Universidade do Minho
Escola de Economia e Gestão

Forecasting volatility using GARCH models

Francisco João Matos Costa

A thesis presented for the degree of

Master in Finance

Research under the supervision of professor Nelson Areal

School of Economics and Management
University of Minho
Portugal

April 2017

Acknowledgements

First and foremost, I would like to acknowledge my adviser, professor Nelson Areal, for all the support and cooperation that for sure made this dissertation better. I am also grateful to all the professors of Master in Finance that were a big influence in the choice of this subject.

I would like to thank to all my friends and colleagues that supported me throughout the last year and always kept pushing me to finish this research and assisted me in the most complicated moments.

I dedicate this research to my family, especially my parents and my brother, that encouraged me unconditionally and allowed me to work in this research without any concerns and distractions.

Resumo

Esta dissertação tem como ponto central a previsão da volatilidade usando vários modelos GARCH (*General autoregressive conditional heteroeskedasticity*) de modo a testar qual tem a melhor capacidade de previsão. O foco desta dissertação é o estudo do mercado dos EUA. Os dados usados para este estudo são cotações do NASDAQ-100, de 1986 até 2016. Neste estudo são considerados três períodos de estimação para os modelos GARCH: 500 dias, 1000 dias e 2000 dias de modo a minimizar a possível presença de mudanças na estrutura dos dados. Regressões lineares (Mincer-Zarnowitz) foram efectuadas de forma a avaliar a performance individual de cada modelo GARCH. Depois disso, de forma a detectar qual o melhor modelo para prever a volatilidade, o teste de SPA de Hansen and Lunde (2005) foi ultizado. Os resultados são conclusivos de que os modelos são semelhantes no que toca à previsão da volatilidade condicional do dia seguinte, com a possível excepção do modelo IGARCH. O modelo GJR não apresenta resultados satisfatórios quando a janela de estimação utilizada na estimação dos modelos é de 1000 dias.

Palavras chave: GARCH, volatilidade, previsão.

Abstract

The purpose of these research is to forecast volatility using different GARCH (General autoregressive conditional heteroeskedasticity) models in order to test which model has best forecasting ability. The focus of this research is the US market. The data is composed by NASDAQ-100 quotations from 1986 to 2016. The study considers three estimation periods for the GARCH family models: 500 days, 1000 days and 2000 days in order to minimize structure changes that might be present in the data. A series of Mincer-Zarnowitz regressions were completed in order to assess the performance of each GARCH model. Afterwards, the SPA test from Hansen and Lunde (2005) is used in order to detect which is the best model. The empirical results show that the GARCH models produce similar results in what comes to forecasting next day conditional volatility, with the possible exception of the IGARCH model. There is also reason to believe that the GJR model does not provide good estimations of volatility when the rolling window used in the estimation of the models is 1000 days.

Key words: GARCH, volatility, forecast.

Contents

Chapter 1

Introduction

Risk has become an important part of financial analysis, both for risk management and for regulatory purposes. Different investors have different levels of risk that they are willing to take. Volatility is not the same as risk although it can be perceived as a measure of risk. There are several definitions for volatility. In finance, volatility is the variation of a financial asset over a certain period of time measured by the standard deviation of returns. It is the risk of change in an asset value. High levels of volatility indicate that the security will have a large expected change, on the other hand, a lower value of volatility means that the change might just not be so dramatic.

As a result of the impact of volatility of financial assets in the economy many studies have been conducted. The ability to forecast financial market volatility is important for portfolio selection and asset management as well as for the pricing of primary and derivative assets (Engle and Ng, 1993). Volatility can not be observed therefore its difficult to assess which models are better in terms of the estimation of volatility itself. Several studies were conducted in order to forecast volatility using different models. Kim et al. (1998), Bollerslev et al. (1992), West and Cho (1995) and Andersen and Bollerslev (1998) are some examples. Bollerslev (1986) and Taylor (1986) introduced the GARCH model, an extension on the ARCH model introduced by Engle (1982), in order to produce better forecasts of conditional volatility and since then several authors introduced additional characteristics to the "traditional" GARCH model with the objective to capture different attributes of returns that have strong influence in the estimation of the conditional volatility. Many authors have used GARCH family models to forecast volatility. Pagan and Schwert (1990), Franses and Dijk (1996), Brailsford and Faff (1996), Corrado and Miller (2005) are some examples.

In this research different GARCH family models are used in order to forecast the one-day-ahead conditional volatility of the NASDAQ-100 stock market. First I conduct a series of Mincer-Zarnowitz regressions in order to understand the forecasting power of each GARCH family model to forecast volatility. After that I evaluate which model is best in terms of forecasting ability. In order to do that I use a test for superior ability introduced by (Hansen, 2005).

