3. ASSESSING VOLATILITY FORECASTING MODELS: WHY GARCH MODELS TAKE THE LEAD[1]

Marius MATEI*

Abstract

The paper provides a critical assessment of the main forecasting techniques and an evaluation of the superiority of the more advanced and complex models. Ultimately, its scope is to offer support for the rationale behind of an idea: GARCH is the most appropriate model to use when one has to evaluate the volatility of the returns of groups of stocks with large amounts (thousands) of observations. The appropriateness of the model is seen through a unidirectional perspective of the quality of volatility forecast provided by GARCH when compared to any other alternative model, without considering any cost component.

Keywords: volatility, GARCH, forecast, correlation, risk, heteroskedasticity
JEL Classification: C3, C53, D81

1. Introduction and the scope of the paper

Although traditional research in financial economics has been concentrated on the mean of stock market returns, the more recent developments in international stock markets have increased the interest for practitioners, regulators and researchers towards the volatility of such returns. The number of crashes and the size of their effects have forced all to look more carefully to the level and stationarity of volatility in time, researchers shifting their attention towards development and then improvement of econometric models able to produce accurate forecasts of such swings in returns' volatility.

The heteroskedastic[2] models developed for such purpose present particular importance due to the extended concern in the both academic and applied literature

---

[1] The paper is part of the PhD thesis " Risk analysis in the evaluation of the international investment opportunities. Advances in modeling and forecasting volatility for risk assessment purposes", coordinator Prof. Mircea Ciumara, NIER.

* Ph.D. Student at ESADE Business School, Department of Finance, Barcelona and at National Institute of Economic Research, Romanian Academy, Bucharest, Contact: marius.matei76@gmail.com, marius.matei@esade.edu.

[2] A sequence or a vector of random variables is heteroskedastic if the random variables have different variances.

for volatility measuring. Volatility represents the conditional standard deviation of the underlying asset return. It has many applications in the financial domain, among which there are the calculation of the value at risk of a financial position in risk management and asset allocation under the mean-variance framework.

Volatility modeling improves the efficiency in parameter estimation and the accuracy in interval forecast. Finally, volatility index can be a useful financial instrument in investment decision. VIX volatility index calculated by the Chicago Board of Option Exchange started to trade in futures beginning March 2006.

Some of the most important univariate volatility models are the autoregressive conditional heteroskedastic (ARCH) model compiled by Engle (1982), generalized ARCH (GARCH) model compiled by Bollerslev (1986), the exponential GARCH (EGARCH) model of Nelson (1991), the conditional heteroskedastic autoregressive moving average (CHARMA) model obtained by Tsay (1987), the random coefficient autoregressive (RCA) model of Nicholls and Quinn (1982), and the stochastic volatility (SV) models compiled by Melino and Turnbull (1990), Taylor (1994), Harvey, Ruiz and Shephard (1994), and Jacquier, Polson, and Rossi (1994).

Each model has its own strengths and weaknesses and having at hand such a large number of models, all designed to serve to the same scope, it is important to correctly distinguish between various models in order to find the one which provides the most accurate predictions.

The paper offers a critical assessment of the main forecasting techniques and to evaluate the superiority of the more advanced and complex models. Ultimately, its scope is to offer support for the rationale behind of an idea: GARCH is the most appropriate model to use when one has to evaluate the volatility of the returns of groups of stocks with large amounts (thousands) of observations. The appropriateness of the model is seen through a unidirectional perspective of the quality of volatility forecast provided by GARCH when compared to any other alternative model. In this context, the quality of the results is seen as the chosen model's ability to comprehend the relationship between the exogenous variables and the endogenous ones, by taking into account the autocorrelations and interaction effects that may exist within the data[3]. However, the superiority of GARCH is debated only on a theoretical ground, by involving more types of arguments. The first one would be the pure theoretical description of the improvements each refining of the forecasting volatility models brings, thus more recent models of GARCH showing improvements that each refining has targeted to. Each newer version was aimed to solve inefficiencies of previous models, and thus we assert that more recent models come with an advance in performance than their previous versions. Then, the second type of argument is a careful review of the most representative papers which attempted to benchmark such models. The choice for this argumentation is based on an extensive review of the literature written on the topic and on the fact that such literature, although not having arrived to a common conclusion regarding the superiority of a certain pattern of models, it contributed to creating a ´common-sense´ belief that envisages superiority of GARCH class of models. For this purpose, I have

[3] The quality factor measures preciseness by comparing the forecast with the real (historical) values. A comment on the methodology used for that will follow.

carefully reviewed the 50 most important papers published before June 2009, papers that had as main scope revealing of ´best´ or ´worst´ volatility forecasting models in different empirical contexts. The conclusion came after counting the papers that came with conclusions as regards ARCH/GARCH class of models, by placing them either as ´best´ or ´worst´, but without mentioning derivations of ARCH/GARCH in both categories. Therefore, papers that mentioned a model of ARCH/GARCH class as belonging to one of the two categories ('best' or 'worst') and at the same time another ARCH/GARCH model in the other category have been excluded. The result was a number of 19 papers which concluded that a derivation of ARCH or GARCH performed better against other tested models of other types (historical volatility models, implied standard deviation models, stochastic volatility models, etc.) while only 6 papers revealed that a model in the same category (ARCH/GARCH) was found to be inferior.

My third point of argumentation in the favor of GARCH takes into consideration one observation belonging to Andersen and Bollerslev (Andersen and Bollerslev, 1998a, Andersen *et al*., 1999) who revealed the latent (or inherently unobserved, stochastically evolving through time) character of volatility. Stock volatility consists of intra-day volatility and variation between days. Unlike price, which is a flow variable and can be measured instantaneously, volatility is a stock variable and therefore has to be measured over a period. This has been constantly a problem for econometricians as volatility is not observable and precisely measured, but rather estimated. Its unobservability makes difficult the forecasting performance assessment of conditional heteroskedastic models. The latent character of volatility transforms the volatility estimation and forecasting problem into a filtering problem in which the "true" volatility cannot be determined exactly, but only extracted at some degree of error. This might raise problems as the volatility given by the models must be compared with the "true" underlying volatility. The errors then can be an effect of the model that makes the forecasts or of how the true volatility is estimated. The previously mentioned papers brought a new point of understating possible sources of such many conflicting findings as regards model performance ranking. They said that the failure of GARCH-class of models to provide good forecasts was not a failure of the GARCH model itself, but rather a failure to specify correctly the true volatility measure against which the forecasting performance was measured. They sustain that the standard way of using *ex post* daily squared returns as the measure of "true" volatility for daily forecasts is flawed as such measure comprises a large and noisy independent zero mean constant variance error term which is unrelated to the actual volatility. Andersen and Bollerslev suggest that cumulative squared-returns from intra-day data be used as an alternative way to express such "true" volatility. Such measure, called "integrated volatility", offers the opportunity for a more meaningful and accurate volatility forecast evaluation. This represents a step forward in forecasting as it indicates the necessity of using high frequency data in empirical estimations.

According to this point of view, it seems that the failure of the GARCH models in the previous empirical tests is not a failure of the models themselves, but rather a methodology error of incorrectly estimating volatility with daily or monthly data. Since GARCH models are more sensitive to the type of the frequency of the data used than others, it is intuitive to think that using low frequency data would affect more their performance than other models (especially than simpler ones). This point of view adds to the previous ones and thus concludes that GARCH models might not perform poorly as in the tests made with low frequency data, if using higher frequency data.

