Forecasting Volatility of Dhaka Stock Exchange: Linear Vs Non-linear models

Masudul Islam [#1], Lasker Ershad Ali [*2], Nahida Afroz [#3]

[#1] *Statistics Discipline, Khulna University, Khulna-9208, Bangladesh*

[2] *Mathematics Discipline, Khulna University, Khulna-9208, Bangladesh*

[3] *Statistics Department, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh*

[1] colinmasudul@gmail.com

[2] liponmath@yahoo.com

[3] nahidafrz@yahoo.com

***Abstract*— Prior information about a financial market is very essential for investor to invest money on parches share from the stock market which can strengthen the economy. The study examines the relative ability of various models to forecast daily stock indexes future volatility. The forecasting models that employed from simple to relatively complex ARCH-class models. It is found that among linear models of stock indexes volatility, the moving average model ranks first using root mean square error, mean absolute percent error, Theil-U and Linex loss function criteria. We also examine five nonlinear models. These models are ARCH, GARCH, EGARCH, TGARCH and restricted GARCH models. We find that nonlinear models failed to dominate linear models utilizing different error measurement criteria and moving average model appears to be the best. Then we forecast the next two months future stock index price volatility by the best (moving average) model.**

***Keywords*— Volatility, Stock index future volatility, EGARCH, TGARCH, Restricted GARCH.**

I. Introduction

Volatility in stock market has been one of the most analyzed issues in the past decades. The term volatility is a key element for pricing financial instruments such as options, a measure of trade off between return and risk for allocating assets and is closely related to portfolio return fractiles, option pricing and risk management. Financial market volatility also has a wider impact on financial regulation, monetary policy and macro economy. The practical importance of volatility modeling and forecasting in many finance applications means that the success or failure of volatility models will depend on the characteristics of empirical data that they try to capture and predict. A high volatility in a stock market creates a bad impact for the country's economy. For this reason the volatility is an important issue that concerns government policy making, market analysis, corporate and financial managers. To make the market be efficient and make reliable to the investor, many businessmen try to forecast the volatility because the stock market is one of the sources for the industry to raise money.

In the empirical finance literature, many linear models are used to describe the stock return volatility. Poterba and Summers (1986) specify a stationary AR (1) process for the volatility of the S&P 500 Index. French, Schwert and Stambaugh (1987) use a non-stationary ARIMA (0, 1, 3) model to describe the volatility of the S&P 500 Index. Schwert (1990) and Schwert and Seguin (1990) use a linear AR (12) as an approximation for monthly stock return volatility. The extensive use of such linear models is not surprising since they provide good first order approximation to many processes and the statistical theory is well developed for linear Gaussian models. However, certain features of a volatility series cannot be described by linear time series models. For example, empirical evidence shows that stock returns tend to exhibit clusters of outliers, implying that large variance tends to be followed by another large variance. Such limitations of linear models have motivated many researchers to consider nonlinear alternatives. The most commonly used nonlinear time-series models in the financial literature are the autoregressive conditional heteroscedastic (ARCH) model of Engle (1982), the generalized ARCH (GARCH) model of Bollerslev (1986), the exponential GARCH (EGARCH) model of Nelson(1991) and Threshold ARCH(TARCH) of Zakoian (1990) and Glosten, Jaganathan, and Runkle (1993). These ARCH-class models have been found to be useful in capturing certain nonlinear features of financial time-series such as heavy-tailed distributions and clusters of outliers. Bera, Bubnys and Park (1993) investigate the validity of the conventional OLS model to estimate optimal hedge ratio using futures contracts. Another complex class of nonlinear models is called component ARCH model (restricted GARCH (2, 2)) suggested by Bollerslev, Engle, and Nelson (1994). They conclude that the component ARCH model is a suitable tool for describing short run movement and long run levels of volatility found in financial time series.

Bangladesh is a developing country where the Stock Market is an economic indicator of the country. But the Stock Market of Bangladesh is not an efficient market. So, making the market efficient and reducing the uncertainty that the investor is invest, the volatility forecast is necessary step for the government and policy makers. The purpose of this paper is to examine the relative ability of various models to forecast daily stock index future volatility on the basis of error measurement and find the best forecasting model which is suitable for Bangladesh.