The literature review is divided in four parts. In the first part the characteristics of returns will be described. Next, the time varying nature of volatility is presented as well as different methods that can be used to model volatility. In the third part there will be a discussion on what model is the most suitable for this type of study and in the final part I present the results of different empirical studies on the subject.

Chapter 3 presents the methodology. In this chapter is explained how returns are calculated and the conditional variance of several GARCH models is presented. Also in this section univariate and multivariate regressions and a test of superior ability are discussed.

Chapter 4 describes the data used in the research.

Chapter 5 presents the results. This chapter is divided in two sections: the measures of forecasting accuracy where Mincer-Zarnowitz regressions results are presented and analyzed in the first part and in the second part the results for the SPA test are exhibited and analyzed.

The last chapter is a conclusion of the most important aspects found in the research.

Chapter 2

Literature review

2.1 Returns and their characteristics

Prices are the base for great part of financial studies. It is common that those prices are converted in returns. They are the gain or loss of a security in a particular period. If investors are willing to take more risk they should be rewarded with higher yields, however there is also a risk of a higher potential loss. Returns are used in finance due to their statistical characteristics, among others they are stationary and only weakly correlated through time. Returns are also used because they are unit-free.

The total return is the sum of the capital gain and any dividend payments. It is possible to adjust a price time series so that the dividends are added back to generate a total return index. If the price were a total return index, returns generated provide a measure of total return that would accrue to a holder of the asset during a period of time (Brooks, 2014).

Returns have three usual characteristics. These characteristics are called stylized facts. Firstly, returns do not follow a normal distribution, in most cases the distribution is symmetric (Taylor, 2005), however, in some cases, distribution is characterized by a asymmetric shape, normally skewned to the left and high Kurtosis which implies that returns have a high peak (Mandelbrot, 1963) and fat tails (Fama, 1965). The second stylized fact of returns is that there is almost no correlation between returns for different days. The last stylized fact is that there is positive dependence between absolute returns on nearby days and thus for squared returns (Taylor, 2005).

2.2 Different ways to model volatility and its variability

Volatility is a measure of return variability over a certain period of time. As reflected in the literature volatility is one of the most important features in finance. It is often used as a measure of total risk of financial assets. It is used in a variety of Value-at-risk models for measuring market risk. It is also an important part of the Black-Scholes (1974) formula for deriving the price of traded options and other theoretical asset pricing models such as the Sharpe (1964) model. Despite many studies on this subject not all of them are focused on the forecasting ability of volatility.

2.2.1 Fluctuations in volatility

Volatility is not constant over time. There seems to be no plausible explanation for the changes in volatility, however there is a small part that can be explained.

It is common that large returns (of any sign) are followed by large returns (of any sign) and vice versa. This phenomena is called volatility clustering and its possible that it occurs due to the fact that information also arrives in a short period of time (Mandelbrot, 1963). High volatility produces more dispersion in returns than low volatility, thus returns are more spread out if volatility is higher. A high volatility cluster will contain several large positive returns as well as negative returns, in the opposite case, there will be few or none in a low volatility cluster (Taylor, 2005). Asset price changes are characterized by periods of low volatility where the change in price is small and alternate in an irregular manner with periods of high volatility where prices tend to increase (Gaunersdorfer and Hommes, 2007). Arguably one aspect that can explain this phenomena is the arrival of new information which affects the price also occurs in bunches rather than being evenly spaced over time (Brooks, 2014).

The announcement effect, or in other words, the arrival of new information to the market of some economic factors such as gross national product (GNP), inflation and unemployment rate could also have an impact in volatility. In the literature there is an ongoing discussion about this matter. Flannery and Protopapadakis (2002) found that macroeconomic factors such as real GNP and money growth, despite that they affect returns, are associated with lower return volatility. In the other hand, Glosten et al. (1993) found that inflation has a strong impact on return volatility.

The leverage effect (Black, 1976), the tendency for volatility to increase more after a large fall in prices than after an increase on the same amount, is also an important feature that impacts volatility. The phenomena can be explained in part by fixed financial and operating costs. For a company that has debt and equity outstanding, the leverage effect is normally higher when the value of the firm falls, which raises equity returns volatility if the whole is constant (Bollerslev et al., 1994).

In periods of crisis volatility is expected to increase and tend to be more stable, thus lower, during periods of economic stability. Schwert (1989) states that the average level of volatility is much higher in periods of recession and that, particularly for financial asset returns, volatility increases during periods of financial market crisis.