All four pieces of argumentation are given in order to describe the sources of ´common-sense´ beliefs as regards higher superiority of GARCH. However, a final paper that would answer in a thorough and final manner to all discrepancies of all pieces of literature that came with contradictory evidence on the superiority of different classes of models has not been written yet. All that can be written in this stage is to bring together pieces of evidence as regards this topic justifying a certain type of argument and to compare it with the ´common-sense´ belief as regards performance advancement. This is the aim of this paper, the conclusion of which is in line with that ´common sense´: GARCH is superior in forecasting ability to other models.

To describe adequately the mixed findings in the literature, I consider that a review of this debate with references to both sides of argumentation would prove useful. However, I will focus on one paper (Brailsford and Faff, 1985) and the reason why its argumentation makes the best conclusions. However, although they use daily data, their methodology considers all studies written before, but is also straightforward and little open to flaws, reason for which I consider that their conclusions that point to higher accuracy of GARCH models correctly stand out. It is then intuitive to think that if tested with higher frequency data but keeping the same time periods and methodology, GARCH models would further gain in accuracy than they did in Brailsford and Faff's (1985) paper.

The role of the empirical exercise to follow in this study is not to test more models in which GARCH would eventually prove to be superior, but rather to show how GARCH effectively works and to evidentiate empirically how it solves the autocorrelation problem in long enough time series. Therefore, the reader will be able to understand the model's general qualities on the basis of which the theoretical argumentation given above with regard to its performance is based.

For the type of factors that concerns the quality of the results, I will assert the important step ahead that GARCH models make against the basic ones. The discussion will start from ARMA models, and will be built on an approach that will justify why each refinement (that most of the time incorporates a generalization) of one model represents an improvement as compared to the previous one. Thus, I will reason why ARCH is better than ARMA and why GARCH is better than ARCH. My conclusion will be that, based on the testing with complex data manipulation, GARCH is the best model.

The paper is structured as it follows: Section 1 is an introduction to the issue and presents the scope of the research enclosed. Section 2 presents a benchmarking of the main volatility forecasting models, stressing the improvements new models bring as compared to the previous ones. Section 3 presents the empirical exercise on the GARCH treatment, by analyzing first the data appropriateness for a GARCH model exercise. As well in this section one may find the model implementation and the post-estimation analysis. Section presents the conclusions and final remarks, as well as making proposals for future research.

2. Assessing the quality of the volatility forecasting techniques

A discussion regarding the quality measuring tools and methodology used for that would be needed at this point. What makes a model better than another? How quality is defined and how do we measure it?

The first main discussion would be that, due to the fact that the conditional evidence is unobserved, it has influenced a lot the design of the volatility models and made it difficult to benchmark between them. One conclusion of this would be that so far models with poor forecasting capacities in all empirical tests have not been yet identified, so this is the primary reason for which so many models have coexisted up to now. Besides, there is no natural and intuitive way to model the conditional heteroskedasticity, so each of such models will try to capture features that their authors considered to be important. Given this, there is a big degree of subjectivism in such a benchmarking analysis.

Furthermore, as mentioned before, when measuring the performance of a volatility model, the unobserved variance is usually replaced by squared returns, and this led to a poor out-of-sample performance. Rather than using squared inter-day returns, known to be highly noisy measures of daily volatility, Andersen and Bollerslev (1998a) used as evaluation methodology an estimated measure of the volatility using intra-day returns, resulted in good out-of-sample performance of volatility models. This is an indicator of the fact that the previously found poor performance can be explained by the use of a noisy measure of volatility.

Conditional volatility has been initially tested starting from real data taken from the US stock market. Later on, the same econometric models were applied in other stock markets, such as the Netherlands (de Jong *et al*., 1992), Japan (Tse, 1991), Singapore (Tse and Tung, 1992) or UK (Poon and Taylor, 1992). The few papers that attempted to test the predictive capacity of ARCH models have found inconsistent results. For example, Akgiray (1989) concluded that a GARCH (1,1) specification showed a better forecasting capacity when compared to other traditional models, when tested with US data. Working with Asian data, Tse (1991) and Tse and Tung (1992) came with opposite results, questioning the superiority of GARCH model. However, all three studies were converging in one result, namely that the exponential weighted moving average (EWMA) model was among the best forecasting models.

To sum up what we stated above, the present literature written on this topic contains contradictory evidence as regards the quality of the market volatility forecasts. The main message is that volatility forecasting is a notoriously complicated undertaking. There is evidence that underlines the superiority of more complex models such as ARCH models, while there is evidence on the other side as well, underlying the superiority of more simple alternatives. This is seen as an extremely problematic fact due to the difficulty that this contradiction rises in choosing the appropriate model in volatility forecasting in decision-making and analysis activities. For the second category of models, I would mention here Dimson and Marsh (1990), who, by using data from the UK equity market, concluded that the simple models offered more accurate forecasts, recommending as well the exponential smoothing and simple regression models. However, their analysis did not include ARCH models.

The same group of studies includes the work of Hansen and Lunde (2001) who used intra-day estimated measures of volatility to compare volatility models. Their objective was to evaluate whether the evolution of volatility measures has led to better forecasts of volatility when compared to the first "species" of volatility models. For this, they compared two different time series, daily exchange rate data and stock prices. Their findings showed that the more advanced models did not provide better forecasts than GARCH (1,1) model.

Hansen and Lunde evaluated the relative performance of the various volatility models in terms of predictive ability of realized volatility by using the tests developed by White (2000) and Hansen (2001) called as data snooping tests. Unfortunately, as pointed out by Bollerslev, Engle and Nelson (1994) and by Diebold and Lopez (1996), it is hard to say which criteria are the best to use when comparing volatility measures. Hansen and Lunde used seven different criteria for such comparison, which included standard criteria such as mean squared error (MSE) criterion, a likelihood criterion, and the mean absolute deviation criterion which was less sensitive to extreme mispredictions, compared to the MSE.

Thus, they considered a benchmark model and an evaluation criterion and tests for data snooping. This allowed them to know whether any of the competing models were significantly better than the benchmark. The benchmark models considered were ARCH (1,1) and GARCH (1,1) models. Their findings showed the superiority of all models as compared to ARCH (1,1), but GARCH (1,1) was not significantly outperformed in each stance. Although the analysis in one data set clearly indicated the existence of one superior model as compared to GARCH(1,1) when using the mean squared forecast error as a criterion, this did not hold up to other type of criteria that seemed to be more robust to outliers, such as the mean absolute deviation criterion.

Among the evidence that highlights the superiority of more complex models (although in some points there are some consistencies in findings with the previous mentioned evidence), there is Brailsford and Faff (1995), who, by using Australian data, showed empirically that more advanced ARCH class models and a simple regression model provided superior forecasts of volatility. A second finding of them would be that the various model rankings are sensitive to the choice of error statistic, used to assess the accuracy of forecasts. Of course, when bringing into discussion the results of Brailsford and Faff and those of Dimson and Marsh, we make a strong assumption that using different pools of data (Australian and UK) does not affect the quality of the models tested. This means that, if doing Brailsford and Faff analysis with UK data and Dimson and Marsh with Australian data, their conclusions would still hold.

