single regime GARCH model to LSTAR-GARCH and LSTAR-LST-GARCH model, we noted that the performance of MLP-GARCH model is almost identical to the LSTAR-GARCH model. The RMSE reported for GARCH model is 0.00082 which decreases to 0.00074 with the LSTAR-GARCH model. RMSE's for MLP-GARCH models are calculated as 0.00077 showing improvement in 80 days ahead foreacasts though the improvement is low. Note that, the above mentioned models allow nonlinear modeling of conditional mean except for the LSTAR-LST-GARCH that allows STAR type nonlinearity both in the mean and in the varianve. On the other hand, if models are augmented with MLP architecture for generalization purposes the improvement is significant. Accordingly, the RMSE values for the The LSTAR-GARCH-MLP and LSTAR-LST-GARCH-MLP models are 0.0000547 and 0.000036 showing almost 10 times improvement in out of sample forecast accuracy. If models are evaluated by rows, in terms of RMSE criteria, LSTAR-LST-APGARCH-MLP (RMSE=0.000025) has the highest forecast capability followed by the LSTAR-LST-GARCH-MLP and LSTAR-LST-FIAPGARCH-MLP models having almost same RMSE values (0.0000361 and 0.0000362). In the last row, LSTAR-LST-GARCH-MLP models provide the highest forecast accuracy followed by the LSTAR-GARCH-MLP models.

In terms of the MAE criteria, LSTAR-LST-APGARCH-MLP model also has the best generalization capacity (MAE=0.000020) followed by the LSTAR-LST-FIAPGARCH-MLP model (0.0000285). Results support that though nonlinear volatility models with STAR type nonlinearity namely, LSTAR-GARCH and LSTAR-LST-GARCH family provide significant gains in in-sample accuracy, though there are significant improvement compared to single regime GARCH models, MLP augmentations in conditional volatility of these models provide significant forecast accuracy improvement. Thus, both model groups, LSTAR-GARCH-MLP and LSTAR-LST-GARCH-MLP show significant gains in terms of generalization in the out-of-sample. Results suggest that nonlinear augmentations of GARCH models for forecasting may provide certain gains, significant improvement in forecasting is achieved following the neural network architecture and modeling techniques in nonlinear modeling of conditional volatility.

Conclusion

The study aimed to investigate linear GARCH, fractionally integrated FI-GARCH and Asymmetric Power APGARCH models and their nonlinear counterparts based on a family of Neural Network models. GARCH models are extended to neural network based structures. In the study, nonlinear augmentations based on STAR type nonlinearity are evaluated and further augmented to MLP modeling methodology and architecture. The models analyzed are in spirit of NN-GARCH architecture of Donaldson and Kamstra (1997) which are enhanced to various NN-GARCH family models in Bildirici and Ersin (2009).

The models in the literature aim augmenting the conditional mean or the conditional variance or both with nonlinear techniques. In the study, we evaluated various forms of models and suggest MLP based augmentations. Baseline models analyzed include GARCH, FIGARCH, APGARCH and FIAPGARCH; their relevant LSTAR in the mean augmentations are LSTAR-GARCH, LSTAR-FIGARCH, LSTAR-APGARCH and LSTAR-FIAPGARCH; smooth transition type nonlinearity in the mean as well as the variance are LSTAR-LST-GARCH, LSTAR-LST-FIGARCH, LSTAR-LST-APGARCH and LSTAR-LST-FIAPGARCH models. Following the literature, we first evaluated modeling the conditional mean with state of the art nonlinear models, MLP model with errors following GARCH, APGARCH, FIGARCH and FIAPGARCH type processes. The obtained models are MLP-GARCH, MLP-FIGARCH, MLP-APGARCH and MLP-FIAPGARCH models. Results show that though there is improvement in terms of insample and out-of sample accuracy as we move from GARCH towards LSTAR-GARCH, LSTAR-LST-GARCH and MLP-GARCH models, the improvement in a forecast horizon of 80 days ahead is not satisfactory. As a result, models are augmented with neural networks in conditional variance processes. The obtained LSTAR-GARCH-MLP and LSTAR-LST-GARCH-MLP model family showed significant improvement in terms of forecast accuracy in the out-of-sample, whereas, the highest gains are obtained from LSTAR-LST-APGARCH-MLP model.