The paper is organized as follows. Section 2 contains some methods that we used to analyse the data. Section 3 describes empirical analysis of selecting forecasting model. We conclude in Section 4.

II. Data and Methodology

Data

Daily closing prices data of DSE-20 index between January 2002 and November 2011 is obtained from Dhaka Stock Exchange. Since Most trading activities take place in near day contract, only near-day contract data are examined. A continuous sequence of 2739 observations of closing price data is gathered over the last ten-year period. The logarithm of price relatives multiplied by 100 is used to calculate price change. i.e., $r_t = 100 \times \ln[P_t / P_{t-1}]$, where the (unconditional) distribution of r_t is leptokurtic and asymmetric(in some cases),correlation between returns is absent or very weak and correlations between the magnitudes of returns on nearby days are positive and statistically significant.

Assessing the distributional properties of daily stock index price change, various descriptive statistics are reported in Table 1 including: mean, standard deviation, skewness, kurtosis and the Kolmogrove-Smirnov (K-S) D statistics normality test. The null hypothesis of normality is rejected at the 1% level using K-S D statistics and the deviation from normality may be gleaned by the Kernel Density graph as well as the sample skewness and kurtosis measures. While skewness is relatively small and kurtosis is very large for both DSE-20Index. Following Poon, Ser-Huang (2005), we estimate the volatility of daily returns by the following equation

$$\hat{\sigma} = \sqrt{\frac{1}{T}\sum_{T=1}^{T}(r_t - \mu)^2} \qquad (1)$$

where, r_t is the return on day *t* and μ is the average return over the *T*-day period. Since, variance is simply the square of standard deviation; it makes no difference whichever measure we use when we compare the volatility of two assets. Since volatility is latent variable, many researchers have resorted to using daily square returns, calculated from daily closing prices, to proxy daily volatility. Lopez (2001) shows that $\in_t^2$ is an unbiased but extremely imprecise estimator of σ_t^2 due to its asymmetric distribution. Let,

$$Y_t = \mu + \in_t \qquad \in_t = \sigma_t z_t$$

and $z_t \sim N(0,1)$ .Then $E[\in_t^2 | \varphi_{t-1}] = \sigma_t^2 E[z_t | \varphi_{t-1}] = \sigma_t^2$.

Since, $z_t^2 \sim \chi^2_{(1)}$.

Methodology

The focus of this paper is on the forecasting accuracy of daily stock price volatility from various statistical models. The basic methodology involves the estimation of various models for an initial period and finds the best model on the basis of error measurement criteria. Then test the best model for the later data (in-sample forecasts) and finally calculate next two month forecasted data, thus for forming out-of –sample forecasts. The linear models employed are: (1) a random walk model, (2) Historical model, (3) an autoregressive model, (4) a moving average model, (5) an exponential smoothing model, (6) a simple regression model. The nonlinear models utilized here are ARCH, GARCH(1,1),EGARCH(1,1), TGARCH(1,1) and Restricted GARCH(2,2)Models.

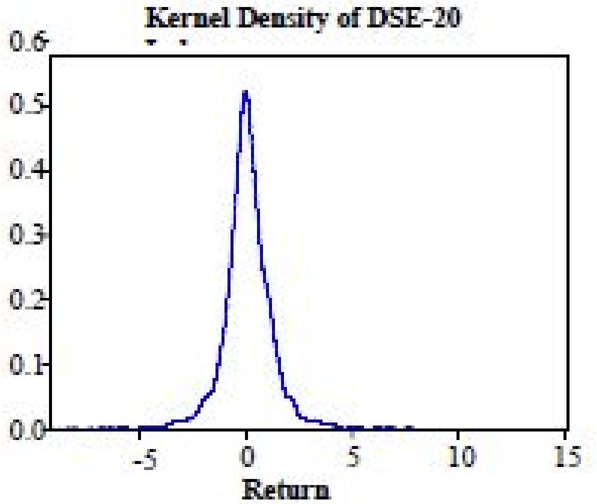

Fig 1. Shape of real distribution of daily returns of DSE-20 Index.

TABLE I
Descriptive statistics on stock index between January 1, 2002 and November 30, 2011.