Most of the literature expresses the quality as a measure between the actual and relative error statistics. The methodology that offers the most complete basis of argumentation and on which this paper is based is the one developed by Brailsford and Faff (1995). The choice for this methodology encompasses the following facts: it uses more (four) characteristics of benchmarking, it follows previous studies (Akgiray (1989), Dimson and Marsh (1990), Tse (1991) and Tse and Tung (1992)) and thus it sums up and investigates all their previously discussed models, and, the last but not least, it is straightforward and due to this, there is little space for flaws or threats. I will shortly describe this method in a few lines below.

In their (Brailsford and Faff's) paper, the quality of one model has been put in evidence by calculating four different error statistics[4] across eleven models used to forecast monthly volatility:

1. mean error (ME) statistic defined by the expression $ME = \frac{1}{90}\sum_{T=1}^{90}(\hat{\sigma}_T^2 - \sigma_T^2)$
2. mean absolute error statistic (MAE) that is a mean absolute error statistic defined by the expression $MAE = \frac{1}{90}\sum_{T=1}^{90}\left|\hat{\sigma}_T^2 - \sigma_T^2\right|$
3. root mean squared error statistic (RMSE) defined by RMSE= $\sqrt{\frac{1}{90}\sum_{T=1}^{90}(\hat{\sigma}_T^2 - \sigma_T^2)}$

   and
4. mean absolute percentage error statistic (MAPE) defined by expression MAPE= $\frac{1}{90}\sum_{T=1}^{90}\left|(\hat{\sigma}_T^2 - \sigma_T^2)/\sigma_T^2\right|$

where $\hat{\sigma}_T^2$ is the raw monthly volatility series and σ_T^2 last month's observed volatility.

They consider for testing the following models: one random walk model, one historical mean model, two moving average models, one exponential smoothing model, one exponentially weighted moving average model, one simple regression model, two standard GARCH models, and two GJR-GARCH models.

It is worth mentioning that the methodology of Dimson and Marsh (1990) differs from that of Brailsford and Faff (1995) by the fact that they standardize each error statistics by the value of the error statistics obtained from the random walk forecast. They chose such a methodology due to the fact that the statistics can be interpreted more easily relative to the benchmark forecast.

But Brailsford and Faff (1995) chose to express each (of the four above-mentioned) error statistics on a relative basis, where the benchmark is the value of the statistics for the worst performing model. Although usually fitting investigations on volatility models are run on the basis of full sample information, for benchmarking purposes these models need to be examined out-of-sample. This means that the authors selected an out-of-the sample of 90 observations (90 months) on which they tried to make predictions using the eleven models selected. So, for each of these eleven

[4] The methodology is based on evaluating and comparing 90 monthly forecast errors generated from each model which are compared by their ME, MAE, RMSE and MAPE. These 90 errors represent the out-of-sample.

models, they calculated the errors made from the difference between reality and forecasts, according to the fours error statistics. For each of the eleven models, they obtained four different error statistics. Each model was benchmarked after the size of discrepancy (size of errors) between forecast and real values. They also obtained relative error statistics, by expressing the actual statistic as a ratio relative to the worst performing model (the one that had the biggest absolute error statistic) for a given error measure. They compared the actual and relative forecast error statistics for each model across the four error measures. As previously said, the quality factor was the difference between these actual and relative errors. For each statistic, the model with the biggest difference was considered to be the benchmark (as the worst performer, since the model showed too large differences), while the model with the smallest difference was the best performing, with the highest quality. Notable to be mentioned is that for each error statistic (and model) we have (potentially) a different benchmarking model. Furthermore, the power of one test against another (by how many percentage points one model is better than another) has been calculated by the following formula:

$$\frac{(\varepsilon_i - \hat{\varepsilon}_i) - (\varepsilon_b - \hat{\varepsilon}_b)}{(\varepsilon_b - \hat{\varepsilon}_b)}$$

where

ε_i = actual forecast error statistic of the best model

$\hat{\varepsilon}_i$ = forecast error statistic of the best model

ε_b = actual forecast error statistic of the benchmark model

$\hat{\varepsilon}_b$ = relative forecast error statistic of the benchmark model

The same result (power) may be obtained by subtracting each relative error statistic from 1. Thus, for each of the four error statistics, there were provided different answers as regards the model that performs better. So, these statistics should be assigned different interpretations and/or different powers in assigning the best/worst model.

In this paper, the error statistics were interpreted and gave results as it follows:

a) ME gives the direction of over/underprediction. All models tested by Brailsford and Faff (1995) were found to be underpredicted with one exception (exponential smoothing model).

b) MAE statistics indicated GJR-GARCH (1,1) as the best model, with 35 percent higher accuracy than the benchmark model, which for this statistic was found to be the exponential smoothing model.

c) RMSE equally favors the historical mean and the simple regression model (23 percent more accurate than the benchmark model). To be noted that for this statistic GJR-GARCH (1,1) ranks fourth.

d) MAPE gives a relative indication of overall forecasting performance. In this case GJR-GARCH (1,1) model has been found with the best (actual) MAPE of 56.9 percent.

Briefly speaking, the ranking of each of the four forecasting models varies depending upon the choice of the error statistics, but it seems that GARCH ranks the best. This variability in rankings underlines the potential hazard of selecting the best model on the basis of arbitrarily chosen error statistics.

However, some consistency exists among the findings of different empirical tests, although methodologies differ. Dimson and Marsh (1990) used, instead of RMSE statistics, the primary error measure. Their conclusion was that the simple regression model is superior. This is relatively consistent with one result of Brailsford and Faff who found that the simple regression model and RMSE equally rank among the first in terms of performance. Furthermore, Dimson and Marsh found that the superiority of the simple model is insensitive to the use of the MAE statistics, which is again consistent with Brailsford and Faff's findings. However, while Dimson and Marsh found an equivalent ranking across all models between their error statistics, Brailsford and Faff's model rankings, while similar, were not entirely robust between RMSE and MAE statistics. This inconsistency was even further exacerbated when other error statistics, like MAPE statistic, were considered.

Although the purpose of the paper is not specifically showing which of the (above mentioned) models is more advanced, I consider Brailsford and Faff's paper important for this discussion as it reveals the superiority of more advanced models (namely GARCH models). However, their paper goes even beyond, by ranking the models among themselves. My paper reconsiders the question of choosing between basic and more complex models, by showing what is new in the model and by showing ultimately how GARCH works and which its improvements and disadvantages are. The discussion will be centered solely on the quality advantages each model brings and not focus on the cost component. The empirical example to follow would then offer an illustration of the advantages discussed in the paper with respect to the GARCH model.

2.1. From ARMA to ARCH model. What is new in ARCH

The autoregressive moving-average (ARMA) models join the concepts of AR and MA models aiming at keeping the number of parameters small. Their importance in finance is given mainly for their use in explaining ARCH and GARCH models, the generalized autoregressive conditional heteroskedastic model being seen as a non-standard ARMA model for an a_t^2 series. The ARMA model has been firstly proposed by Box, Jenkins and Reinsel (1994).

An autoregressive model, in its simplest form, is a model in which one uses the statistical properties of the past behavior of a variable y_t to predict its behavior in the future. In other words, we can predict the value of the variable y_{t+1} by just taking into account the sum of the weighted values that y_t took in the previous period plus the error term ε_t.

The simplest form of an ARMA model is that given by (1,1) univariate form.

r_t follows an ARMA(1,1) process if it verifies the following equation:

$$r_t - \varphi_1 r_{t-1} = \varphi_0 + a_t - \theta_1 a_{t-1},$$

where a_t is a white noise series and φ_0 is a constant; $\varphi_1 \neq \theta_1$

$r_t - \varphi_1 r_{t-1}$ gives the AR component of the model, while $\varphi_0 + a_t - \theta_1 a_{t-1}$ gives the MA component.

a_t is also called shock or innovation of an asset return at moment t.