2.2.2 Modeling volatility

Estimation of volatility, is arguably one of the most important topics in finance. The volatility of financial assets is a key feature for measuring risk underlying many investment decisions in financial practice (Gaunersdorfer and Hommes, 2007). Volatility can be estimated using different models.

Historical volatility is an easy approach to estimate volatility. This method consist in determining the variance/standard deviation of returns over a period of time which then serves as the estimate volatility of the forthcoming periods (Brooks, 2014). Implied volatility is another method that can be used in order to obtain volatility forecasts. Based on the price of an option we can calculate the volatility of the option that is implied by the value of the option itself (Dumas et al., 1998) using, for instance, the Black-Scholes (1974) formula.

Volatility can also be modeled through the use of exponentially moving average models. This method is an extension of the historical model which consists in allowing more recent observations to have more impact than older ones. Volatility is more likely to be affected by recent events and that is not accounted for in the simpler historical model. This model also allows for a smoother transfer to shocks, in other words an historical approach could lead to an artificial level of volatility that can lead to a erroneous market expectation (Hunter, 1986).

Autoregressive volatility models can also be used to model volatility. An autoregressive process is a stochastic process that operates under the assumption that past values have an effect on current values. The autoregressive model is an extension of the random walk that includes passed terms. This model depends linearly on passed terms and is basically a regression model where the previous terms are the predictors. The moving average model is similar to the autoregressive model except it is a linear combination of the past white noise terms (uncorrelated variables, stationary and mean 0). The difference between the two is that the MA(1) model will only take into account the last shock instead of all previous ones that are considered by the AR(1) model. The ARMA model is a combination of the AR and MA models with the intention of capturing both market participant effects (momentum and mean reversion effects) and characterize shock information (unexpected event). AR, MA and ARMA models are not conditionally heteroskedastic an so they do not take in consideration volatility clustering (Fabozzi et al., 2014).

There is strong evidence that suggests that GARCH models with non-normal distributions are more robust in what comes to volatility forecasting than other historical models (Liu and Morley, 2009). The most common methods of modeling volatility are the ARCH (autoregressive conditional heteroskedastic) (Engle, 1982) and GARCH (general autoregressive conditional heteroskedastic) models (Bollerslev, 1986; Taylor, 1986). The GARCH model is an extension of the ARCH model that recognizes the difference between conditional and unconditional variance allowing for the conditional variance to change over time as a function of past errors (Bollerslev, 1986). The GARCH model also allows both for a longer memory and a more flexible lag structure. These models are non-linear models.

The assumption of the classic linear regression model that the variance of the errors is constant is known as homoscedasticity, in other words it is assumed that var$(\varepsilon_t = \sigma^2)$. On the other hand, if the variance of the errors is not constant this would be known as heteroscedasticity. Its expected that the variance of the errors will not be constant over time, thus a model that does not assume that the variance is constant is more appropriate.

If the errors are heteroskedastic but assumed homoscedastic it could lead to a wrong standard error estimation (Brooks, 2014).

Volatility clustering is taken in consideration in the following models.

2.2.3 ARCH

ARCH models were introduced by Engle (1982). The model assumes that the variance of the current error term is related to the size of previous period error terms.

Assuming that the return on an asset is given by

$$r_t = \mu + \sigma_t \varepsilon_t \tag{2.1}$$

where ε_t is a sequence of N(0,1) i.d.d. random variables. The residual term at time t, $r_t - \mu$, can be defined as

$$a_t = \sigma_t \varepsilon_t \tag{2.2}$$

In the ARCH model,

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 \tag{2.3}$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$ to ensure positive variance and $\alpha < 1$ to ensure that the model is stationary. The returns are conditional on all information up to time $t-1$ and are uncorrelated but not i.i.d. The kurtosis is defined as

$$\text{Kurt}(a_t) = \frac{E[a_t^4]}{(E[a_t^4])^2}$$

If a_t were normally distributed the Kurtosis would be 3. However,

$$\begin{aligned} \text{Kurt}(a_t) &= \frac{E[\sigma_t^4]E[\varepsilon_t^4]}{(E[\sigma_t^2])^2(E[\varepsilon_t^2])^2} \\ &= \frac{3E[\sigma_t^4]}{(E[\sigma_t^2])^2} \end{aligned}$$

meaning that kurt(a_t)>3.