In conclusion at first step, models with fractional integration and asymmetric power GARCH provided gains compared to simple GARCH models. Specifications such as LSTAR-GARCH and LSTAR-LST-GARCH further augmented to their fractionally integrated and asymmetric power GARCH variants such as LSTAR-LST-FIAPGARCH in this study also provide gains in modeling. For MLP-GARCH model we obtained low improvement compared to the previous models. However, neural network augmented LSTAR-LST-GARCH-MLP models provided significant gains in forecast accuracy. Therefore, the NN extended versions of ST and FI type volatility models are shown to provide improved forecast results. Results suggest that volatility clustering, asymmetry and nonlinearity characteristics are modeled more efficiently and provide better forecast accuracy with neural networks based LSTAR-GARCH-MLP and LSTAR-LST-GARCH-MLP models.

References

Abramson, A., Cohen, I. (2007), “On the Stationarity of Markov-Switching GARCH Processes,” *Econometric Theory 23*, 485–500.

Abhyankar, A., Copeland, L. S., Wang, W. (1997). Uncovering nonlinear structure in real time stock market indexes: The Sand P 500, the DAX, the Nikkei 225, and the FTSE-100. Journal of Business and Economic Statistics, 15(1), 1–14.

Anderson, H.M., Nam, K. and Vahid, F., 1999. Asymmetric nonlinear smooth transition GARCH models. Rothman, P., editor, Nonlinear time series analysis of economic and fi nancial data, Kluwer, Boston, 191-207.

Alexander, C. Lazar, E. (2008), “Markov Switching GARCH Diffusion”, *ICMA Centre Discussion Papers in Finance* 2008-01.

Ane, T. and Ureche-Rangau, L. (2006), Stock market dynamics in a regime switching asymmetric power GARCH model, *International Review of Financial Analysis* 15, 109-129

Ang, A., Bekaert G.(2002), “International Asset Allocation with Regime Shifts”, Review of Financial Studies 15, 1137-1187.

Ang, A., Bekaert, G. (2001), “Regime Switches in Interest Rates,” Journal of Business and Economic Statistics,

Bartlett, P.L. (1997), "For valid generalization, the size of the weights is more important than the size of the network," in Mozer, M.C., Jordan,M.I., and Petsche, T., (eds.) Advances in Neural Information Processing Systems 9, Cambridge, MA: The MIT Press, pp. 134-140.

Bauwens, L., Rombouts, J. (2004), “Econometrics,” In J. Gentle,W. Hardle and Y. Mori (Eds.), *Handbook of Computational Statistics: Concepts and Methods*, 951–79. Heidelberg: Springer.

Bauwens, L. Preminger, A., Rombouts, J. (2006), “Regime Switching GARCH Models,” *CORE Discussion Paper* 2006/11, Universite Catholique de Louvain, Louvain La Neuve.

Bauwens, L. Preminger, A., Rombouts, J. (2007), “Theory and Inference for a Markov-Switching GARCH Model,” *CIRPEE Working Papers* No. 07-33.

Bauwens, L. Preminger, A., Rombouts, J. (2010), “Theory and Inference for a Markov-Switching GARCH Model,” *Econometrics Journal* 13, 218–244.
doi: 10.1111/j.1368-423X.2009.00307.x

Billing SA, Jamaluddin SA, Chen S. (1990), “A comparison of the back propagation and recursive prediction error algorithms for training neural networks”. Mech Sys Signal Process, International Journal Control ;55: 233–55.

Bildirici, M., Ersin, Ö. (2009), “Improving Forecasts of GARCH Family Models with the Artificial Neural Networks: An Application to the Daily Returns in Istanbul Stock Exchange”, Expert Systems with Applications, 36: 7355-7362

Binner J. M. , Elger C. T., Nilsson B., Tepper J. A., (2006), “Predictable non-linearities in U.S. inflation”, Economics Letters 93; 323–328

Bollen, N., Gray, S., Whaley, R. (2000), “Regime-Switching in Foreign Exchange Rates: Evidence From Currency Option Prices,” *Journal of Econometrics 94*, 239–76.

Bollerslev, T., (1986), “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31(3), April, 307–327.

Bollerslev, T., Mikkelsen, O., (1996) “Modeling and Pricing Long Memory in Stock Market Volatility,” *Journal of Econometrics*, 73(1), 151–184.

Brooks, R. D., Faff, R. W., McKenzie, M. D., and Mitchell, H. (2000), “A Multi-Country Study of Power ARCH Models and National Stock Market Returns,” *Journal of International Money and Finance*, 19, 377–397.

