MPRA

Munich Personal RePEc Archive

Nonlinear Combination of Financial Forecast with Genetic Algorithm

Ozun, Alper and Cifter, Atilla

Marmara University

01. February 2007

Nonlinear Combination of Financial Forecast with Genetic Algorithm

Alper Ozun[a*],   Atilla Cifter [b†]

[a] *Marmara University, Department of Informatics, and Isbank, Board of Inspectors/Market Risk Group, Istanbul, Turkey*
[b] *Marmara University, Department of Econometrics, and Deniz Yatirim-Dexia Group, Financial Reporting, Istanbul, Turkey*

February 2007

Abstract

Complexity in the financial markets requires intelligent forecasting models for return volatility. In this paper, historical simulation, GARCH, GARCH with skewed student-t distribution and asymmetric normal mixture GRJ-GARCH models are combined with Extreme Value Theory Hill by using artificial neural networks with genetic algorithm as the combination platform. By employing daily closing values of the Istanbul Stock Exchange from 01/10/1996 to 11/07/2006, Kupiec and Christoffersen tests as the back-testing mechanisms are performed for forecast comparison of the models. Empirical findings show that the fat-tails are more properly captured by the combination of GARCH with skewed student-t distribution and Extreme Value Theory Hill. Modeling return volatility in the emerging markets needs "intelligent" combinations of Value-at-Risk models to capture the extreme movements in the markets rather than individual model forecast.

JEL classification: C52; C32; G0

Keywords: *Forecast combination, Artificial neural networks, GARCH models, Extreme value theory, Christoffersen test*

1. Motivation

Risk has been numerated since 1996 in which JPMorgan introduced first Value-at-Risk (VaR) methodology known as RiskMetrics. The VaR has been widely used in risk management and capital allocation under different normality assumptions and algorithms. Difference in methodology leads the model risk to appear in risk measurement. To minimize the model risk, alternative methodologies in the calculation of risk amount and back-testing mechanism

---

[*] PhD. in Banking, PhD. Student in Informatics, Marmara University, and Inspector, Board of Inspectors/Market Risk Group, Is Bank, Istanbul Turkey Tel: +90-212-316 4009 Address: Turkiye Is Bankasi A.S. Genel Müdürlük Teftis Kurulu Baskanligi Kat:34 Is Kuleleri 34330 Levent Istanbul Turkey, *E-Mail address*: Alper.Ozun@isbank.com.tr

[†] PhD Student in Econometrics, Marmara University, and Financial Reporting Vice President, Deniz Yatırım-Dexia Group, Istanbul Turkey Tel: +90-212-354 8373 Address: Denizbank Genel Mudurluk Kat:16 Buyukdere Cad. No:106 34394 Esentepe Istanbul Turkey, *E-Mail address:* Atilla.Cifter@denizyatirim.com

should be compared and combined by using intelligent systems, which are capable of capturing the complexity in the financial markets.

There are mainly three categories for VaR estimation, which are namely parametric methods, non-parametric methods and semi-parametric methods. Parametric methods estimate VaR by fitting some particular distribution to observed returns. In general GARCH and its derivatives with normal or asymmetric distributions are used for the volatility calculations in the parametric models. The parametric methods have two main drawbacks. First of all, they calculate VaR by using assumptions on return distribution, which are not proper for the tail estimation. Secondly, parametric methods model the central observations while the risk managers want to know the excess loss (fat-tail) from their capital allocated for a specific portfolio. Estimating the observations on the center of the return distribution is not what the portfolio managers are interested in.

Non-parametric methods, the most widely used one is historical simulation, estimate VaR by reading off it from an appropriate histogram of returns. In that perspective, historical simulation has two main drawbacks, as well. Firstly, it is not interested in what the tails should look like, and secondly, does not give any clue about VaR beyond the sample range.

On the other hand, Extreme Value Theory (EVT) as a semi-parametric model designed for tail estimation is free of the backwards that parametric models have. McNeil (1996) argues that the distribution of extreme returns converges asymptotically to:

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp(-[1+\xi(x-\mu)/\sigma]^{-1/\xi}) \\ \exp(-e^{-(x-\mu)/\sigma}) \end{cases}, \tag{1}$$

If μ and σ are respectively the mean and standard deviation, while ξ indicates the size of the tails in that the higher ξ, the heavier the tail. The EVT is based on the central limit theorem applied to the extremes rater than mean. It limits the distribution of extreme returns always has the same form without relying on the distribution of the parent variable. Neftci (2000) lists three main reasons to use EVT rather than parametric methods in VaR estimation. First of all, a histogram for a time series can always be estimated accurately at or near the center of the distribution. For that reason, there is less need for a priori models yielding closed-form formulas for the underlying distributions. However, extremes are rare and located at the tails. Secondly, the distribution of asset price increments is heavy-tailed and asymmetric. For that reason, EVT approximating the tail areas asymptotically might be more successful than imposing an explicit functional form. Lastly, extreme movements in asset prices might be caused by mechanisms that are structurally different from the usual dynamics of markets. Extreme observation might be the result of a major default or a speculative bubble. During these extreme conditions, the distributional characteristics of the time series might shift, which requires separating tail estimation from estimation of the rest of the distribution.

Whether a parametric or non-parametric model is used for the VaR estimation, there is always a model risk if the model does not fit the dynamics of the markets. The scientific problem in this article is if different methods can be used for the same financial variable in a combined perspective to increase the forecast power. Some behaviours of time series of a financial variable might be forecasted by parametric or semi-parametric methods while some of them is

done by non-parametric methods. Therefore, an "intelligent" combination for the models might be a solution to deal with the model risk in risk measurement.

In this article, we firstly use GARCH with Gaussian distribution, GARCH with skewed Student-t distribution, asymmetric normal mixture GRJ-GARCH, historical simulation and EVT-Hill models individually to forecast daily returns of the Istanbul Stock Exchange. Then, we combine the individual parametric and non-parametric models with EVT-Hill by using artificial neural networks with genetic algorithm. The forecast performances of the models are compared by two different back-testing mechanisms, namely Kupiec and Christoffersen tests.