It has a general form (general ARMA model) like:

$$r_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i r_{t-i} + a_t - \sum_{1=1}^{q} \theta_i a_{t-1}$$

with a_t as white noise series and p and q as non-negative integers.

The ARCH model assumes that r_t follows a simple time series model such as a stationary ARMA(p,q) model with some explanatory variables. It has the form:

$$r_t = \mu_t + a_t\,,\ \mu_t = \varphi_0 + \sum_{i=1}^{k} \beta_i x_{it} + \sum_{i=1}^{p} \varphi_i r_{t-i} - \sum_{1=1}^{q} \theta_i a_{t-1}$$

with x_{it} explanatory variables, while k, p and q non-negative integers; μ_t is the mean equation of r_t .

ARCH models are simple and easy to handle, and take care of clustered errors, as well as of nonlinearities. One characteristic of ARCH models is the "random coefficient problem": the power of forecast changes from one period to another.

But ARCH has some weaknesses as well. It assumes that positive and negative shocks have similar effects on volatility because it depends on the square of the previous shocks. This is rather an extreme simplification of the reality, since the price of a financial asset responds differently to positive and negative effects. Another weakness is that the ARCH model is rather restrictive. One example would be that α_i^2 of an ARCH (1, 1) model must be in the interval [0,1/3] if the series has a finite moment. The ARCH model does not contribute significantly to better understanding the source of volatility in financial time series. It only provides a mechanical method to describing the behavior of the conditional variance. But it does not explain many of the causes of such behavior. The last but not the least, the ARCH models are likely to overpredict the volatility because they respond slowly to large isolated shocks to the return series.

2.2. From ARCH to GARCH model. What is new in GARCH

Although the ARCH model has a basic form, one of its characteristics is that it requires many parameters to describe appropriately the volatility process of an asset return. Thus, alternative models must be further searched, one of them being the one developed by Bollerslev (1986) who proposed a useful extension known as the generalized ARCH.

As against the ARCH model, the Generalized Autoregressive Centralized Heteroskedastic Model (GARCH) has only three parameters that allow for an infinite number of squared roots to influence the current conditional variance. This feature allows GARCH to be more parsimonious than ARCH model, which feature explains the wide preference for use in practice, as against ARCH.

While ARCH incorporates the feature of autocorrelation observed in return volatility of most financial assets, GARCH improves ARCH by adding a more general feature of conditional heteroskedasticity. Simple models - low values of parameters *p* and *q* in GARCH(*p,q*) - are frequently used for modeling the volatility of financial returns; these models generate good estimates with few parameters. Like everything else, however, GARCH is not a "perfect model", and thus could be improved - these improvements are observed in the form of the alphabet soup that uses GARCH as its prime ingredient: TARCH, OGARCH, M-GARCH, PC-GARCH etc.

Similarly to the ARCH model, the conditional variance determined through GARCH is a weighted average of past residuals. The weights decline but never reach zero. Essential to GARCH, is the fact that it permits the conditional variance to be dependent upon previous own lags.

The model can be written as it follows. Let's assume a log return series r_t and $a_t = r_t - \mu_t$ be the innovation at time *t*. We say that a_t follows a GARCH (*m,s*) model if

$$a_t = \sigma_t \varepsilon_t \,, \ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-1}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \ ,$$

where ε_t is a sequence of iid random variables with mean 0 and variance 1,

$$\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0 \text{ and } \sum_{i=1}^{max(m,s)} (\alpha_i + \beta_i) < 1 . \tag{1}$$

Here it is understood that $\alpha_i = 0$ for i>m and $\beta_j = 0$ for *j*>*s*. The latter constraint on $\alpha_i + \beta_i$ implies that the unconditional variance of a_t is finite, whereas its conditional variance σ_t^2 evolves over time.

2.3. Extensions of the GARCH model

a) The Exponential GARCH (EGARCH) Process

The GARCH process fails in explaining the "leverage effects" which are observed in the financial time series. First observed by Black (1976), the leverage effects represent the tendency of variation in the prices of stocks to be negatively correlated with changes in the stock volatility. In other words, the effect of a shock upon the volatility is asymmetric, meaning that the impacts of "good news" (positive lagged residual) and of "bad news" (negative lagged residual) are different. The Exponential GARCH (EGARCH) model of Nelson (1991) accounts for such an asymmetric response to a shock and has the following form for (1,1):

$$\log(\sigma_t^2) = \alpha_0 + \alpha_1 \left| \frac{u_{t-1}}{\sigma_{t-1}} \right| + \beta_1 \log(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sigma_{t-1}}$$

The leverage effects are represented by γ that accounts for the asymmetry of the model. The reason for considering this asymmetric effect is that it allows the volatility to react more promptly to reductions in the prices (that represent the "bad news") rather than to the corresponding increases (that stand for "good news").

b) The Threshold GARCH (TARCH) Process

EGARCH is not the only model that accounts for the asymmetric effect of the news. Threshold GARCH (TARCH) model developed by Zakoian (1994), Glosten, Jaganathan and Runkle (1993) is similar, but the leverage effect is expressed in a quadratic form while in the case of EGARCH it is expressed in the exponential form.

A TARCH (*p,q*) process may be specified as it follows:

$\sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{k=1}^{r} \gamma_k u_{t-k}^2 I_{t-k}^-$ where I_{t-k}^- = 1, if $u_t < 0$ and = 0 otherwise. $u_{t-i} > 0$ represents the "good news" and $u_{t-i} < 0$ represents the "bad news". They have different outcomes on the conditional variance. The impact of the news is asymmetric and the leverage effects exist when $\gamma_k \neq 0$. For $\gamma_k = 0$ (for all *k*), TARCH takes the form of a standard GARCH model.

3. Experimental study

3.1 Data setting

The objective of the empirical exercise is to estimate the volatility of a particular price index (S&P500, Dow Jones Industrial Average and NASDAQ)[5] using the GARCH model and to reveal the usefulness of this model in detecting and eliminating autocorrelations. The selection of these three indices has been driven by the fact that GARCH model works best when there are long-term series (and we have daily values of these indices over a long period of time) and when there is a considerable correlation in the variables; the preference for them is that there is a strong belief in the autocorrelation within each of the three US stock indices.

Table 1

Summary statistics of DJIA, NASDAQ and S&P 500 over the sample period

	DJIA	NASDAQ	S&P 500
Average daily return	0.06%	0.07%	0.05%
Daily return volatility	1.05%	1.30%	1.04%
Excess kurtosis	36.631	9.344	55.497

Source: www.yahoofinance.com.

However, we know that the NASDAQ is predominantly focused on the "new economy", that the DJIA is weighted towards the "old economy", and the S&P strikes a balance between the two (composed as it does of the top 500 firms in the US) and this might give a hint of possible correlations among the series as well.

[5] I shall hereafter refer to the S&P500 and Dow Jones Industrial Average as simply S&P and DJIA respectively.

3.2. Data sample

There have been selected three indices[6]: DJIA, NASDAQ and S&P beginning January 1st 1980 and running up to August 27th 2008. This gives us a total of 7232 return observations. As discussed earlier, the choice for these equity indices is reasoned by their high (as expected) autocorrelations, a fact that makes their cases as ideal for applying GARCH. Let's first familiarize ourselves with the data.