Brooks, C. (1998). Predicting stock index volatility: Can Market Volume help? *Journal of Forecasting,* 17, 59–98.

Cai, J. (1994): “A Markov Model of Switching–Regime ARCH,” *Journal of Business and Economic Statistics*, 12, 309–316.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, 108, 905–936.

Campbell, J.Y., A.W. Lo and A.C. MacKinlay, 1998, *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Castiglione, F. (2001). Forecasting price increments using an arti.cial neural network. Advances in Complex Systems, 4(1), 45–56.

Chan, K. (1993), “A Review of Some Limit Theorems of Markov Chains and Their Applications,” *In* H. Tong (Ed.) *Dimension Estimation and Models*, River Edge, NJ: World Scientific Publishing.

Chan F., 2011, Billy Theoharakis, Estimating *m*-regimes STAR-GARCH model using QMLE with parameter transformation, Mathematics and Computers in Simulation 81: 1385–1396

Chan F., and McAleer M., 2002. "Maximum likelihood estimation of STAR and STAR-GARCH models: theory and Monte Carlo evidence," Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 17(5): 509-534

Chan F., and McAleer M., 2003. "Estimating smooth transition autoregressive models with GARCH errors in the presence of extreme observations and outliers," Applied Financial Economics, Taylor and Francis Journals, vol. 13(8):581-592

Chang, B., Tsai, H. (2009), “Training Support Vector Regression by Quantum-Neuron-Based Hopfield Neural Net with Nested Local Adiabatic Evolution,” *International Journal of Innovative Computing, Information and Control* 5 (4), 1013-1026.

Chen Q and Li C.D. , (2006), “Comparison of Forecasting Performance of AR, STAR and ANN Models on the Chinese Stock Market Index”,

Chung, C.-F. (1999): “Estimating the fractionally intergrated GARCH model,” National. Taïwan University working paper.

Coakley, J. R., Brown, C. E. (2000). Artificial neural networks in accounting and .nance: Modeling issues. International Journal of Intelligent Systems in Accounting, Finance and Management, 9, 119–144.

Conrad and Haag (2006), “Inequality Constraints in the Fractionally Integrated GARCH Model”, *Journal of Financial Econometrics*, 4 (3), 413–449.

Conrad, Christian & Rittler, Daniel & Rotfuß, Waldemar, 2010. "Modeling and Explaining the Dynamics of European Union Allowance Prices at High-Frequency," Working Papers 0497, University of Heidelberg, Department of Economics

Davidson, J. (1994), *Stochastic Limit Theory*, New York: Oxford University Press.

Diebold, F. (1986), “Comment on Modelling the Persistence of Conditional Variances,” *Econometric Reviews 5*, 51–56.

Diebold, F. and A. Inoue (2001), “Long Memory and Regime switching”, *Journal of Econometrics 105*, 131–59.

Ding, Z., Granger, C.W.J., Engle, R.F. (1993). “ A Long Memory Property of Stock Market Returns and a New Model,” *Journal of Empirical Finance*, 1, 83-106.

Donaldson, R.G., M. Kamstra (1997), “An Artificial Neural Network-GARCH Model for International Stock Return Volatility,” *Journal of Empirical Finance* 4, 17-46.

Elman, J., 1990. Finding structure in time. Cognitive Science 14, 179–211.

Dufr´enot, G., V. Marimoutou and A. P´eguin-Feissolle (2002). ”LSTGARCH effects in stock returns : the case of US, UK and France”, International Conference on Forecasting Financial Markets, London, May

Engle, R.F. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of theVariance of United Kingdom Inflation,” *Econometrica* , 50, 987-1007.

Engle R.F., Bollerslev T. (1986), Modelleing the Persistence of Conditional Variances, Econometric Reviews, 5(1): 1-50

Engle, R.F., Ng V.K. (1993), “Measuring and Testing the Impact of News on Volatility,” *Journal of Finance*, 48, 1749-1778.

Engle, Robert F, (1990), Stock Volatility and the Crash of '87: Discussion," Review of Financial Studies, Oxford University Press for Society for Financial Studies, vol. 3(1), pages 103-06.

Faraway, J. & C. Chatfield (1998), "Time-series forecasting with neural networks: A comparative study using the airline data," *Applied Statistics*, **47**, 231-250.