In the next chapter, recent literature on VaR and EVT is reviewed. In the third part, after explaining the GARCH models, historical simulation and EVT Hill methodologies and back-testing methods, the combination models with artificial neural networks tested in this article are presented in the third part, as well. Empirical results of the back-tests are discussed in a comparative perspective in the fourth part. The article ends with certain suggestions for the future research on the combination models and emphasizes the importance of intelligent systems in risk modeling.

2. Alternative Models for Value-at-Risk Estimation

Value-at-Risk is defined as the conditional quantile of the portfolio loss distribution for a time-horizon and at a significant level. On the other hand, Extreme Value Theory, which deals with what the distribution of extreme values look like within given limits, concentrates on the expected loss beyond the VaR. Methodologies of the alternative forecast models are examined in the following.

2.a. Parametric Approaches: GARCH Models With Alternative Assumptions on Return Distribution

Parametric models estimate VaR by fitting a particular distribution to a set of observed returns. In general, GARCH and its derivatives are proper to use in volatility calculations for the parametric VaR models. GARCH models might be applied with normal or asymmetric distribution. However, regardless of their assumptions on the return distribution, they are restricted models in that they estimate VaR with a scope of given distributional assumption. In this paper, we use three different GARCH models to estimate VaR, namely, GARCH with Gaussian distribution, GARCH with skewed Student's-t distribution and Normal Mixture GRJ-GARCH.

In a static linear model ($y_i = \alpha + \beta \chi_i + \varepsilon_i$), the error term ($\varepsilon_i$) is a random variable with normal distribution. The variance on the other hand, is constant as showed in the Equation 2.

$$E(\varepsilon_i - 0)^2 = E(\varepsilon_i)^2 = \sigma_\varepsilon^2 \tag{2}$$

Engle (1982) creates ARCH model to display the time-varying variance.

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_i^2 = \omega + \alpha(L)\varepsilon_t^2 \tag{3}$$

In the Equation 3, σ_i is the conditional variance of ε_t and chances on time. The ARCH model is restricted because the sum of $\alpha_i > 0$ and α should be 1. Bollerslev (1986) introduces GARCH model to estimate volatility with negative variance. The model captures the effects of both the linear variance and conditional variance of the past.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sum_{t=2}^{n} \varepsilon_{t-1}^2 + \beta_1 \sum_{t=2}^{n} \sigma_{t-1}^2 \tag{4}$$

GARCH models are used with alternative assumptions. In normal distribution, skewness and kurtosis take the value of (0, 3). Peters (2001) formulizes the normal distribution on the Equation 4 where T denotes observation number.

$$L_T = -\frac{1}{2} \sum_{T=1}^{T} \left[In(2\pi) + In(\sigma_t^2) + z_t^2 \right] \tag{5}$$

Bollerslev(1987), Hsieh(1989), Baillie and Bollerslev(1989), and Palm and Vlaar(1997) empirically display that the performance of fat-tailed distributions like Student's-t distribution is better in capturing higher observed kurtosis. The log-likelihood function of the student's-t distribution is presented on the Equation 6 (Saltoglu, 2003).

$$l^{t-dist}(\theta) = T \left\{ In\Gamma(\frac{v+1}{2}) - In\Gamma(\frac{v}{2}) - \frac{1}{2} In[\pi(v-2)] \right\} \tag{6}$$

$$-\frac{1}{2} \sum_{t=1}^{T} (In(h_t) + (1+v)In(1+\frac{\varepsilon_t^2}{v-2})$$

Since both Gaussian and Student's-t distributions are symmetric; Fernandez and Steel (1998) uses asymmetric Student's-t distribution in Value-at-Risk estimation, which captures both asymmetry and fattailedness. The log-likelihood of a standardized skewed student's-t is displayed on Equation 7 where $\Gamma(.)$ is the gamma function (Peters, 2001).

$$l_{skewed-st} = T \left\{ In\Gamma(\frac{\eta+1}{2}) - In\Gamma(\frac{\eta}{2}) - 0.5In[\pi(\eta-2)] + In(\frac{2}{\xi+\frac{1}{\xi}}) + In(s) \right\}$$
$$-0.5 \sum_{t=1}^{T} \left\{ In\sigma_t^2 + (1+\eta)In\left[1 + \frac{(s\varepsilon+m)^2}{\eta-2} \xi^{-2I_t} \right] \right\} \tag{7}$$

An alternative GARCH model with asymmetric distribution is constructed by Gloslen, Jagannathan and Runkle (1993). They argue that asymmetry in the return volatility might be captured by adding a dummy variable into GARCH model. The model is called as GJR-GARCH model and presented on the Equation 8.

$$\sigma_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 u_{t-1}^2 I_{t-1} + \beta_1 \sigma_{t-1} \tag{8}$$

Palm (1996), Pagan (1996) and Bollerslev, Chou, and Kroner (1992) show that the use of a fat-tailed distributions is improved the forecast performance. Bollerslev (1987), Hsieh (1989), Baillie and Bollerslev (1989) and Palm and Vlaar (1997) also state that these distributions perform better in order to capture the higher observed kurtosis.

In a recent study, Christoffersen and Jacobs (2004) show that a simple asymmetric GARCH model that captures the leverage effect performs best of all GARCH models. Bekaert and Wu (2000) and Wu (2001) empirically prove that the leverage effect in stocks determines a strong negative correlation between returns and volatility which might be the most important source of skewness in returns. Christoffersen, Heston and Jacobs (2004) and Bates (1991) also focus on the relationship between time-variability in the physical conditional skewness and the empirical characteristics of option implied volatility skews.

Recently, to capture switching, shocks and long-term memory in the stock returns, Normal Mixture GARCH (NM-GARCH) model is constructed by Alexander and Lazar (2005). In the NM-GARCH model, the individual variances are only tied with each other through their dependence on the error term.