Statistics	DSE-20 Index
Sample size	2739
Mean	0.0646
Median	0.0500
Standard. deviation	1.6380
Skewness	0.6443
Kurtosis	78.761
Kolmogrove-Smirnov Test	0.099

*Indicates statistical significance at the 0.01 level.

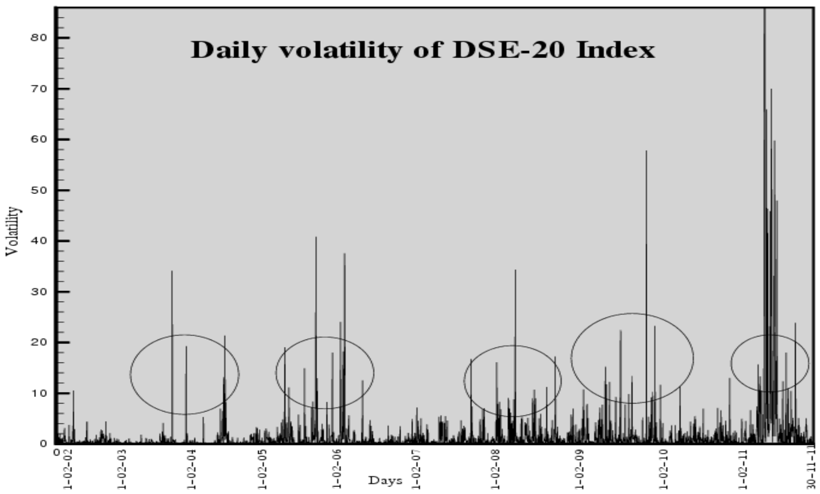

Fig 2. DSE-20 index volatility for the period from January 2002 to November'11.

Table II
Descriptive statistics on stock index volatility series of near day contrasts between January 1, 2002 andNovember30,2011.

Statistics	DSE-20 Index
Sample size	2739
Mean	1.646
Median	0.309
Standard. deviation	6.375
Skewness	18.176
Kurtosis	526.96
Kolmogrove-Smirnov Test	0.358

*Indicates statistical significance at the 0.01 level.

Linear models
Random walk model
According to Random walk model, the best forecast of today's volatility depends on yesterday's observed volatility.

$$\hat{\sigma}_t^2 = \sigma_{t-1}^2 + \varepsilon_t \tag{2}$$

where, σ_t^2 is the daily volatility measure defined in equation (1).

Historical average model
Under historical average model, the conditional expectation of volatility is assumed to be constant and the optimal forecast of future volatility would be the historical average.

$$\hat{\sigma}_t^2 = \frac{\sum_{t=2}^{T} \sigma_{t-1}^2}{T} + \varepsilon_t \tag{3}$$

where, σ_t^2 is the daily volatility measure defined in equation (1).

Moving average model
For a moving average model of order q the forecast of volatility at time point t is the average of the recent q volatilities; that is,

$$\hat{\sigma}_t = \frac{\sigma_{t-q} + \sigma_{t-q+1} + \quad \ldots \quad + \sigma_{t-1}}{q} + \varepsilon_t \tag{4}$$

Exponential smoothing model
Exponential smoothing is a simple method of adaptive forecasting. Single exponential smoothing forecast is given by,

$$\hat{\sigma}_t = (1-\alpha)\sigma_{t-1} + \alpha\hat{\sigma}_{t-1} + \varepsilon_t \tag{5}$$

where, $0 < \alpha < 1$ is the smoothing factor. By repeated substitution, the recursion can be rewritten as

$$\hat{\sigma}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j \sigma_{t-j} + (1-\alpha)^{t-1} + \varepsilon_t \tag{6}$$

Simple regression model
This is a one-step ahead forecast based on the simple linear regression of the volatility at period t on the volatility at period *t+1*. The expression is given by,

$$\hat{\sigma}_{t+1}^2 = \beta_1 + \beta_2\sigma_t^2 + \varepsilon_{t+1} \tag{7}$$