The unconditional variance of a_t is

$$\begin{aligned}
\text{Var}(a_t) &= E[a_t^2] - (E[a_t])^2 \\
&= E[a_t^2] \\
&= E[\sigma_t^2 \varepsilon_t^2] \\
&= E[\sigma_t^2] \\
&= \alpha_0 + \alpha_1 E[a_{t-1}^2]
\end{aligned}$$

and since a_t is a stationary process, Var(a_t)=Var(a_{t-1})=E[a_{t-1}^2], so

$$\text{Var}(a_t) = \frac{\alpha_0}{1-\alpha_1}$$

2.2.4 GARCH family

The GARCH model was introduced by Bollerslev (1986) and Taylor (1986). The model allows the conditional variance to depend on the previous lags.

In a GARCH(1,1) model,

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{2.4}$$

where α_0>0, α_1>0, β_1>0, and $\alpha_1 + \beta_1$<1.

An ARCH(1,1) is an ARMA(1,1) model on squared residuals by substituting $v_t = a_t^2 - \sigma_t^2$ in equation 2.4, so

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
a_t^2 - v_t &= \alpha_+ \alpha_1 a_{t-1}^2 + \beta_1 (a_{t-1}^2 - v_{t-1}) \\
a_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) a_{t-1}^2 + v_t - \beta_1 v_{t-1}
\end{aligned}$$

which is an ARMA(1,1) process on the squared residuals.

The unconditional variance of a_t is

$$\begin{aligned}
\mathrm{Var}(a_t) &= E[a_t^2] - (E[a_t])^2 \\
&= E[a_t^2] \\
&= E[\sigma_t^2 \varepsilon_t^2] \\
&= E[\sigma_t^2] \\
&= \alpha_0 + \alpha_1 E[a_{t-1}^2] + \beta_1 \sigma_{t-1}^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1) E[a_{t-1}^2]
\end{aligned}$$

and since a_t is a stationary process,

$$\mathrm{Var}(a_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

provided that $a_t = \sigma_t \varepsilon_t$, the unconditional variance of returns is also $\alpha_0/(1 - \alpha_1 - \beta_1)$.

A GARCH (1,1) model can be writen as an ARCH(∞),

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta(\alpha_0 + \alpha_1 a_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_0 \beta_1 + \alpha_1 \beta_1 a_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2 \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_0 \beta_1 + \alpha_1 \beta_1 a_{t-2}^2 + \beta_1^2(\alpha_0 + \alpha_1 a_{t-3}^2 + \beta_1 \sigma_{t-3}^2) \\
&\vdots \\
&= \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{i=0}^{\infty} a_{t-1-i}^2 \beta_1^i
\end{aligned}$$

so that the conditional variance at time t is the weighted sum of past squared residuals and the weights decrease as the time goes back.

This model has some drawbacks (the non-negativity conditions may be violated by the estimated model; despite that they can account for volatility clustering GARCH models can not account for leverage effects; and the model does not allow for any direct feedback between the conditional variance and the conditional mean) and thus many extensions of the simpler GARCH model were developed.

There is non-stationary in variance when $\alpha + \beta \geq 1$, meaning that the unconditional variance is not defined. If the persistent parameters sum up to 1 there will be the case of unit root in variance also designated IGARCH which would have undesirable properties. The simpler GARCH model assumes that the volatility response to positive and negative shocks is symmetric. There is evidence that positive shocks have less impact in volatility than negative shocks of the same amount (Brooks, 2014). Glosten et al. (1993) and Nelson (1991) have developed the GJR and EGARCH models, respectively, to account for this phenomena. It is argued in finance that investors who take higher risks should be rewarded with higher returns. The GARCH-M model introduced by Engle et al. (1987) allows for the conditional variance of asset returns to enter in the conditional mean equation enabling the risk to take part in the return of an asset.

There are several other models that derive from the ARCH and GARCH process. These models will be presented in section 3.1.1 with more detail.

2.3 Which model is more adequate for the study?

The use of models that do not account for heteroskedasticity and volatility clusters is not a viable choice. The most common models used in these types of studies are the ARCH and GARCH models because of the aspects that were pointed previously.

Between ARCH and GARCH models there are some differences. Andersen and Bollerslev (1998) find that ARCH and stochastic volatility models provide good volatility forecasts. However, GARCH models tend to outperform autoregressive conditional heteroskedastic models (Akgiray, 1989) due to the fact that they are less likely to breach negativity constraints. GARCH models also tends to be more parsimonious meaning that it accomplishes a better prediction with less variables, and avoids overfitting (Brooks, 2014). Because of these, I will use GARCH family models to predict volatility.

2.4 Empirical evidence

Over the years several studies have been conducted on this subject. The main focus of these section is the research on the forecasting ability of the GARCH family models.