The asymmetric normal mixture GARCH model has one equation for the mean and K conditional variance components representing different market conditions. The error term has a conditional normal mixture density with zero mean as a weighted average of K normal density functions with different means and variances Alexander and Lazar (2005).

$$\varepsilon_t / I_{t-1} \sim NM(p_1,.....,p_K,\mu_1,.....,\mu_K,\sigma_{1t}^2,......,\sigma_{Kt}^2), \quad \sum_{i=1}^{K} p_i = 1, \quad \sum_{i=1}^{K} p_i \mu_i = 1 \qquad (9)$$

The conditional density of the error term is calculated from the Equation 9:

$$\eta(\varepsilon_t) = \sum_{i=1}^{K} p_i \varphi_i \qquad (10)$$

On the Equation 9, φ denotes normal density functions with different constant means μi and different time varying variances σ_{it}^2 for i = 1,…, K.

The model assumes that K variances follow normal mixture GARCH processes. The NM-GARCH is formulated in the Equation 11.

$$\sigma_{it}^2 = \alpha_0 + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2 \qquad \text{for } i=1,....,K \qquad (11)$$

NM-GJR GARCH based on Gloslen, Jagannathan and Runkle (1993) is given by Equation 12.

$$\sigma_{it}^2 = \alpha_0 + \alpha_i \varepsilon_{t-1}^2 + -\lambda_i d_{t-1}^- \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2 \quad \text{for i=1,....,K;} \qquad (12)$$

where $d_t^- = 1$ if $\varepsilon_t < 0$, and 0 otherwise.

For both models, the overall conditional variance is

$$\sigma_t^2 = \sum_{i=1}^{K} p_i \sigma_{it}^2 + \sum_{i=1}^{K} p_i \mu_i^2 \tag{13}$$

According to Alexander and Lazar (2005), when K is bigger than 1, the existence of second, third and fourth moments are assured by imposing less stringent conditions than in the single component in which K is equal to 1. For asymmetric NM-GARCH models, the conditions for the non-negativity of variance and the finiteness of the third moment are shown on the Equation 14.

$$0 < p_i < 1, \quad i=1,....,K\text{-}1, \quad \sum_{i=1}^{K-1} p_i < 1, \quad 0 < \alpha_i, \quad 0 \le \beta_i < 1 \tag{14}$$

To construct NM-GRJ GARCH Model, The Equation 14 can be used:

$$m = \sum_{i=1}^{K} p_i \mu_i^2 + \sum_{i=1}^{K} \frac{p_i \omega_i}{(1-\beta_i)} > 0, \; n = \sum_{i=1}^{K} \frac{p_i(1-\alpha_i - 0.5\lambda - \beta_i)}{(1-\beta_i)} > 0 \tag{15}$$

$$\text{and } \omega_i + (\alpha_i + 0.5\lambda)\frac{m}{n} > 0$$

The NM-AGARCH and NM-GRJ GARCH models have dynamic asymmetry when the λ_i parameters in the component variance processes capture time-varying short-term asymmetries arising from the leverage effect. If λ_i is positive, the conditional variance is higher following a negative unexpected return at time t − 1 than following a positive unexpected return. In the markets, since negative news corresponds to a negative unexpected return, positive λ_i should be expected (Alexander and Lazar, 2005).

2.b. Non-Parametric Approach: Historical Simulation

Whether it employs Gaussian or asymmetric distribution, the GARCH models are restricted because of their assumptions on distributional characteristics of returns. Historical simulation method as a non-parametric model, which uses historical observed values, on the other hand, does not have any assumptions on the statistical characteristics of the data set.

Historical simulation method consists of going back in time, and applying current weights to a time-series of historical returns. The return does not represent an actual portfolio but rather reconstructs the history of a hypothetical portfolio using the current position. The calculation of VaR with historical simulation is simple and requires only for each risk factor, a time-series of actual movements; and positions on risk factors.

By following Liu (2005), we can formulize historical simulation method as follows. In the model, portfolio return on t+1 is defined as $R_{p,t+1}$. The model assumes that the distribution of tomorrow's portfolio returns, $Rp;t+1$ is identical to the empirical distribution of the past m periods' portfolio return, $(R_{p,t+1 - \tau})^m \tau = 1$. The VaR with coverage rate p, is simplified as the 100pth percentile of the sequence $(R_{p,t+1 - \tau})^m \tau = 1$. When the returns $(R_{p,t+1 - \tau})^m \tau = 1$ are sorted in ascending order, the 100p percentile of the sequence is the $VaR^p p;t+1$.

Return distributions might be non-normal but they are all normally distributed, the VAR obtained under the historical-simulation method should be the same as that under the GARCH normal method. Since the historical method might not adequately represent future distributions, it is not widely used in complex markets. In general, according to the simulations constructed by Liu (2005), the GARCH models adapts structural break rapidly but is less stable, on the other hand, historical simulation method adapts the break slowly but is more stable.

2.c. Semi-Parametric Approach: Extreme Value Theory For Tail Estimation

Extreme Value Theory is introduced by McNeil and Frey (2000) under the assumption that the tail of the conditional distribution of the GARCH is approximated by a heavy-tailed distribution. The model supposes that the tail of the conditional distribution of error term ε_t is approximated by the distribution function

$$F(z) = 1 - L(z) z^{-1/\zeta} \sim 1 - cz^{-1/\varepsilon} \tag{16}$$

Whenever $\varepsilon_t > u$, where $L(z)$ is a slowly changing function approximating with a constant c, and ζ is a positive parameter. u is a threshold value such that all observations above u is used in the estimation of ζ. The Hill estimator (Hill, 1975), $\hat{\zeta}$, is the MLE estimator of ς under the assumption that the standardized residuals $\hat{\varepsilon}_t$ are approximately i.i.d. The Hill Estimator is formulized as

$$\hat{\zeta} = 1/Tu \sum_{t=1}^{Tu} \ln(\hat{\zeta}_{(T-t+1)}) - \ln(u) \tag{17}$$