Francq, C., M. Roussignol and J.-M. Zakoian (2001), "Conditional heteroskedasticity driven by hidden Markov chains," *Journal of Time Series Analysis 22*, 197–220.

Francq, C. and J.-M. Zakoian (2002), "Comments on the paper by Minxian Yang: 'Some properties of vector autoregressive processes with Markov-switching coefficients," *Econometric Theory 18*, 815–18.

Francq, C. and J.-M. Zakoian (2005), "The *l*2-structures of standard and switching-regime GARCH models," *Stochastic Processes and their Applications 115*, 1557–82.

Franses, P.H.B.F., Neele, J. & Dijk, D.J.C. van (1998). Modelling asymmetric volatility in weekly Dutch temperature data. Econometric Institute 9840/A

Freisleben, B. (1992), Stock market prediction with back propagation networks, Proceedings of the 5th international conference on industrial and engineering application of arti.cial intelligence and expert system, 451–460

Frömmel, M. (2007), "Volatility Regimes in Central and Eastern European Countries' Exchange Rates," *Faculteit Economie En Bedrijfskunde Working Paper*, 2007/487

Gallant, R., Rossi, P. E., and Tauchen, G. (1992). Stock prices and volume. Review of Financial Studies, 5, 199–242.

Garcia, R., Luger R., and Renault E. (2003), "Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables," Journal of Econometrics 116.

Gelfand, A. and A. Smith (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association 85*, 398–409.

Gerlach, R. and F. Tuyl (2006), "MCMC methods for comparing stochastic volatility and GARCH models," *International Journal of Forecasting 22*, 91–107.

Giraitis, L., Robinson, P.M., Surgailis, D., 2004. LARCH, leverage and long memory. Journal of Financial Econometrics 2,
177–210.

Giraitis, L., Leipus, R., Surgailis, D., 2005. Recent advances in ARCH modelling. In: Kirman, A., Teyssie`re, G. (Eds.), Long-memory in Economics. Springer, Berlin

Glosten, L.R., Jagannathan, R., Runkle, D.(1993). "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779-1801.

Goldfeld, Stephen M.,. Quandt R.E.,(1973), A Markov Model for Switching Regressions, Journal of Econometrics 1, 3-16.

Gonzalez-Rivera, G.,(1998), Smooth transition GARCH models. Studies in Non-linear Dynamics and Econometrics 3, 61.78.

Gray, S. (1996), .Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process., Journal of Financial Economics 42, 27-62.

Gupta A, Lam M (1998) The weight decay backpropagation for generalizations with missing values. Ann Oper Res 78:165–187.

Haas, M., Mittnik, S., and Paolella, M. S. (2004b): “A New Approach to Markov– Switching GARCH Models,” *Journal of Financial Econometrics*, 2, 493–530.

Hagerud, G.E. (1997), .A Smooth Transition ARCH Model for Asset Returns., SSE/EFI Working Paper Series in Economics and Finance, No.162.

Hamilton, J. D. and Susmel, R. (1994): “Autoregressive Conditional Heteroskedasticity and Changes in Regime,” *Journal of Econometrics*, 64, 307–333.

Hao, W., Yu, S. (2006), “Support Vector Regression for Financial Time Series Forecasting,” *in* International Federation for Information Processing (IFIP), Vol. 207, Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management, eds K. Wang, Kovacs, G., Wozni M., Fang, M., Boston: Springer, 825-830.

Hasanov, M. and Omay, T., (2007) “Are the transition stock markets efficient? Evidence from non-linear unit root tests”, Central Bank Review, vol:7(2), 1-7.

Hasanov, M. and Omay, T.,(2008) “Nonlinearities in Emerging Stock Markets: Evidence from Europe’s Two Largest Emerging Markets” Applied Economics. 40,(23),p.2645-2658.

Hentschel, L. (1995). “All in the Family: Nesting Symmetric and Asymmetric GARCH Models**,**” *Journal of Financial Economics*, 39, 71-104.

Higgins, M. L, Bera, A. K. (1992). “A Class of Nonlinear ARCH Models,” *International Economic Review*, 33, 137 -158.

Hiemstra, C., Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. The Journal of Finance, 49, 1639–1664.

Hilebrand, E. (2005), “Neglecting parameter changes in GARCH models. *Journal of Econometrics 129*, 121–38.

Hu, L., Shin, Y. (2007), “Optimal Test for Markov Switching GARCH Model,” *Studies in Nonlinear Dynamics & Econometrics* 12(3), Article 3.