When estimating parameters of a composite conditional mean/variance model, one may confront with convergence problems. Thus, the estimation may appear to stall, or show little or no progress. To avoid these difficulties, it is recommended to perform a pre-fit analysis. The main scope of this is to mitigate against any kind of convergence problems, by choosing the most appropriate model that describes the data. In our case, the scope is to find, before performing GARCH, if the data is appropriate for a GARCH-type model.

3.3 Step one: Pre-fit analysis

There are two steps in this pre-fit analysis:

A. Plotting the return series and analyzing the autocorrelation function (ACF) and the partial autocorrelation function (PACF).
B. Performing preliminary tests, such as Engle's ARCH test or the Q-test.

A. *Plotting the return series and analyzing the autocorrelation function (ACF) and the partial autocorrelation function (PACF).*

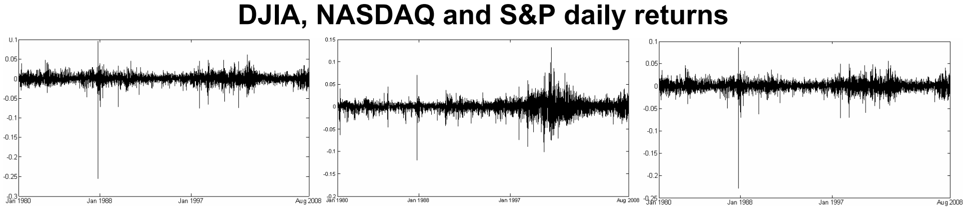
A.1. Because GARCH modeling assumes working with returns we need to convert prices into returns.

a) *price2ret* function is used to obtain the return series out of prices.
b) By using the *plot* function of Matlab, we obtain a graphical representation of these return series.

In the graphs, one may see a first hint for autocorrelations, specifically clustering effect around various data.

Figure 1

DJIA, NASDAQ and S&P daily returns



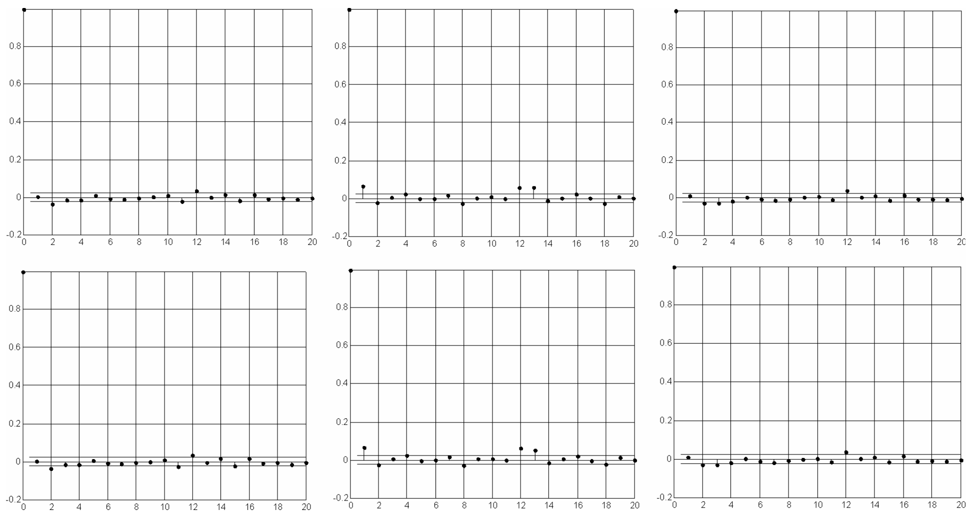
Source: www.yahoofinance.com

A.2. We check for correlation in the return series by performing the autocorrelation function to compute and display the sample ACF of the returns and by plotting the partial correlation functions.

[6] Source www.yahoofinance.com.

Figure 2

DJIA, NASDAQ and S&P autocorrelation and partial autocorrelation functions



Source: www.yahoofinance.com.

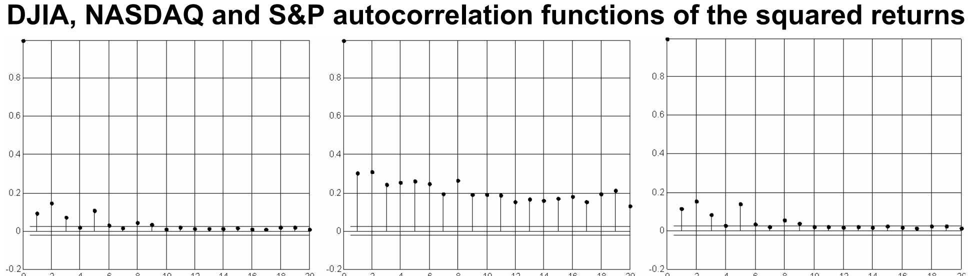
ACF and PACF graphs give some useful information on the broad characteristics of the returns. They provide indication if one needs to use any correlation structure in the conditional mean. In this particular case, we can see that ACF and PACF display some autocorrelation, but much lower than in the case of the graphs of the volatility of the returns at the previous point (A.1.).

A.3. Check for correlation in the squared returns. We need this also because although ACF of the observed returns exhibits little correlation, the ACF of the squared returns may still indicate significant correlation and persistence in the second-order moments.

We check for this by plotting the autocorrelation functions of the squared returns (Matlab code: *autocorr(variable, ^2)).*

Figure 3

DJIA, NASDAQ and S&P autocorrelation functions of the squared returns



Source: www.yahoofinance.com

It can be observed that the autocorrelation has increased for all the indices. For DJIA and S&P, it's more accentuated for smaller lags, while for NASDAQ is significantly higher for all the lags. One may notice that the ACF in all graphs appears to die out slowly, showing the possibility of a variance process close to being nonstationary.

As we can see in the previous figures that reveal the case of index daily returns, data shows clustered volatility, indicating possible correlations between present and previous volatilities. But this is more evident in the case of volatilities (the first group of charts) than in the case of autocorrelations between the daily returns (second and third group of charts).

In conclusion, there has been detected significant clustering in all cases, for all indices, thing that is a good indicator of the fact that these indices are an appropriate choice to reveal the usefulness of the GARCH as purpose of the present study.

B. Performing preliminary tests, such as Engle's ARCH test or the Q-test.

However, the pre-estimation analysis has not finished. Although the autocorrelation has been detected visually through the graphs, we have to quantify it. We can quantify the preceding qualitative checks for correlation using formal hypothesis checks, like Ljung-Box-Pierce Q-test and Engle's ARCH test.

By performing a Ljung-Box-Pierce Q-test, we can verify, at least approximately, the presence of any significant correlation in the raw returns when tested for up to 20 lags of the ACF at the 0.05 level of significance. The lbqtest function performs a lack-of-fit model misspecification, based on Q statistics. Under the null hypothesis that the model fit is adequate, the test statistics is asymptotically chi-square distributed. The rejection or acceptance of the null hypothesis is given by the decision vector H: 0 indicates the acceptance of the null hypothesis that the model fit is adequate (meaning that no serial correlation at the corresponding element of lags), 1 means rejection.

The results for LBPQ are as follows. We can thus check that no significant correlation is present in the raw returns when tested for up to 20 lags of the ACF. However, since we are more interested in how more recent data influence future variation, there will be performed both Ljung-Box-Pierce Q-test and Engle's ARCH test at 3, 5 and 7 lags, by default chosen alpha of 0.05.