where $\hat{\zeta}_{(t)}$ denotes the t-th oder statistics of $\hat{\zeta}_t$; and T_u denotes the number of observations that exceed u. Given $\hat{\zeta}$, an estimate of the tail distribution F is obtained by choosing $c = (T_u/T)u^{1/\hat{\zeta}}$, which derives from imposing the condition $1 - F(u) = T_u/T$. We obtain the following estimate of F:

$$\hat{F}(z) = (1 - (T_u/T))(z/u)^{-1/\hat{\zeta}} \tag{18}$$

The EVT relies on $\hat{F}(z)$ to estimate the constants $c_{1,p}$ and $c_{2,p}$. In particular, the estimate of $c_{1,p}$ is equal to $(\hat{F})^{-1}_{1-p}$, the (1-p)th-quantile of the tail distribution $\hat{F}$. We can show that

$$\hat{c}^{Hill}_{1,p} = u(p(T_u/T))^{-\hat{\zeta}} \tag{19}$$

Similarly, to compute an estimate of $c_{2,p}$ we use $\hat{F}(z)$ to compute $E(\varepsilon \mid \varepsilon > \hat{F}^{-1}_{1-p})$, where $\varepsilon \sim$ i.i.d. $\hat{F}$. We can show that the following closed form expression holds true

$$E(\varepsilon \mid \varepsilon > \hat{F}^{-1}{}_{1\text{-}p}) = (\hat{F}^{-1}{}_{1\text{-}p}) / (1\text{-} \hat{\zeta}). \tag{20}$$

This implies the following Hill's estimate of $c_{2,p}$:

$$\hat{c}^{\text{Hill}}{}_{2,p} = (\hat{c}^{\text{Hill}}{}_{1,p}) / (1\text{-} \hat{\zeta}). \tag{21}$$

The Hill's estimates of $\text{VaR}^{p}{}_{T+1}$ and $\text{ES}^{p}{}_{T+1}$ are given by

$$\text{Hill-VaR}^{p}{}_{T+1} = \hat{\sigma}_{T+1} + \hat{c}^{\text{Hill}}{}_{1,p} \tag{22}$$
$$\text{Hill-EVT}^{p}{}_{T+1} = \hat{\sigma}_{T+1} + \hat{c}^{\text{Hill}}{}_{2,p} \tag{23}$$

respectively.

2.d. Artificial Neural Networks As The Combination Algorithm

It might not be possible to capture the behaviours of the financial markets with indvidual models due to high complexity in the market dynamics. Asymmetric information flow, non-linear behaviours, turbulences and chaos in the markets require advanced modeling methodologies. Combination of forecast methods might be a solution to capture multi-characteristics of the financial markets. In that point, with their flexible structure, artificial neural networks might be a proper solution as a combination algorithm. Ability of the artificial neural networks to model both deterministic and random characteristics makes it ideal for capturing chaotic patterns. Maasoumi, Khotanzad, and Abaye (1994) state that it is necessary to have non-linear platforms to discern relationships among time-series due to dynamic nature of the financial time series; and artificial neural networks are most proper systems to discover discern relationships. Researches display that artificial neural networks have many advantages over econometric methods. Since artificial neural networks do not have any assumptions about the nature of the distribution of the returns, they are free of bias in their analysis. Instead of certain assumptions about the underlying population, the networks with hidden layers use the data to develop an internal representation of the relationship between the variables. In this way, better results can be performed with neural networks when the relationship among the variables does not fit an assumed model (Ozun, 2006).

In this research, we use multilayer perceptron network to combine the VaR models. The power of neural networks comes when neurons are combined into the multi-layer structures. Hecht and Nielsen (1990) argues that a three layer multi-layer perceptron neural network is capable of approximating any mapping function. White (1994) states that multi-layer perceptron models are non-linear neural network models that can be used to approximate almost any function with a high degree of accuracy. A multi-layer perceptron neural network figure is presented below (Ozun, 2006).
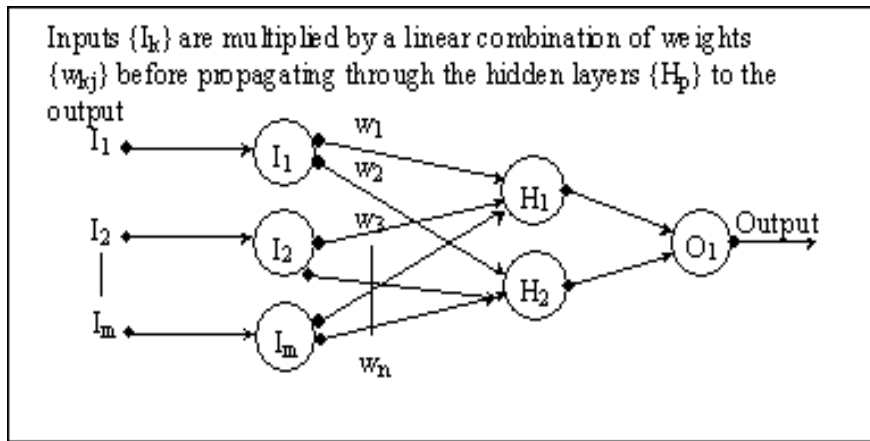
Fig. 1. A Multi-Layers Perceptron Neural Network
Source: Sheikh, S., Understanding and Implementing Neural Networks, Neuropsychology and Instructional Design, 2005

Rumelhart and McClelland (1986) define a neural network as a mapping function from a set of input variables to a set of output values. The nodes are designed in layers: i) an input layer ii) hidden layers, and iii) an output layer. The input layer transfers the information. The layers those where the nodes process the information passed to them by the input layer are named as hidden layers. The hidden layer includes at least three of four input nodes and the initial weights of the connections can be chosen randomly. Finally, The layer where an output pattern from a given input pattern processing through the preceding layers is called as the output layer. In this research, the candidate VaR forecasts are the input, the combined VaR forecasts are the output. Liu (2005) states that the difference between the VaR-Artificial Neural Networks model and standard Artificial Neural Networks is that the latter deals in with mean forecast and the cost function is a symmetric differentiable function, while for VaR models, we are interested in the quantile forecast and the cost function is an asymmetric non-differentiable tick loss function. For that technical reason, we follow Liu (2005) and employ Genetic Algorithm to train the combination.