Huang, S., Wu, T. (2008), “Integrating GA-based Time-scale Feature Extractions with SVMs for Stock Index Forecasting,” *Expert Systems with Applications* 35, 2080-2088.

Hutchinson J.M., Lo A.W., Poggio T., (1994), "A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks", *The Journal of Finance*, Vol. 49, No. 3, 851-889

Jeanne, Olivier and Paul Masson (2000), "Currency Crises, Sunspots, and Markov-Switching Regimes," Journal of International Economics 50, 327-350.

Jeng, J., Chuang, C., Su, S. (2003), "Support Vector Interval Regression Networks for Interval Regression Analysis," *Fuzzy Sets and Systems* 138, 283-300.

Jordan, M., 1986. Attractor Dynamics and Parallelism in a Connectionist Sequential Machine. Proceedings of the 8th Annual Conference of the Cognitive Science Society, pp. 531–545.

Jorian P. (2000).Value at risk: The New Benchmark for Managing Financial Risk, *2ed.,* NewYork: McGraw Hill.

İnce, H., Trafalis, T. (2006), "A Hybrid Model for Exchange Rate Prediction," *Decision Support Systems* 42, 1054-1062

Karanasos, M., Psaradakis, Z., Sola, M., 2004. On the autocorrelation properties of long-memory GARCH processes. Journal of Time Series Analysis 25, 265–281

Kaufman, S. and S. Frühwirth-Schnatter (2002), "Bayesian analysis of switching ARCH models," *Journal of Time Series Analysis 23*, 425–58.

Klaassen, F. (2002), "Improving GARCH volatility forecasts with regime-switching GARCH," *Empirical Economics 27*, 363–94.

Kılıç, R. (2010), "Long memory and nonlinearity in conditional variances: A smooth transition FIGARCH model," *Journal of Empirical Finance* (2010), doi:10.1016/j.jempfin.2010.11.007

Kramer, W. (2008), "Long Memory with Markov-Switching GARCH", Cesifo WP no. 2225

Krogh, A., Hertz, J. (1995), "A Simple Weight Decay Can Improve Generalization,"In *Advances in Neural Information Processing Systems 4,* J. Moody, S. Hanson and R. Lippmann, eds. Morgan Kauffmann Publishers, San Mateo: California.

Lamoureux, C. and W. Lastrapes (1990), "Persistence in variance, structural change, and the GARCH model," *Journal of Business and Economic Statistics 8*, 225–34.

Lapedes, A. and Farber, R. (1987). Nonlinear signal processing using neural networks. Proceedings of the IEEE conference on neural information processing system-natural and synthetic.

Lee J. and Degennaro, R. 2000. Smooth transition ARCH models: Estimation and testing. Review of Quantitative Finance and Accounting 15, 5-20.

Lundberg, S., and TerÄasvirta, T., 1998. Modeling economic high-frequency time series with

STAR-STGARCH models. Department of Economics, Stockholm School of Economics.

Luukkonnen, R., Saikkonnen, P., Terasvirta, T. (1988), “Testing Linearity Against Smooth Transition Autoregressive Models,” Biometrika, 75, 491-499.

Maheu, J. and Z. He (2009), “Real time detection of structural breaks in GARCH models,” Working paper, Department of Economics, University of Toronto.

Mandic D., Chambers J.A., (2001), *Recurrent Neural Networks for Prediction ,* John Wiley and Sons

Mashor, M. Y. (2000). “Hybrid Multilayered Perceptron Networks”, *International Journal of System Science*, Vol. 31. No. 6. pp. 771-785.

McKenzie, M. & Mitchell, H., (2002),"Generalized Asymmetric Power ARCH Modelling of Exchange Rate Volatility," *Applied Financial Economics*, 12(8): 555-64.

McMillan, D.G. (2003) “Non-linear Predictability of UK Stock Market Returns”, Oxford Bulletin of Economics and Statistics, vol. 65, issue 5, pp. 557-573

Meitz, S. and P. Saikkonen (2008), “Stability of nonlinear AR-GARCH models,” *Journal of Time Series Analysis 29*, 453–75.

Mikosch, T. and C. Starica (2004), “Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects,” *Review of Economics and Statistics 86*, 378–90.

Moeanaddin, R. and H. Tong (1990), “Numerical evaluations of distribution of non-linear autoregression,” *Journal of Time Series Analysis 11*, 33–48.