Pagan and Schwert (1990) studied the forecasting ability for monthly US returns of GARCH, exponential GARCH (EGARCH), Markov switching regime (MRS-GARCH) and three non-parametric models. They find that the GARCH and EGARCH models produce moderate results and the other models have a poor performance. Franses and Dijk (1996) compare GARCH, Quadratic GARCH (QGARCH) and GJR models for forecasting weekly volatility of European stock market indices. They find that non-linear GARCH models do not outperform the normal GARCH model, if the sample contains extreme observations. Brailsford and Faff (1996) state that GJR and GARCH models are superior to other models for predicting Australian monthly stock index volatility. Hansen and Lunde (2005) compare several models in order to understand if the evolution of volatility models has lead to better forecasting predictions. They find that there is conclusive evidence that GARCH (1,1) model is not outperformed by other models for the daily US return. Marcucci (2005) analyses the prediction ability of Markov switching regime and GARCH models at horizons that range from one day to one month. The author finds evidence that the MRS-GARCH significantly outperform standard GARCH models in shorter horizons. Chong et al. (1999) examines the performance of a set of GARCH models using five daily observed stock market indices in the Kuala Lumpur Stock Exchange. They found that EGARCH model outperforms the other variations of GARCH models and that the IGARCH model can not be recommended for forecasting volatility. Awartani and Corradi (2005) had similar findings for the S&P 500 index adjusted for dividends both for one-step ahead and longer forecast horizons. In accordance with that, Kişinbay (2010) found that asymmetric volatility models, including EGARCH models, provide improvements in forecasting compared to the GARCH model for forecasting at short-to-medium-term horizons.

Chapter 3

Methodology

3.1 Return

Time series of prices are the starting point of many studies in finance. As previously pointed out, in section 2.1, due to statistical reasons it is preferable not to work directly with price series so they are converted into series of returns, that is the gain or loss of a security in a particular period. Time series of prices collected was a total return index, meaning that returns will account for dividends.

Asset returns can be calculated in two different ways. There are the discrete returns that will not be discussed and the continuously compounded returns which are the ones that will be used in the research. As long as returns are relatively small (tends to happen with daily returns) continuously compounded returns and discrete returns are similar.

Let the daily return r_t be defined as follows

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

with P_t being the price of a security in time t and P_{t-1} being the price of the security at $t-1$.

3.1.1 GARCH family models

In this section the specifications of the conditional variance of several GARCH family models will be introduced. Due to a short time frame and also easier implementation of the models I will only use, for this research, a subset of models used in Hansen and Lunde (2005). I will use the following models in the research: simple GARCH model, exponential GARCH model (EGARCH), Glosten-Jagannathan-Runkle model (GJR-GARCH), integrated GARCH model (IGARCH), nonlinear asymmetric GARCH model (NGARCH), threshold GARCH model (TGARCH) and the asymmetric power GARCH (APARCH) model.

GARCH models tend to outperform ARCH models and other models that are more traditional, as referred in section 2.3. The ARCH model has some drawbacks, for example they are relatively long lagged in the conditional variance and they have to impose a fixed lag structure to avoid problems with negative variance (Bollerslev, 1986). In order to minimize those problems Bollerslev (1986) extended the ARCH class to allow for longer memory and more flexible lag structure.

As pointed out previously in section 2.2.4 simple GARCH models also present some other problems. Black (1976) found evidence that stock returns are negatively correlated with changes in returns volatility. This means that volatility tends to increase in response to bad news and decrease in response to good news. GARCH models do not account for that since they only assume that the magnitude and not the positive or negative sign of excess returns influences the conditional volatility. Another limitation is that non-negativity constraints may be violated due to the fact that volatility is not constant over time and they can create difficulties in estimating GARCH models. Engle and Bollerslev (1986) focused on studying the persistence of shocks and their impact on conditional variance. If they persist indefinitely, there is the risk that they have significant impact in long lived capital goods (Poterba and Summers, 1986). These are the three main drawbacks of simple GARCH models according to Nelson (1991), thus in order to address to the drawbacks he adopted an exponential GARCH model.

The GJR model introduced by Glosten et al. (1993) is somewhat similar to the EGARCH model in the sense that both of them embody the asymmetries in volatility in response to negative and positive shocks. Glosten et al. (1993) consider a more general specification of the GARCH-M model of Engle et al. (1987). They incorporate dummy variables in the GARCH-M model to capture seasonal effects, include the nominal interest rate in the conditional variance equation and, as pointed out previously, they allow for asymmetries in the conditional variance equation. Glosten et al. (1993) consider the EGARCH model with the modifications that they incorporated in the GARCH-M model.