Auto regressive model
The first-order autoregressive model is define as,

$$\hat{\sigma}_t = \lambda\sigma_{t-1} + \varepsilon_t \tag{8}$$

The general form of AR model of order p is

$$\sigma_t = \sum_{i=1}^{p} \lambda_i \sigma_{t-i} + \varepsilon_t \tag{9}$$

Nonlinear models
ARCH model
The ARCH model, first introduced by Engle (1982), has been extended by many researchers and extensively surveyed in Bera and Higgins (1993), Chou and Kroner (1992), Bollerslev and Die-bold and Lopez (1995). In contrast to the historical volatility models described , ARCH models do not make use of the past standard deviations, but formulate conditional variance, h_t ,of asset returns via maximum likelihood procedures. We follow the ARCH literature here by writing $\sigma_t^2 = h_t$ .To illustrate this, first write returns, r_t , as

$$r_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t}\, z_t \tag{10}$$

where, $z_t \sim N(0,1)$ is a white noise. The process z_t is scaled by h_t (the conditional variance) which in turn as a function of
past squared residual returns. In the ARCH (q) process proposed by Engle (1982),

$$h_t = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \tag{11}$$

with $\omega>0$ and $\alpha_j \geq 0$ to ensure h_t is strictly positive variance. Typically, q is of high order because of the phenomenon of volatility persistence in financial markets.

GARCH (1, 1) model
In the standard GARCH (1, 1) specification:

$$y_t = x_t\gamma + \varepsilon_t \tag{12}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{13}$$

the mean equation given in (12) is written as a function of exogenous variables with an error term. The (13) in GARCH (1, 1) refers to the presence of a first-order GARCH term (the first term in parentheses) and a first-order ARCH term (the second term in parentheses).

TARCH model
Threshold ARCH was introduced independently by Zakoian(1990) and Glosten, Jaganathan and Runkle (1993).The specification for the conditional variance is,

$$\sigma_t^2 = w + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 d_{t-1} + \beta\sigma_{t-1}^2 \tag{14}$$

where, d_t =1 if ε_t <0, and d_t =0 otherwise. In this model, good news (ε_t <0), and bad news(ε_t >0), have differential effects on the conditional variance—good news has an impact of α , while bad news has an impact of $\alpha+\gamma$ If $\lambda > 0$,the leverage effect exists. For higher order specifications of the TARCH model,

$$\sigma_t^2 = w + \sum_{i=1}^{q} \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 d_{t-1} + \sum_{j=1}^{p} \beta\sigma_{t-1}^2 \tag{15}$$

EGARCH model
The EGARCH or Exponential GARCH model was proposed by Nelson (1991). The specification for the conditional variance is

$$\log(\sigma_t^2) = w + \beta\log(\sigma_{t-1}^2) + \alpha\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \tag{16}$$

where, the left-hand side is the log of the conditional variance. This implies that the leverage effect is exponential, rather t quadratic and that forecasts of the conditional variance are guaranteed to be nonnegative. The presence of leverage effects can be tested by the hypothesis that, $\gamma < 0$. For higher order specifications of EGARCH models,

$$\log(\sigma_t^2) = w + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2) + \sum_{i=1}^{q} (\alpha_i \left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| + \gamma_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}}) \quad (17)$$

Restricted GARCH (2, 2) or Component ARCH model

The conditional variance in the GARCH (1, 1) model,

$$\sigma_t^2 = \overline{w} + \alpha(\varepsilon_{t-1}^2 - \overline{w}) + \beta(\sigma_{t-1}^2 - \overline{w}) \quad (18)$$

shows, mean reversion to $\overline{w}$ which is a constant forall time. By contrast, the component model allows mean reversion to a varying level q_t , model as,

$$\sigma_t^2 - q_t = \alpha(\varepsilon_{t-1}^2 - \overline{w}) + \beta(\sigma_{t-1}^2 - \overline{w}) \quad (19)$$

$$q_t = w + \rho(q_{t-1} - w) + \phi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \quad (20)$$

Here σ_t is still the volatility, while q_t is the time varying long run volatility. The first equation describes the transitory component, $\sigma_t^2 - q_t$, which converges to zero with powers of $\alpha + \beta$. The second equation describes the long run component

q_t , which converges to w with powers of ρ . Combining the transitory and permanent equations asfollows,

$$\sigma_t^2 = (1-\alpha-\beta)(1-\rho)w + (\alpha+\phi)\varepsilon_{t-1}^2 - (\alpha\rho + (\alpha+\beta)\phi)\varepsilon_{t-2}^2 + (\beta-\phi)\sigma_{t-1}^2 - (\beta\rho - (\alpha+\beta)\phi)\sigma_{t-2}^2 \quad (21)$$

which, shows that the component model is a (nonlinear) restricted GARCH(2,2) model.