Müller, K., Smola, A., Raetsch, G., Shoelkopf, B., Kohlmorgen, J., Vapnik, V. (1997), “Predicting Time Series with Support Vector Machines,” *Proceedings of ICANN’97,* Springer, LNCS 1327, 999-1004.

Nam K., Pyun C.S. and Arize A.C., 2002, Asymmetric mean-reversion and contrarian profits: ANST-GARCH approach, Journal of Empirical Finance 9 (2002) 563– 588

Nelson, D. B. (1990), “Stationarity and persistence in the GARCH(1,1) model,” *Econometric Theory 6*, 318–34.

Nelson, D.B. (1991). “Conditional Heteroscedasticity in Assets Returns: A New Approach,” *Econometrica*, 55, 703-708.

Nelson, D.B. (1992), “Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model,” *Journal of Econometrics*, 52, 61-90.

Nelson, M., Hill, T., Temus, W., O’Connor, M. (1999). Time series forecasting using neural networks: Should the data be deseasonalized .rst? Journal of Forecasting, 18, 359–367.

Pai, P., Lin, C., Hong, W., Chen, C. (2006), “A Hybrid Support Vector Regression for Exchange Rate Prediction,” *Information and Management Sciences* 17(2), 19-32.

Phua, P. K. H., Ming, D. (2003). Parallel nonlinear optimization techniques for training neural networks. IEEE Transactions on Neural Networks, 14(6), 1460–1468.

Phua, P. K. H., Zhu, X., Koh, C. (2003). Forecasting stock index increments using neural networks with trust region methods. Proceedings of IEEE International Joint Conference on Neural Networks, 1, 260–265.

Potter, SM., 1995a, A nonlinear approach to U.S. GNP, Journal of Applied Econometrics 10, 109-125
.
Potter, S.M. 1995b, Nonlinear impulse response functions, Mimeo, Department of Economics, UCLA.

Refenes, A. P. N., Burgess, A. N., and Bentz, Y. (1997). Neural networks in financial engineering: a study in methodology. IEEE Transactions on Neural Networks, 8(6), 1222–1267.

Refenes, A. N., Zapranis, A., Francies, G. (1994). Stock performance modeling using neural networks: a comparative study with regression models. Neural Networks, 5, 961–970.

Vapnik, V. (1995), The Nature of Statistical Learning Theory, Springer Verlag.

Taylor, M. (1994), "Historical Surface Level of the Great Salt Lake". Master's Thesis, Utah State University, Logan, Utah, USA.

Teräsvirta, Timo. (1994) "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models". *Journal of the American Statistical Association.* c. 89. s. 425: 208-218.
Tse, Y.K. and Tsui, A.K.C. (1997) "Conditional Volatility in Foreign Exchange Rates : Evidence fro the Malaysian Ringgit and Singapore Dollar" *Pacific-Basin Finance Journal* (5) pp. 345 – 356

Resta, M. (2000). Towards an arti.cial technical analysis of financial markets. Proceedings of the IEEE–INNS–ENNS international joint conference on neural networks, 5, 117–122.

Robert, C. and G. Casella (2004), *Monte Carlo Statistical Methods*, New York: Springer.

Schwert, G. (1989), "Why does stock market volatility change over time?" *Journal of Finance 44*, 1115–53.

Spezia, L., Paroli, R. (2008), "Bayesian Inference and Forecasting in Dynamic Neural Networks with Fully Markov Switching ARCH Noises," *Communications in Statistics—Theory and Methods*, 37: 2079–2094.

Stachurski, J. and V. Martin (2008), "Computing the distributions of economic models via simulations," *Econometrica 76*, 443–50.

Sims, Christopher, and Tao Zha (2004), "Were There Switches in U.S.Monetary Policy?", working paper, Princeton University.

Sitte, R., and Sitte, J. (2000). Analysis of the predictive ability of time delay neural networks applied to the Sand P 500 time series. IEEE Transactions on Systems, Man and Cybernetics – Part C: Applications and Review, 30(4), 568–572.

Swanson N. and H. White, 1995. A model selection approach to assessing the information in the term structure using linear models and artificial neural networks, Journal of Business and Economics Statistics, 13, 265-275.

Tanner, M. and W. Wong (1987), "The calculation of the posterior distributions by data augmentation," *Journal of the American Statistical Association 82*, 528–40.