The integrated GARCH was introduced by Engle and Bollerslev (1986). The model is closely related to traditional random walk with a unit root conditional mean, where the prediction of the mean s steps in the future is equal to the level today. These model focuses on the fact that current information remains important for the forecasts of conditional variance for all horizons or in other words, the model concentrates on the "persistent variance". Thus shocks to the conditional variance will not be forgotten when using the integrated GARCH.

The news impact curve relates past stock return shocks (news) to current volatility, it measures how new information is incorporated into volatility estimates (Engle and Ng, 1993). Engle and Ng (1993) referred that the news impact curve of some models, such as the EGARCH is symmetric, implying a reduced response to extreme news if γ[1] < 2 and that the news impact of the GJR is centered at $\varepsilon_{t-1} = 0$, but has different slopes for its positive and negative sides. Engle and Ng (1993) introduced the NGARCH model, that allows either the slope of two sides of the news impact curve to differ or the center of the news impact curve to locate at a point where ε_{t-1} is positive.

In order to account for the asymmetry of returns (conditional variance tends to be higher after a decrease in return than after an increase of the same amount) Zakoian (1994) introduced the TGARCH. This model is closely related to the GJR model (Wu, 2010). It is a counterpart of the GJR model in the case where the entity to be modelled is the conditional standard deviation instead of the conditional variance (Teräsvirta, 2009).

There is the possibility that a squared power term may not be necessarily optimal (McKenzie and Mitchell, 2002). Ding et al. (1993) stated that there is no reason to assume that the conditional variance is a linear function of lagged squared returns/residuals or the conditional standard deviation a linear function of lagged absolute returns/residuals. Thus Ding et al. (1993) introduced a new class of ARCH models denominated power ARCH model. This model estimates the optimal power term instead of imposing a structure on the data, allowing an infinite range of transformations (McKenzie and Mitchell, 2002). The asymmetric power ARCH was introduced to account for the asymmetry response of volatility to positive and negative shocks (Ding et al., 1993).

Table 3.1 introduce the conditional mean specifications and table 3.2 present the condi-

[1] γ is a constant parameter of the EGARCH model

tional variance specifications of the models used in the research.

Table 3.1: Conditional mean [1]

Non-zero constant mean	$\mu_t = \mu_0$

[1]Conditional mean equation was obtained from Hansen and Lunde (2005)

Table 3.2: Alternative GARCH models conditional variance [1]

GARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
EGARCH	$\log(\sigma_t^2) = \omega + \sum_{i=1}^{q} [\alpha_i e_{t-i} + \gamma_i(\lvert e_{t-i}\rvert - E\lvert e_{t-i}\rvert)] + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2)$
GJR	$\sigma_t^2 = \omega + \sum_{i=1}^{q} [\alpha_i + \gamma_i I_{(\varepsilon_{t-i}>0)}]\varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$
IGARCH	$\sigma_t^2 = \omega + \varepsilon_{t-1}^2 + \sum_{i=2}^{q} (\varepsilon_{t-i}^2 - \varepsilon_{t-1}^2) + \sum_{j=1}^{p} \beta_j (\sigma_{t-j}^2 - \varepsilon_{t-1}^2)$
NGARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (\varepsilon_{t-i} + \gamma_i \sigma_{t-i})^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
TGARCH	$\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i [(1-\gamma_i)\varepsilon_{t-i}^{+} - (1+\gamma_i)\varepsilon_{t-i}^{-}] + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$
APARCH	$\sigma^\delta = \omega + \sum_{i=1}^{q} \alpha_i [\lvert\varepsilon_{t-i}\rvert - \gamma_i \varepsilon_{t-i}]^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta$

[1]Conditional variance equations were gathered from Hansen and Lunde (2005)

Conditional volatility estimation

One erroneous assumption is that the parameters of the model are constant over time. The economic environment changes over time. Thus it is not plausible to assume that the parameters of the model are constant (Zivot and Wang, 2006). Granger (1996) stated that structural changes are one of the most important problems that researchers have to face. In order to deal with structural changes and changes of parameters over time conditional GARCH volatility was estimated using a fixed rolling window estimation. The rolling estimators are based on a changing sub-sample of fixed length that moves sequentially from the beginning of the sample to the end (Xiao-lin et al., 2016). In each step we determine the conditional daily volatility. Being l the number of fixed steps of the rolling window, the full sample is converted into a sequence of $T - l$ sub-samples, or $\tau - l + 1, \tau - 1, \ldots, T$, for $\tau = l + 1, \ldots, T$.

I have opted to use the daily frequency to measure returns and three different estimation windows: 500 days, 1000 days and 2000 days. The use of three different estimation windows will allow me to account for possible structural changes in the data and assess their impact and also to try to understand which of them is the optimal sample size to forecast next day volatility.