Table 2

Ljung-Box-Pierce Q-test output for heteroskedasticity

DJIA- LBPQ

H	P-value	Statistic	Critical value
1.0000	0.0085	11.6953	7.8147
1.0000	0.0177	13.6908	11.0705
1.0000	0.0303	15.4820	14.0671

NASDAQ-LBPQ

H	P-value	Statistic	Critical value
1.0000	0.0000	32.8859	7.8147
1.0000	0.0000	37.2295	11.0705
1.0000	0.0000	39.0801	14.0671

S&P 500- LBPQ

H	P-value	Statistic	Critical value
1.0000	0.0037	13.5066	7.8147
1.0000	0.0059	16.3478	11.0705
1.0000	0.0083	18.9556	14.0671

Source: www.yahoofinance.com

Table 3

Engle's test output for heteroskedasticity

DJIA- Engle

H	P-value	Statistic	Critical value
1.0000	0.0000	215.0972	7.8147
1.0000	0.0000	278.4061	11.0705
1.0000	0.0000	280.7218	14.0671

NASDAQ-Engle

H	P-value	Statistic	Critical value
1.0000	0.0000	1.1449	0.0078
1.0000	0.0000	1.3259	0.0111
1.0000	0.0000	1.3702	0.0141

S&P 500- Engle

H	P-value	Statistic	Critical value
1.0000	0.0000	253.6795	7.8147
1.0000	0.0000	360.2872	11.0705
1.0000	0.0000	363.0812	14.0671

Source: www.yahoofinance.com

All tests show H=1 and all parameters are higher than their critical values, thing that makes us conclude that we reject the null hypothesis. Thus, some serial correlation exists at the corresponding elements of Lags.

Engle's test shows significant evidence in support of the GARCH effects, like heteroskedasticity. Under the null hypothesis that a time series is a random sequence of Gaussian disturbances (i.e., no ARCH effects exist), this test statistic is also asymptotically Chi-Square distributed. Like in the LBPQ case, the H vector is a Boolean decision flag. When 0, it implies the existence of no significant correlation (not rejection of the decision null hypothesis) and when 1 means that significant correlation exists (rejection of the null hypothesis). The Matlab code for it is *archtest.* The results for the Engle's test are displayed as in the above tables.

We can see that for the DJIA, NASDAQ and S&P, we reject the null hypothesis, so we have significant correlation in each time series.

After performing Ljung-Box-Pierce and Engle tests for heteroskedasticity it can be concluded that all these series are heteroskedastic (of course, some more than others). This indicates that the returns index for each of the three cases may be an ideal case for GARCH treatment.

And with this, we finish the pre-estimation part of the GARCH model. Before starting to perform this model, I restate the main objective of the empirical exercise: to see if GARCH is able to eliminate autocorrelations just enough to make LBPQ and Engle tests indicate no further heteroskedasticity.

3.4 Step two: Performing GARCH technique

The first step in running a GARCH algorithm consists in performing a univariate GARCH and check the necessity of adding extra variables. Necessity of adding extra variables would come up if the chosen exogenous variables would not be sufficient to explain the endogenous one (which in our case is the returns of each price index). You may find the results below:

Table 4

Summary of the GARCH(1,1) model for (a) DJIA (b) NASDAQ (c) S&P 500

	Value	T-stat		Value	T-stat		Value	T-stat
C	5.61×10^{-4}	5.61	**C**	7.21×10^{-4}	7.33	C	5.24×10^{-4}	5.34
α_0	1.39×10^{-6}	10.42	$\boldsymbol{\alpha_0}$	1.82×10^{-6}	11.87	$\boldsymbol{\alpha_0}$	1.1×10^{-6}	9.93
α_1	9.17×10^{-1}	306.40	$\boldsymbol{\alpha_1}$	8.71×10^{-1}	184.02	$\boldsymbol{\alpha_1}$	9.23×10^{-1}	346.28
β_1	7.21×10^{-2}	47.01	$\boldsymbol{\beta_1}$	1.20×10^{-1}	27.54	$\boldsymbol{\beta_1}$	6.86×10^{-2}	45.49
(a)			(b)			(c)		

Source: www.yahoofinance.com

The rationale of this is to use a parsimonious model if it is "good enough", where the goodness of the model depends on our prescribed requirements. In our case, the rationale is to use the best possible univariate GARCH model. This translates into finding if the coordinates *m* and *n* of GARCH(*m,n*) should be selected in order to optimize the trade-off between the extra parameters and the extra predictive ability achieved. The selection of the variables *m* and *n* is optimized independently of the other models under consideration. The Matlab code used is *garchfit(djiaretrecent); garchdisp(coeff,errors)* (for DJIA index).

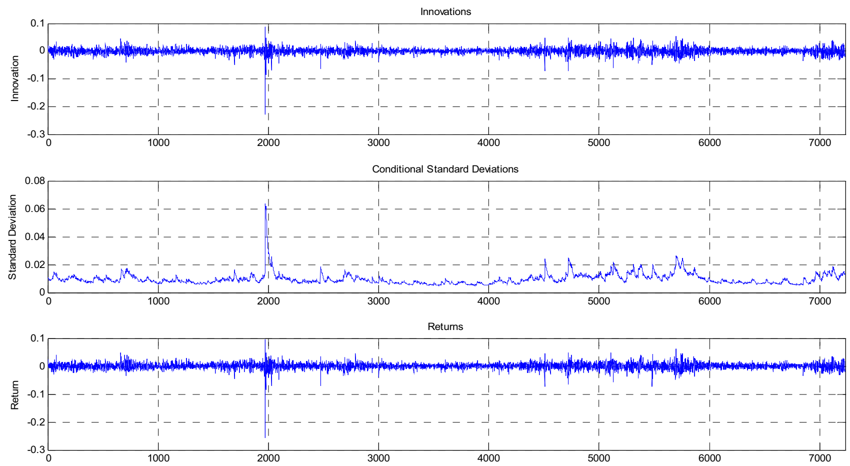
The results obtained from the univariate GARCH(1,1) models are summarized in Table 4. Recall that the GARCH(1,1) model is $y_t = C + \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 + \beta_1\varepsilon_{t-1}^2$.

We thus see from the Table 4 that we can reject the null hypothesis that α_0 and α_1 are separately equal to zero (since the t-values are outside +/-1.96 interval, thus we are in the rejection region). In other words, it is appropriate to model the time series of volatility as a GARCH(1,1). We pause to consider the "visual effect" of the GARCH(1,1) decomposition (Figure 4).