Genetic Algorithm as a solution for optimization problems based on natural selection is introduced by Holland (1965). Genetic Algorithm keeps an initial population of solution candidates and evaluates the quality of each solution candidate according to a specific cost function. GA repeatedly modifies the population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution.

In the finance literature, there are not so many researches examining the forecast combination with intelligent systems. Shanming and Bao (1993) provide an approach to nonlinear combination of forecasts based on neural networks. The test result suggests that the nonlinear combination of forecasts with neural networks is an effective way for combining forecasts. Palit and Popovic (2000) propose the nonlinear combination of time series forecasts based on isolated use of neural networks, fuzzy logic and neuro-fuzzy systems. It is showed that the nonlinear combination of a group of forecasts based on intelligent systems is able to produce a single better forecast than any individual forecasts involved in the combination. Liu (2005) empirically show that based on MLC, Violation Ratio and Christofferson's conditional coverage test, the ANNs combinations have superior forecast performance than the individual VaR models. Chang, Lo, Chen and Huang (2006) develop an integrated forecasting model, which combines artificial neural network and grey system theory. The results show that the

integrated forecasting model outperforms the original artificial neural network forecasting model and obtained better forecasting performance in foreign exchange rates forecasting.

To compare the forecast performance of both the individual and "artificial" models, we use Kupiec and Christoffersen tests as the backtesting mechanism.

The LR of Kupiec test with chi-square distribution is formulized on the following equation (Kupiec, 1995).

$$LR = 2\left\{\log\left[f^x(1-f)^{T-x}\right] - \log\left[\alpha^x(1-\alpha)^{T-x}\right]\right\} \tag{24}$$

f is defined as the ratio of the number of observations exceeding Var(x) to the number of total observation (T) and pre-specified VaR level as α (Tang and Shieh, 2006).

Christoffersen test (Christoffersen, 1998) based on testing whether $\Pr(r_t < v_t) = p$ after conditioning on all information available at time t, on the other hand, focuses on the probability of failure rate (Sarma $et\ all$, 2001). The importance of testing conditional coverage arises with volatility clustering in financial time series. Christoffersen test might be superior in detecting fat-tail in the time series as compared to Kupiec test.

Christoffersen test can be applied as follows (Saltoglu, 2003). Define $p^\alpha = \Pr(y_t < VaR_t(\alpha))$ to test $H_0 : p^\alpha = \alpha$ against $H_1 : p^\alpha \neq \alpha$. Consider $\left\{1(y_t < VaR(\alpha)\right\}$ which has a binomial likelihood $L(p^\alpha) = (1 - p^\alpha)^{n_0}(p^\alpha)^{n_1}$.

where $n_0 = \sum_{t=R}^{T} 1(y_t > VaR_t(\alpha))$ and $n_1 = \sum_{t=R}^{T} 1(y_t < VaR_t(\alpha))$.

Under the null hypothesis, it becomes $L(\alpha) = (1-\alpha)^{n_0}\alpha^{n_1}$. Thus the likelihood ratio test statistics is in equation below.

$$LR = -2In(L(\alpha))/L(\hat{p})) \xrightarrow{\ d\ } \chi(1) \tag{25}$$

We estimate VaR with α =0.01 confidence interval and backtest VaR models with Kupiec Christoffersen out-of-sample forecasting test. We chose %99 confidence level in accordance to Basel II requirement.

4. Data and Empirical Results

Data

Istanbul Stock Exchange Rate (ISE-100 Index) is from Bloomberg. Our dataset covers 2412 daily observations from 01/10/1996 to 11/07/2006. We constituted the series in log-differenced level. Figure 2 shows ISE Index in log-differenced series. By performing Augmented Dickey–Fuller (Dickey and Fuller, 1981) test we found that ISE Index is stationary at log differenced level (as Augmented Dickey-Fuller test of I(1) with 0 lags is equal to -48.2929 {<%1}). The estimation process is run using 10 years of data (1996-2006) while the remaining 300 observations is used for historical simulation rolling,

500 observations is used for artificial neural network training and last 1342 observations is used for out-of-sample forecasting. Figure 3 shows density plot of ISE and normal distribution and as it is shown that distribution of data is not approximate normal distribution even with log-differenced transformation. As a result, extreme value distribution and fat-tailed distributions such as skewed student-t can improve forecasting.
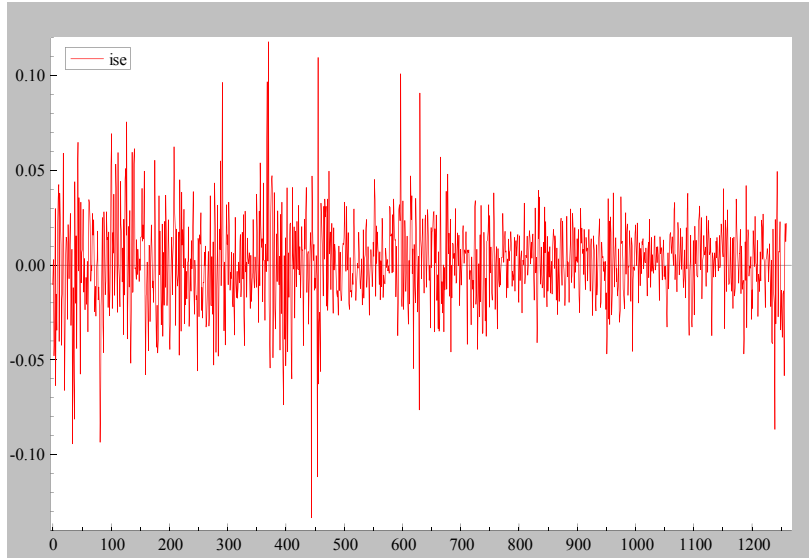