Since there are three different estimation windows for the GARCH models it is not reasonable to study the residuals for the whole period, meaning that they would be different for the three estimation periods. The GARCH models were estimated assuming two distributions, normal and t-Student.

3.2 Forecast Evaluation

3.2.1 Realized Variance

Realized volatility measures what actually happened in the past. The realized variance (RV), which is the sum of squared intraday returns is a perfect estimate of volatility in the ideal situation where prices are observed continuously and without measurement error (Hansen and Lunde, 2006). Given continuously observed quote data (with no transaction costs) the realized variance may be measured without error along with the realized return (Andersen and Benzoni, 2008). This suggests that volatility during a period of time can be estimated in a more precise way if the frequency of returns increases, provided that intraperiod returns are uncorrelated and certain other conditions apply (Taylor, 2005).

Realized variance data was collected form Oxford-Man Institute Realized Library (see section 4). The database contains daily financial returns $r_1, r_2, \ldots, r_T$ and a corresponding sequence of daily realized measures $RM_1, RM_2, \ldots, RM_T$.

Following the approach proposed by Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) the realized measure is:

$$RM_t = \sum_{j=0}^{n} x_{j,t}^2$$

where $x_{j,t}$= $X_{tj,t} - X_{tj-1,t}$ and $t_{j,t}$ are the times of trades or quotes on the t-th day and X is the asset traded or quoted. This is justified by the fact that if prices are observed without noise then as $min_j|t_{j,t} - t_{j-1,t}| \downarrow 0$ estimates the quadratic variation of prices process on t-th day.

Realized volatility is the square root of the realized variance. Thus, $\sigma = \sqrt{RM_t}$.

3.2.2 Univariate and Multivariate regressions

Several studies use univariate and multivariate regressions. To test the forecast ability of different models the approach will be similar to the one used in Christensen and Prabhala (1998), Corrado and Miller (2005) and Jiang and Tian (2005). However in the mentioned studies the authors try to predict realized volatility using different methods such as implied volatility and in the case of this study the objective is to identify which GARCH model yields better results in forecasting the next day volatility.

One way to evaluate the volatility models is through the R^2 from a Mincer-Zarnowitz (MZ) regression

$$r_t^2 = a + bh_t^2 + u_t$$

or a logarithmic version

$$\log(r_t^2) = a + b\log(h_t^2) + u_t$$

Both of these models are used in Pagan and Schwert (1990). In the first case there are some major drawbacks. The regression will be very inefficient and will have misleading standard errors, due to the fact that r^2 is heteroskedastic. Secondly, the r^2 is a noisy estimate of the volatility to be estimated. This leads to a very low R^2. This method is also not reliable because it measures the level of variance errors instead of the more realistic proportional errors (Engle and Patton, 2001). The use of this methods is also not recommended due to the fact that it does not penalize a biased forecast leading to a higher R^2 that is not true (Hansen and Lunde, 2005). Andersen and Bollerslev (1998) implied that the use of alternative measures of volatility should ease this problem. In order to mitigate these drawbacks instead of using r^2 as a benchmark I will use the realized volatility which is a better estimate of volatility in comparison to the one used in the original MZ regression thus the equation becomes

$$\sigma_t^2 = a + bh_t^2 + u_t$$

Using linear regressions allows to test three hypothesis: First if conditional volatility contains some information about future volatility, second if conditional volatility is an unbiased forecast of realized volatility and if so one should find that $a = 0$ and $b = 1$ and ultimately if conditional volatility is efficient where u_t should be white noise and uncorrelated with any variable (Christensen and Prabhala, 1998).

The coefficient of determination is defined as the proportion of variance explained by the regression model makes it useful as a measure of success of predicting the dependent variable from the independent variable (Nagelkerke, 1991).

Evaluating volatility using loss functions

There are other ways that are more reliable to evaluate volatility models. Hansen and Lunde (2005) used six different loss functions in their paper. The functions were similar, however in half of them they used variance instead of standard deviation, these may result in some problems since small errors in variance may lead to a very large loss function. I use the same loss functions as the ones used by Hansen and Lunde (2005):

$$MSE_1 \equiv n^{-1}\sum_{t=1}^{n}(\sigma_t - h_t)^2 \qquad MSE_2 \equiv n^{-1}\sum_{t=1}^{n}(\sigma_t^2 - h_t^2)^2$$

$$QLIKE \equiv n^{-1}\sum_{t=1}^{n}(log(h_t^2) + \sigma_t^2 h_t^{-2}) \qquad R^2LOG \equiv n^{-1}\sum_{t=1}^{n}[log(\sigma_t^2 h_t^{-2})]$$

$$MAE_1 \equiv n^{-1}\sum_{t=1}^{n}|\sigma_t - h_t| \qquad MAE_2 \equiv n^{-1}\sum_{t=1}^{n}|\sigma_t^2 - h_t^2|$$