Figure 4

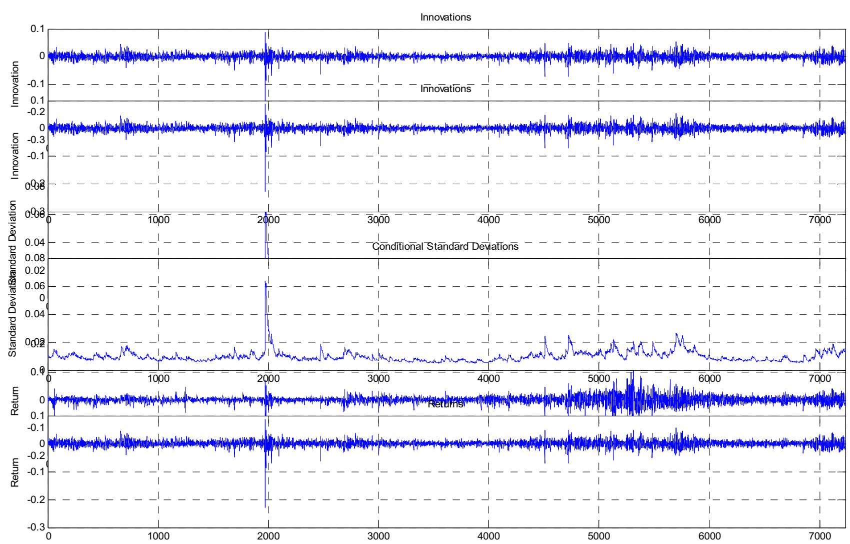
GARCH's split of the variance into variance innovations and conditional standard deviations

DJIA:



NASDAQ:

S&P:



Source: www.yahoofinance.com

3.5. Step three: Obtaining residuals from GARCH(1,1) and standardizing them

From the 3.4, it has been obtained that the GARCH(1,1) models for DJIA, NASDAQ and S&P 500 are as it follows:
DJIA:

$$y_t = 5.61 \times 10^{-4} + \varepsilon_t$$

$$\sigma_t^2 = 1.390 \times 10^{-6} + 0.917\sigma_{t-1}^2 + 0.0721\varepsilon_{t-1}^2$$

NASDAQ:

$$y_t = 7.21 \times 10^{-4} + \varepsilon_t$$

$$\sigma_t^2 = 1.82 \times 10^{-6} + 0.871\sigma_{t-1}^2 + 0.120\varepsilon_{t-1}^2$$

S&P 500:

$$y_t = 5.24 \times 10^{-4} + \varepsilon_t$$

$$\sigma_t^2 = 1.10 \times 10^{-6} + 0.923\sigma_{t-1}^2 + 0.0686\varepsilon_{t-1}^2$$

For each day (of the 7232 days of our sample), we calculate the volatility forecast and call this σ_t. We use this calculated variance forecast to obtain the standardized residuals of the daily returns for each day. In other words, we calculate $\frac{y_t - \overline{y_t}}{\sigma_t}$ as for each t we know the return y_t. This step is necessary in order to test if the innovations provided by GARCH, in their standardized form, still exhibit autocorrelation. If they would still show autocorrelation, that would mean that GARCH is not a proper model to be used.

Now, we are ready for our post-estimation analysis. In this part we will, first, compare the residuals, conditional standard deviations, and returns, after which we will plot and compare the correlation of the standardized innovations. Finally, we will quantify and compare the correlation of the standardized residuals.

3.6. Step four: Post-estimation analysis

Post-estimation analysis consists of three steps:

1. *Compare residuals, conditional standard deviations, and returns*

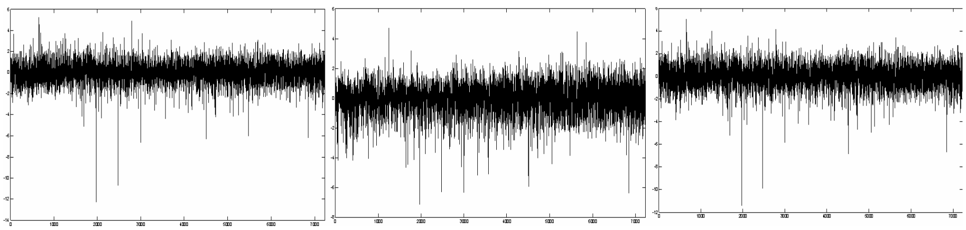
By using the Matlab function *garchplot(innovations, sigmas, nasdaqret)*, we split the variance into variance innovations and conditional standard deviations. The GARCH test uses this step in order to investigate if the fitted innovations exhibit volatility clustering.

From the graph of each index, we notice some volatility clustering in innovations and returns, but much less in innovations than in returns. We want to see if by performing GARCH, the autocorrelation of the standardized innovations disappears, indicating the effectiveness of GARCH model.

2. Plot and compare the correlation for the standardized innovations
We saw that the previous fitted innovations display some volatility clustering. But if we plot the standardized innovations (the innovations divided by their conditional standard deviation), however, they appear generally stable with little clustering. The Matlab code for calculating the standardized invocations is *djiainnrecent=innovations./sigmas* and for graphing them is *plot(innovations./sigmas).*

Figure 5

Correlation of the standardized innovations for DJIA, NASDAQ and S&P

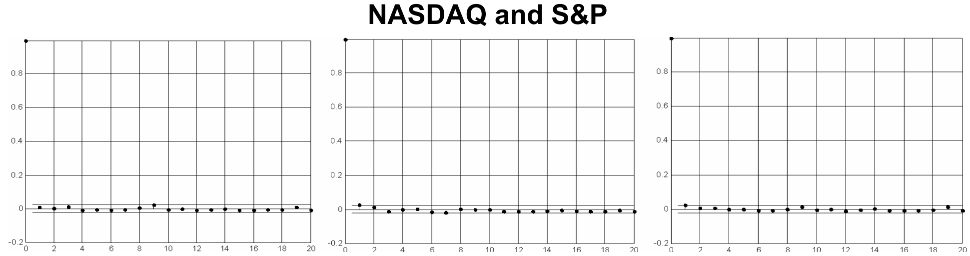


Source: www.yahoofinance.com

As well, if we plot the ACF of the squared standardized innovations, we will not find any further correlation. The Matlab code for it is: *autocorr((innovations./sigmas).^2)*

Figure 6

Autocorrelation function of the standardized innovations for DJIA, NASDAQ and S&P



Source: www.yahoofinance.com

By observing the above ACF plots, we see no further correlation. Furthermore, if we compare the ACF of the squared standardized innovations in this figure to the ACF of the squared returns prior to the fitting the default model, we see that this GARCH model sufficiently explains the heteroskedasticity in the raw returns.

3. Quantifying and comparing the correlation of the standardized innovations
At this phase, we compare the results of the Q-test and ARCH-test with the results of the same tests performed in the pre-estimation analysis. I will use this time the standardized residuals. By this action, I want to see if GARCH has treated efficiently the data (Matlab code for the Q-test: *lbqtest((innovations./sigmas).^2,[3 5 7]',0.05))*

Q-test results:

Tables 5 and 6

Ljung-Box-Pierce Q-test output and Engle's test output for heteroskedasticity for standardized residuals

DJIA- LBPQ / *DJIA- Engle*

H	P-value	Statistic	Critical value	H	P-value	Statistic	Critical value
0.0000	0.6242	1.7576	7.8147	0.0000	0.6274	1.7432	7.8147
0.0000	0.7735	2.5201	11.0705	0.0000	0.7686	2.5524	11.0705
0.0000	0.8842	3.0092	14.0671	0.0000	0.8815	3.0381	14.0671

NASDAQ- LBPQ / *NASDAQ- Engle*

H	P-value	Statistic	Critical value	H	P-value	Statistic	Critical value
0.0000	0.2604	4.0100	7.8147	0.0000	0.0736	6.9487	7.8147
0.0000	0.5437	4.0401	11.0705	0.0000	0.2151	7.0756	11.0705
0.0000	0.6438	5.1326	14.0671	0.0000	0.1318	11.1589	14.0671

S&P 500- LBPQ / *S&P 500- Engle*

H	P-value	Statistic	Critical value	H	P-value	Statistic	Critical value
0.0000	0.0732	6.9610	7.8147	0.0000	0.2677	3.9428	7.8147
0.0000	0.2138	7.0932	11.0705	0.0000	0.5473	4.0146	11.0705
0.0000	0.1254	11.3174	14.0671	0.0000	0.6537	5.0515	14.0671

Source: www.yahoofinance.com

Although in the pre-estimation analysis both Q-test and ARCH test indicated rejection of their null hypothesis, now we find out that when using standardized innovations (provided by GARCH) based on the estimated model, the same tests indicate acceptance (H=0) of the same null hypothesis. These results confirm the explanatory power of the default model and the existence of the GARCH effects.