III. Empirical Results

To evaluate the performance of the linear and non-linear models in describing stock index futures volatility, we compare their out-of –sample forecasts with our benchmark model (1). The post-sample forecast comparisons are carried out as follows. First, we reserve the last 60 observations for forecast comparison. Secondly, all the models used in forecasting are estimated using the first 2664 observations. Such a scheme provides 60 one-step ahead forecasts. The objective is to evaluate forecasting capability of different models during the lowland high volatility periods on the basis of error measurement criteria. We summarize the forecast performance by considering the root mean squared error(RMSE), mean absolute percentage error (MAPE), Theil-U and LINEX loss function which are defined as follows:

$$RMSE: \sqrt{\frac{1}{N}\sum_{t=1}^{N} \varepsilon_t^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} (\hat{\sigma}_t - \sigma_t)^2} \quad (22)$$

$$MAPE: \frac{1}{N}\sum_{t=1}^{N} \frac{|\varepsilon_t|}{\sigma_t} = \frac{1}{N}\sum_{t=1}^{N} \frac{|\hat{\sigma}_t - \sigma_t|}{\sigma_t} \quad (23)$$

$$Theil\text{-}U = \frac{\sum_{t=1}^{N} (\hat{\sigma}_t - \sigma_t)^2}{\sum_{t=1}^{N} (\hat{\sigma}_t^{BM} - \sigma_t)^2} \quad (24)$$

where, $\hat{\sigma}_t^{BM}$ is the benchmark forecast.

$$LINEX = \frac{1}{N}\sum_{t=1}^{N} \left[\exp\{-a(\hat{\sigma}_t - \sigma_t)\} + a(\hat{\sigma}_t - \sigma_t) - 1\right] \quad (25)$$

In table 3, we see root mean square, mean absolute percentage error, Linex10 are smallest for moving average model, Theil-U is smallest for ar(1) but 2nd best is moving average, and Linex20 are smaller for random walk model but 2nd best is moving average .So, an examination of Table 3 reveals that within the linear models the moving average model dominates all of the models using RMSE, MAPE of all the models, linear and nonlinear, the moving average model to all the models followed closely by random walk model and restricted GARCH. Thus, all error measurement criteria clearly identify the linear class models and moving average model as superior to all linear and non-linear models.

Table III
Error measurement of linear and nonlinear models for DSE-20 Index.

	RW	HM	S R	MA	ES	Ar(1)
Root mean square	6.969	6.350	3.823	3.485	5.344	5.827
Mean absolute persent error	38.986	41.294	24.546	14.95	31.999	34.973
Theil-U	0.018	1207	0.229	0.093	21.816	0.018
Linex10	0.038	0.113	0.141	0.032	0.233	0.179
Linex20	0.010	0.2434	0.300	0.078	0.483	0.375
	ARCH(1)	GARCH(1,1)	GARCH(2,2)	TGARCH	EGARCH	Restrictd GARCH
Root mean square	6.116	5.992	5.982	5.902	6.129	5.943
Mean absolute persent error	39.489	37.555	37.758	36.71	38.619	37.594
Theil-U	1591	3932	2384	1051	1459	1635
Linex10	0.357	0.291	0.156	.318	0.493	0.152
Linex20	0.131	0.599	0.331	0.652	1.002	0.321

Then we forecast last 60 observations and compare with actual values. We find that the forecasted value almost closer to the actual value which indicates the moving average model forecast well for DSE-20.