Terasvirta, T.; Anderson, H.M. (1992), Characterizing nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics, 7,119–136

Thierry, A., Ureche-Rangau, L., (2006), "Stock Market Dynamics in a Regime-Switching Asymmetric Power GARCH Model," *International Review of Financial Analysis*, Elsevier, vol. 15(2), pages 109-129.

Tino, P., Schittenkopf, C., and Dorffner, G. (2001). "Financial Volatility Trading using Recurrent Neural Networks," IEEE Transactions on Neural Networks, 12(4), 865–874.

Tjostheim, D. (1990), "Non-linear time series and Markov chains," *Advances in Applied Probability 22*, 587–611.

Tsay, R. (2005), *Analysis of Financial Time Series*, New York: John Wiley.

Wang, H. (2005). Flexible .ow shop scheduling: Optimum, heuristics, and AI solutions. Expert Systems, 22(2), 78–85.

Wang, H., Jacob, V., and Rolland, E. (2003). Design of e.cient hybrid neural networks for flexible .ow shop scheduling. Expert Systems, 20(4), 208–231.

Wang, X., Phua, P. H. K., and Lin, W. (2003). Stock market prediction using neural networks: Does trading volume help in short-term prediction? Proceedings of IEEE International Joint Conference on Neural Networks, 4, 2438–2442.

Weigend, A. S., Rumelhart, D. E., & Huberman, B. A. (1991). Generalization by weight-elimination with application to forecasting. In: R. P. Lippmann, J. Moody, & D. S. Touretzky (eds.), Advances in Neural Information Processing Systems 3, San Mateo, CA: Morgan Kaufmann.

Weigend, B.A. Huberman, D.E. Rumelhart, Predicting sunspots and exchange rates with connectionist networks, in: Proceedings of the 1990 NATO Workshop on Nonlinear Modeling and Forecasting, Addison Wesley, Santa Fe, NM, USA. 1991.

Weigend A.S., Mangeas M., Srivastava A.N., "Nonlinear gated experts for time series: discovering regimes and avoiding overfitting", International Journal of Neural Systems 6 (1995) 373-399.

White H. (1992), *Artificial Neural Networks: Approximation and Learning Theory*, Blackwell: Oxford.

Yang, M. (2000), "Some properties of vector autoregressive processes with Markov-switching coefficients," *Econometric Theory 16*, 23–43

Yao, J. T., and Poh, H.-L. (1996), "Equity Forecasting: A Case Study on the KLSE Index," In A.-P. N. Refenes, Y. Abu-Mostafa, J. Moody, and A. Weigend (Eds.), Neural Networks in Financial Engineering, *Proceedings of the 3rd International Conference on Neural Networks in the Capital Markets* (pp. 341–353). World Scientific.

Yao, J. T., and Tan, C. L. (2000), "Time Dependent Directional Profit Model for Financial Time Series Forecasting," *Proceedings of the IEEE–INNS– ENNS International Joint Conference on Neural Networks*, 5, 291–296.

Yao, J. T., Tan, C. L., and Poh, H. L. (1999), "Neural Networks for Technical Analysis: A Study on KLCI," *International Journal of Theoretical and Applied Finance*, 2 (2), 221–241.
Yao, J. (2001), "On square-integrability of an AR process with Markov switching," *Statistics and Probability Letters 52*, 265–70.

Yao, J. and J.-G. Attali (2000), "On Stability of Nonlinear Process with Markov Switching," *Advances in Applied Probability 32*, 394–407.

Zaffaroni, P., 2004. Stationarity and memory of ARCH(l) models. Econometric Theory 20, 147–160.

Zakoian, J.M., (1994). "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control*, 18, 931-955

Zhang, G., Patuwo, B.E., Hu, M.Y. (1998), "Forecasting with Artificial Neural Networks: the State of the Art," *International Journal of Forecasting* 14, 35–62.

Zhang, M. Y., J. Russell and R. Tsay (2001), "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data," *Journal of Econometrics 104*, 179–207.
Zhang, Yue-Jun & Wei, Yi-Ming, 2010. "An overview of current research on EU ETS: Evidence from its operating mechanism and economic effect," Applied Energy, Elsevier, vol. 87(6):1804-1814.

Zhu, X., H. Wang, L. Xu, H. Li (2007), "Predicting Stock Index Increments by Neural Networks: The Role of Trading Volume under Different Horizons," *Expert Systems with Applications*. doi:10.1016/j.eswa.2007.06.023 [www.elsevier.com/locate/eswa].