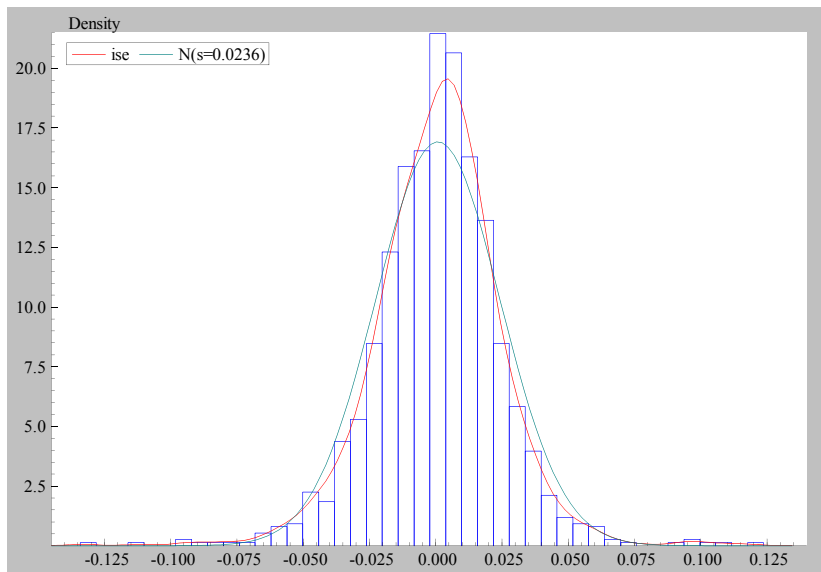**Fig. 2. ISE Log-differenced series**

Fig. 3. Density Plot of ISE and Normal Distibution

Empirical Results

In this subsection, we report estimation and Kupiec and Christoffersen tests results for Historical Simulation, Garch(1,1), Garch(1,1) with skewed student-t distribution, NM-GRJ(1,1,), EVT Hill and combined models detailed in Methodology section. We used Matlab and Ox programming language (see Doornik, 1999) and parameters are estimated using Quasi Maximum Likelihood technique (Bollerslev and Woolridge, 1992) and BFGS quasi-Newton method optimization algorithm used. Estimation of Asymmetric Normal Mixture Garch is performed with modified version of Alexander and Lazar(2006) codes and Garch models, historical simulation and EVT Hill models is carried out with Matlab.

11

Table 1 shows Garch(1,1), Garch(1,1) with skewed student-t and Normal Mixture GRJ(1,1) estimation results. α and β_1 parameters for all of the models statistically significant. For the skewed student-t distribution, the asymmetric parameters (ξ) are negative and statistically significant for Garch model. Thus show that the density distribution of ISE skewed to to left. ω, α, β_1 and normal mixture γ (Gamma) parameter statistically significant for Asymmetric Normal Mixture Model. These results show that Asymmetric Normal Mixture Garch model may improve forecasting.

Table 1. Estimation Results from GARCH(1,1) and NM-GRJ(1,1)

	Garch	Garch- Skew	NM-GRJ
ω	0.139**	0.181**	0.199841 **
	(4.52)	(3.62)	(3.732)
α	0.110**	0.115**	0.122624 **
	(12.26)	(7.49)	(7.542)
β_1	0.880**	0.871**	0.861064**
	(98.00)	(56.45)	(52.75)
v-*Student t*	-	-	-
ξ-*Skewness*	-	-0.056*	-
		(-2.05)	
v-*Skewness*	-	6.508**	-
		(7.62)	
γ-*Normal Mixture*			0.005237 **
			(2.825)
Volatility	0.0400219	0.0368966	0.0378326
LogLike	5245.17	5304.89	5309.1297
AIC	-4.375	-4.423	-4.42665

*, ** are %5 and %10 confidence level respectively.

Out of sample forecasting graphs of univariate and ANN combined models are shown in Figure 4 and ANN combined models are shown in Figure 5. As it is seen in Figure 5, Garch(1,1)-EVT Hill and HS-EVT Hill combinations do not capture tail loss where Garch(1,1) with Skewed t-EVT Hill and NMGarch(1,1)-EVT Hill captures tail loss.
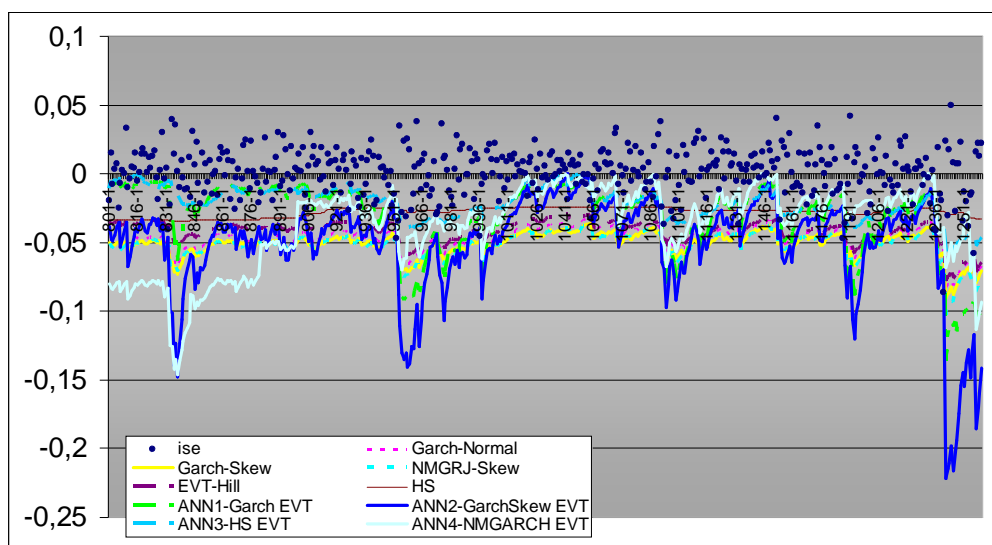