We have GARCH effects and also correlation between innovations that disappear after treating the data. In conclusion to the post-estimation part, the GARCH model is a proper model to be used to explain the variances of the three indices. Thus, our intuitive choice of the three indices is justified, and GARCH proved to be an efficient model to be used for long enough and highly autocorrelated time series.

4. Conclusions and future research

The many advantages of GARCH forecasting techniques, among them their flexibility and accuracy, place them in a unique position to fulfill many of the requirements of the practitioners, especially in the back office risk management and front office trading systems. However, their use is restricted to the long time series (1000 observations proved to be a small sample, and fewer than this does not provide any signal picked up; 5000 observations is not a very large sample in terms of accuracy with which parameters are estimated, but it is a reasonable length with which it can be worked with). GARCH models require several years of daily data in order to be trustworthy.

Among further shortcomings to be mentioned we find that the model takes into account only the size of the movement of the returns (magnitude), not the direction as well. Investors behave and plan their actions differently depending on whether a share moves up or down, which explains why the volatility is not symmetric in the stance of the directional movements. Market declines forecast higher volatility than comparable market increases. This represents the leverage effect described by Gourieroux and Jasiak (2002). Both GARCH and ARCH have this limitation that impedes them from very accurate forecasts.

Further research should be done for quantifying the advantages of some models against others not only from the point of view of their results' quality, but in terms of the costs involved. Some models might need too long time for data processing and too many resources involved, for a quality level that is not necessary. A trade-off between quality and amount of resources required might thus prove a new perspective of benchmarking such models.

References

Akgiray, V. (1989), "Conditional heteroskedasticity in time series of stock returns: Evidence and forecasts", *Journal of Business*, 62: 55-80.

Albu, L. L. (2003), "Short-Term Forecast", *Romanian Journal of Economic Forecasting*, 4(1): 197-199.

Andersen, T. G. and Bollerslev, T. (1998a), "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts", *International Economic Review*, 39(4): 885-905.

Andreica, C., Andreica, M., Andreica, R., Miclaus, I. and Ungureanu, M. (2007), "Conclusion on Using the Statistical Methods in Forecasting the Structure Evolution of an Economic Indicator System", *Economic Computation and Economics Cybernetics Studies and Research*, 41(1-2): 147-157.

Andreica, M. (2006), "A Model to Forecast the Evolution of the Structure of a System of Economic Indicators", *Romanian Journal of Economic Forecasting*, 7(1): 65-73.

Black, F. (1976), "Studies of stock price volatility changes", *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 177-181.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics,* 31: 307-327.

Bollerslev T., Engle, R. F. and Nelson, D.B. (1994), "ARCH Models", *Handbook of Econometrics*, Amsterdam: Elsevier Science B. V.

Box G. E. P., Jenkins G. M. and Reinsel G. C. (1994), *Time Series Analysis: Forecasting and Control*, New Jersey: Prentice Hall, Inc.

Brailsford, T.J. and Faff, R. W. (1995), "An evaluation of volatility forecasting techniques", *Journal of Banking and Finance*, 20: 419-38.

de Jong, F., Kemna, A. and Kloek, T. (1992), "A contribution to event study methodology with an application to the Dutch stock market", *Journal of Banking and Finance*, 16: 11-36.

Diebold, F. X. and Lopez, J. A. (1996), "Forecast Evaluation and Combination", *Handbook of Statistics*, North-Holland : Elsevier Science B.V.

Dimson, E. and Marsh, P. (1990), "Volatility forecasting without data snooping", *Journal of Banking and Finance*, 14: 399-21.

Dosescu, T. and Raischi, C. (2008), "The Econometric Model of a Random System that Generates Time Series Data", *Economic Computation and Economics Cybernetics Studies and Research*, 42(3-4): 85-97.

Dospinescu, A. S. (2005), "Combining the Forecasts Using a Statistical Approach", *Romanian Journal of Economic Forecasting*, 6(2): 72-84.

Engle, R. F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, 50(4): 987-1007.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), "On the relation between the expected value and the volatility of the nominal excess return on stocks", *Journal of Finance*, 48: 1179-1801.

Gourieroux, C. and Jasiak, J. (2002), "Nonlinear autocorrelograms with application to intertrade durations", *Journal of Time Series Analysis*, 23: 127-154.

Hansen, P. R. (2001), "An unbiased and powerful test for superior predictive ability", *Working Papers* 2001-2006, Brown University, Department of Economics.

Hansen, P. R. and Lunde, A. (2001), "A comparison of volatility models: does anything beat a GARCH(1,1)?", *Working Paper Series* No. 84, Aarhus School of Business, Centre for Analytical Finance, 1-41.

Harvey, A. C., Ruiz, E. and Shephard, N. (1994), "Multivariate stochastic variance models", *Review of Economic Studies*, 61(2): 247-64.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics*, 12(4): 371-89.

Lupu, R. (2006), "Using BDS to Check the Forecasting Power of the Stochastic Volatility Model", *Economic Computation and Economics Cybernetics Studies and Research*, 40(1-2): 223.

Melino, A. and Turnbull S.M. (1990), "Pricing foreign currency options with stochastic volatility", *Journal of Econometrics*, 45: 239-265.

Năstac, I., Dobrescu, E. and Pelinescu, E. (2007), "Neuro-Adaptive Model for Financial Forecasting", *Romanian Journal of Economic Forecasting*, 8(3): 19-41.

Nelson, D. B. (1991), "Conditional heteroskedasticity in asset returns: A new approach", *Econometrica*, 59: 347-370.

Nicholls, D. F. and Quinn, B. G. (1982), *Random Coefficient Autoregressive Model: An introduction*, New York: Springer-Verlag.

Poon, S. H. and Taylor, S. J. (1992), "Stock returns and volatility: An empirical study of the UK stock market", *Journal of Banking and Finance*, 16: 37-59.

Scutaru, C., Sâman, C. and Stănică, C. (2008), "Predictability and Complexity in Macroeconomics. The case of Gross Fixed Capital Formation in the Romanian Economy", *Romanian Journal of Economic Forecasting*, 9(4): 196-205.

Taylor, S. J. (1994), "Modeling stochastic volatility: a review and comparative study", *Mathematical Finance*, 4(2): 183-04.

Tsay, R. S. (1987), "Conditional heteroskedastic time series models", *Journal of American Statistical Association*, 82: 590-604.

Tse, Y. K. (1991), "Stock return volatility in the Tokyo stock exchange", *Japan and the World Economy*, 3: 285-298.

Tse, Y. K. and Tung, S. H. (1992), "Forecasting volatility in the Singapore stock market", *Asia Pacific Journal of Management*, 9: 1-13.

White, H. (2000), "A reality check for data snooping", *Econometrica*, Econometric Society, 68(5): 1097-1126.

Zakoian, J.M. (1994), "Threshold Heteroskedasticity Models", *Journal of Economic Dynamics and Control*, 15: 931-955.

Yahoo Finance (2008), S&P 500, NASDAQ and Dow Jones Industrial Average data bases, http://www.yahoofinance.com, Accessed in August 2008.