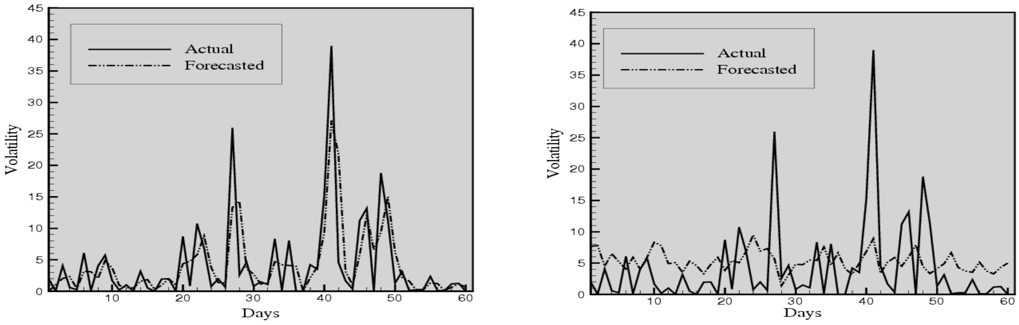

Fig 3. Comparison graph of Moving Average Model & TGARCH Model

IV. Conclusions

The purpose of this paper is to examine the relative ability of various models to forecast daily stock index futures volatility. Understanding and modeling stock volatility is important since volatility forecasts have many practical applications. Invest decisions and asset pricing models depend heavily on the assessment of future returns and risk of various assets. The expected volatility of a security return also plays an important role in the option pricing theory. The six linear models considered here are: (1) random walk, (2) historical average, (3) moving average, (4) simple regression, (5) exponential smoothing, (6) autoregressive model. It is found that among linear models of stock index volatility, the moving average model ranks first using RMSE, MAPE, Theil-U, Linex loss function error criteria. We also examine five nonlinear models. These models are ARCH, GARCH, EGARCH, TGARCH and Restricted GARCH model. We find that linear model dominates nonlinear models utilizing different error statistics and moving average appears to be best model for forecasting stock index future volatility followed closely by random walk model.

References

[1] Akgiray V. (1989). Conditional heteroscedasticity in time series of stock returns: evidence and forecasts, *Journal of Business* 62(1),p.55-80.

[2] Baillie, R.T.,Bollerslev,T.(1989). The Measure in Daily Exchange Rates: A Conditional Variance Tale, *Journal of Business & Economic Statistics*, Vol.7, p. 60-68.

[3] Bautista, C. C. (1998). Stock Market Volatility in the Philippines, College of Business, Administration, University of the Philipines, Diliman, Quezon City, Philippines.

[4] Bera, A., Bubnys, E. & Park, H. (1993). ARCH Effects and Efficient Estimation of Hedge Ratios for Stock Index Futures, *Advances in Futures and Options Research*, 6, p.313-328.

[5] Bollerslev, T, Engle. Nelson, D.B. (1994). ARCH Models,*Handbook of Econometrics*, Vol. IV, Elsevier Science B.V.

[6] Brokwell, P. J. & Davis, R. A. (1996). *Introduction to Time Series and Forecasting*, Springer, New York.

[7] Chu, C.C. & Bubnys, E. L.(1990). A Likelihood Ratio Test of Price Volatilities: City of London, *Scottish Journal of Political Economy*, 36.1, p.90-95.

[8] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, Vol. 50, p.987–1007.

[9] French K.R., Schwert W.S. and Stambaugh R.F. (1987). Expected stock returns and volatility, *Journal of Financial Economics* 19, p. 3-29.

[10] Glosten, L.R., Jaganathan R. & Runkle D.E.(1993). On the relation between the expected value and the volatility of the nominal excess returns on stocks, *Journal of Finance* 48, p.1779-1801.

[11] Gujarati, D. N.(2003). *Basic Econometrics*, 4th ed, McGraw-Hill.

[12] Hill J., Jain A. &Wood R., A. Jr. (1988). Insurance: Volatility Risk and Futures Mispricing, *Journal of Portfolio Management* 14/2: p. 23-29.

[13] Nelson, D. B., 1991. Conditional Heteroscedasticity in asset Returns: A new Approach, *Econometrica* 59, p. 347-370.

[14] Poon, Ser-Huang (2005). *A practical Guide to Forecasting Financial Market Volatility*, John Wiley & Sons, Ltd.

[15] Poterba, J. M. & L. H. Summers (1986). The Persistence of volatility and Stock Market Fluctuations, *American Economic Review* 76, p. 1143-1151.

[16] Schwert, G. W. & P. J. Seguin (1990). Heteroscedasticity in Stock Returns, *Journal of Finance*45, p.1129-1155.