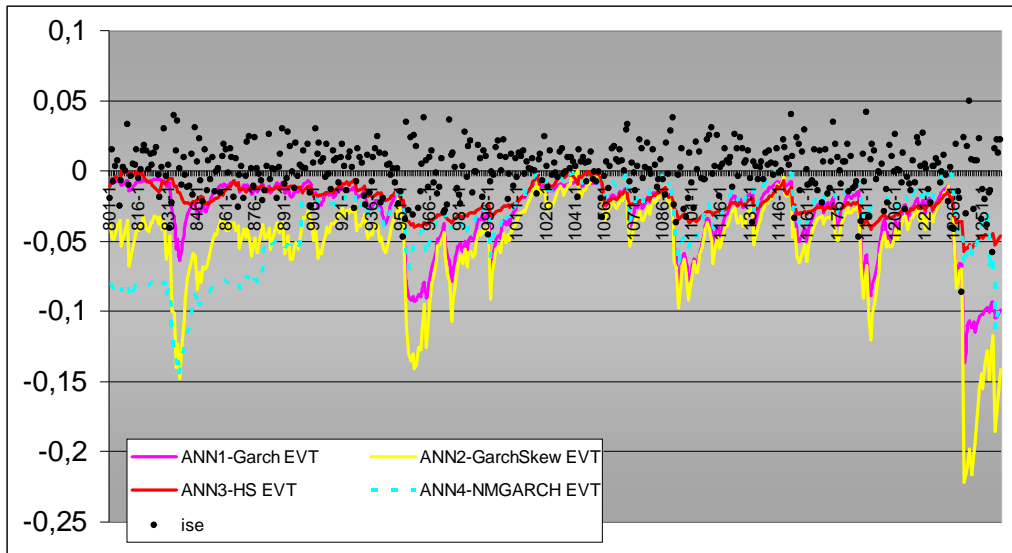**Fig. 4. Out of Sample Forecasting**

Fig. 5. Combined Forecasting Models with Artificial Neural Networks

The estimated results for five univariate and four combined ANN model of Kupiec test are collected in Table 2. The empirical results indicate that Garch(1,1) with normal distribution performs better based on Kupiec LR test including all ANN models. Garch(1,1) with skewed t and EVT hill ANN model perfoms better in ANN models. Thus indicate that ANN combinations do not improve forecasting based on Kupiec test.

Table 2. Kupiec Test- Out of Sample Forecasting

Model	Failure Rate
Garch	**0.9969**
GarchSkew	0.9978
HS	0.9630
EVT-Hill	0.9854
NMGARCH	0.9851
ANN1- Garch EVT	0.8696
ANN2- GarchSkew-EVT	**0.9630**
ANN3-HS EVT	0.8370
ANN4-NMGRJ EVT	0.8935

The estimated results for five univariate and four combined ANN model of Christoffersen test are collected in Table 3. The empirical results indicate that Garch(1,1) with skewed t-EVT Hill ANN combination model performs better based on Christoffersen LR test including univariate models. Thus shows shat fat-tailed models like skewed student-t and EVT combination improves forecasting.

Table 3. Christoffersen Test-Out of Sample Forecasting

Model	LR$_{UC}$
Garch	21,471
GarchSkew	24,821
HS	26,695
EVT-Hill	15,925
NMGARCH	15,104

13

ANN1- Garch EVT	10,401
ANN2- GarchSkew-EVT	**0,6305**
ANN3-HS EVT	5,7223
ANN4-NMGRJ EVT	15,380

5. Conclusion

Financial behaviours in the capital markets are so complex to be estimated with restricted or one-dimensional modeling techniques. Especially in case of emerging equity markets, high volatility in stock prices, structural breaks and regulations on taxing create an environment in which estimating stock returns are nearly impossible with linear and single models.

In computational finance, with the help of improvements in software, intelligent models have been applied to model the price behaviours at the capital markets. What is more, instead of linear or Gaussian models, nonlinear models with asymmetric normality assumptions are proposed to capture the fat tail in the return distribution.

In this research, we combine the EVT Hill with the parametric and non-parametric models by using genetic algorithm as a combination mechanism. In that context, firstly, historical simulation, GARCH, GARCH with skewed student-t distribution, asymmetric normal mixture GRJ-GARCH model and Extreme Value Theory Hill are used to estimate the VaR for the Istanbul Stock Exchange 100 Index. Then, the historical simulation and GARCH models are individually combined with EVT Hill by using genetic algorithm.

The estimated results for five univariate and four combined ANN model of Kupiec test show that GARCH (1,1) with normal distribution performs better than the alternative models including combined ones. On the other hand, the Kupiec test might not be easily detect the fat tails in the time series. Therefore, we apply Christoffersen test for the models. The empirical results indicate that GARCH (1,1) with skewed t-EVT Hill ANN combination model performs better based on Christoffersen LR test including univariate models. Thus shows shat fat-tailed models like skewed student-t and EVT combination improves forecasting.

The empirical findings prove that the advanced and combined models are successful on capturing the fat tails in the returns distribution in the emerging capital markets. The future research might concentrate on the alternative combined models, such as combination of wavelets and parametric models to capture the long-term memory effects in the stock prices.

References

Alexander, C. and Lazar, E., (2005) "*The Equity Index Skew, Market Crashes and Asymmetric Normal Mixture Garch,*" ISMA Center, Mimeo

Alexander, C. and Lazar, E., (2006) "*Normal Mixture GARCH(1,1):Applications to Exchange Rate Modeling,*" Journal of Applied Econometrics 21(3), pp.307-336.

Baillie, R. T. and Bollerslev, T., (1989) "*The Message in Daily Exchange Rates: A Conditional-Variance Tale,*" Journal of Business and Economic Statistics, 7, pp.297-305

Bates, D. S., (1991) "*The Crash of '87: Was It Expected? The Evidence from Options Markets,*" Journal of Finance, 46, pp.1009-1044

Bekaert, G., and Wu, G., (2000) "*Asymmetric Volatility and Risk Equity Markets,*" The Review of Financial Studies, 13(1), pp.1-42

Bollerslev, T., (1986) "*Generalized Autoregressive Conditional Heteroskedasticity,*" Journal of Econometrics, 31, pp.307–327.

Bollerslev, T. (1987) "*A Conditional Heteroskedasticity Time Series Model for Speculative Prices and Rates of Return,*" Review of Economic and Statistics. 69, pp.542-547

Bollerslev, T., Chou, R.Y. and Kroner, K.F., (1992) "*ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence,*" Journal of Economics and Statistics, 69, pp.542-547

Bollerslev, T. and Woolridge, J. M. (1992) "*Quasi-maximum Likelihood Estmation Inference in Dynamic Models with Time-varying Covariances,*" Econometric Theory, 11, pp.143-172

Chang, Lo, Chen and Huang (2006) "*Journal of Information & Optimization Sciences,*" 27 (2006), No. 3, pp. 615–628

Christoffersen, P.F. (1998) "*Evaluating Interval Forecasts,*" International Economic Review, 39, pp.841-862.

Christoffersen, P. F., and Jacobs, K., (2004) "*Which Garch Model for Option Valuation?,*" Management Science, 50, pp.1204-1221

Christoffersen, P. F., Heston, S. and Jacobs, K. (2004) "*Option Valuation with Conditional Skewness,*" Forthcoming in The Journal of Econometrics.

Dickey, D. A., Fuller, W.A. (1981) "*Likelihood ratio statistics for autoregressive time series with a unit root,*" Econometrica, 49, pp.1057–1072.

Doornik, J.A. (1999) "*An Object Oriented Programming Language,*" Timberlake Consultant, Third Ed.

Engle, F., (1982) "*Autoregressive Conditional Heteroscedasticity with Estimate of the Variance of United Kingdom Inflation,*" Econometrica, 50, pp.987-1007

Fernandez, Carmen, and Mark Stell. (1998) "*On Bayesian Modeling of fat tails and skewness,*" Journal of the American Statistical Association, 93, pp.359-371

Glosten, L. R., Jagahannathan, R., and Runkle, D. E., (1993) *"On the Relationship between the Expected Value and The Volatility of the Nominal Excess Return on Stocks,"* Journal of Finance, 48, pp.1779-1801

Hecht-Nielsen.R (1990) Neurocomputing, Wokingham, England: Addison-Wesley Publishing Company Ltd.

Hill, B. M. (1975) *"A simple general approach to inference about the tail of a distribution,"* Annals of Statistics, 3, 1163-1174.

Holland, J. (1965) *"Universal spaces: A basis for studies of adaptation,"* In Automata Theory. Caianiello, E. R. (ed.) Academic Press. 218-30.

Hsieh, D. A. (1989) *"Modeling Heteroskedasticity in Daily Foreign Exchange Rates,"* Journal of Business and Economic Statistics, 7, pp.307-317

Kupiec, P. H. (1995) *"Techniques for Verifying the Accuracy of Risk Measurement Models,"* Journal of Derivatives, winter, pp. 73-84

Liu (2005) *"Value-at-Risk Model Combination Using Artificial Neural Networks,"* Emory University Working Paper Series.

Maasoumi, E., A. Khotanzad, and A. Abaye (1994) *"Artificial Neural Networks for Some Macroeconomic Series : A First Report,"* Econometric Reviews, 13, No. 1.

McNeil, A. J., (1996) "Estimating The Tails of Loss Severity Distributions Using Extreme Value Theory, Mimeo. ETH Zentrum, Zürich.

McNeil A. J. and Frey, R. (2000) *"Estimation of tail-related risk measures for Heteroscedastic Financial time series: An extreme value approach,"* Journal of Empirical Finance, 7, 271-300.

Neftci, S., (2000) *"Value at Risk Calculations, Extreme Events, and Tail Estimation,"* The Journal of Derivatives, Spring 2000.

Ozun, A., (2006) Theoretical Importance of Artificial Neural Networks For The Efficiency of Financial Markets, Proceedings of 5th International Finance Symposium: Integration in the Financial Markets, Vienna University&Marmara University with Cooperation of Istanbul Stock Exchange, 25-26 May, 2006, Istanbul

Palit, AK and Popovic, D. (2000) *"Nonlinear Combination of Forecasts Using Artificial Neural Networks, Fuzzy Logic and Neuro-Fuzzy Approach,"* FUZZ-IEEE, Vol. 2, pp. 566-571, 2000.

Pagan, A., (1996) *"The Econometrics of Financial Markets,"* Journal of Empirical Finance, 3, pp.15-102.

Palm, F., (1996) *"Garch Models of Volatility", in Handbook of Statistics*, ed. By G.Maddala, and C.Rao, pp.209-240, Elsevier Science, Amsterdam.

Palm, F. and Vlaar, P. JG., (1997) *"Simple Diagnostics Procedures for Modeling Financial Time Series,"* Allgemeines Statistisches Archiv, 81, pp.85-101

Peters, JP., (2001) Estimating and Forecasting Volatility of Stock Indices Using Asymmetric Garch Models and (Skewed) Student-t Densities, Mimeo, Ecole d'Admin. des Affaires, Unv.of Li`ege.

Rumelhart, D.E. & McClelland, J.L. (1986) "*PDP Models and General Issues in Cognitive Science*". In D.E. Rumelhart & J.L. McClelland (Eds.), Parallel Distributed Processing. Vol. 1. Cambridge, MA: MIT Press/Bradford Books.

Saltoğlu, B. (2003) A High Frequency Analysis of Financial Risk and Crisis: An Empirical Study on Turkish Financial Market, Yaylım Publishing, Istanbul

Sarma, M., Thomas, S. and Shah, A. (2001) Selection of Value-at-Risk Models, Mimeo

Shanming Shi, Li D Xu, Bao Liu. (1996) "*Application of artificical neural networks to the nonlinear combinmation of forecasts,*" Expert Systems,1996,13(3):195-201

Tang, T.L. and Shieh, S. J., (2006) "*Long-Memory in Stock Index Futures Markets: A Value-at-Risk Approach,*" Phsica A, Vol.366, pp.437-448

White, H. (1994) *"Neural Networks,"* Econometric Reviews, Vol.13, No.1.

Wu, G., (2001) "T*he Determinants of Asymmetric Volatility,*" The Review of Financial Studies. 14(3), pp.837-859