MSE_1 was employed by Allen (1971) and measures the difference between the estimator and what is estimated corresponding to the expected value of the squared error loss. MSE_2, $QLIKE$ and R^2LOG were proposed by Bollerslev et al. (1994). MSE_2 will penalize conditional variance estimates which are different from the realized squared residuals, regardless this loss function does not penalize the method for negative or zero variance (Bollerslev et al., 1994). Other options to investigate this issue are the $QLIKE$ corresponding to the percentage absolute errors or the loss function implied by the Gaussian likelihood and the R^2LOG penalises volatility forecasts asymmetrically in low volatility and high volatility periods (Bollerslev et al., 1994; Hansen and Lunde, 2005; Lorde and Moore, 2008). The other loss functions MAE_1 and MAE_2 are more robust to outliers than the others used in the research (Hansen and Lunde, 2005).

3.2.3 Test of superior predictive ability (SPA)

Testing for superior predictive ability is useful for a forecaster in order to identify which, if any, forecasting models is better than the model that the forecaster is currently using. The goal is to understand if there are any alternatives significantly more accurate than the benchmark.

It is obvious that GARCH family models should outperform more basic models as said previously. In order to identify the best model a test of superior predictive ability is needed. The reality check (RC) test was introduced by White (2000) with the intent to provide a method that avoided data snooping obtained by testing if the model obtained had no predictive superiority over a given benchmark model, avoiding than the "pure luck" forecasts. Hansen (2005) developed a test for SPA. The test is similar to the RC test, however in this test a different test statistic is used and a sample-dependent distribution is used under the null hypothesis. The RC test is "sensitive to the inclusion of poor and irrelevant models" and "less powerful" (Hansen and Lunde, 2005, p. 880). Despite that, it is not possible to ensure that it is an optimal test. I will use the SPA test proposed by Hansen (2005). The following description of the method closely follows the presentations of the method given by Hansen (2005).

Let $\delta_{k,t-h}, k = 0, 1, ..., m$, be different point forecasts of a real random variable ξ. The first model, $k = 0$ is referred to as the benchmark. Decisions are evaluated with a loss function, $L(\xi_t, \delta_{k,t-h})$ where ξ_t is a random variable that represents the aspects of the decision problem that are unknown at the time the decision is made. Forecasts are evaluated in terms of their expected loss.

Define $d_{k,t}$ which is the performance of model k relative to other benchmark at time t

$$d_{k,t} \equiv L(\xi_j, \delta_{0,t-h}) - L(\xi_j, \delta_{k,h-k}),\ \ k = 1, ..., m$$

stacking these variables in the vector of relative performances $\mathbf{d}_t = (d_{1,t}, ..., d_{m,t})'$. Provided that $\mu \equiv E(\mathbf{d}_t)$ is well defined, we can formulate the null hypothesis as $H_0 : \mu \leq 0$.

The main assumption is that the model k is better than the benchmark model if and only if $E(d_{k,t} > 0)$, therefore $\mathbf{d}_t, t = 1, \ldots, n$ is viewed as the data, thus one can state all assumptions in terms of $\mathbf{d}_t$.

Assumption 1. The vector of relative loss variables, $\{\mathbf{d}_t\}$, is strictly stationary and α-mixing of size $-(2+\delta)(r+\delta)/(r-2)$, for some $r > 2$ and $\delta > 0$, where $E|\mathbf{d}_t|^{r+\delta} < \infty$ and $\text{var}(d_{k,t}) > 0$ for all $k = 1, \ldots, m$. This assumption ensures that certain population moments such as μ are well defined.

The SPA test is based on the test statistic

$$T_n^{SPA} \equiv \max\left[\max_{k=1,\ldots,m} \frac{n^{1/2}\bar{d}_k}{\hat{\omega}_k}, 0\right] \tag{3.1}$$

where $\hat{\omega}_k^2$ is some consistent estimator of $\omega_k^2 \equiv \text{var}(n^{1/2}\bar{d}_k)$. It invokes a null distribution that is based on $N_m(\hat{\mu}^c, \hat{\Omega})$ where $\hat{\mu}^c$ is a carefully chosen estimator for μ that conforms with the null hypothesis.

$$\hat{\mu}_k^c = \bar{d}_k \psi_{\{n^{1/2}\bar{d}_k/\hat{\omega}_k \leq -\sqrt{2\log\log n}\}},\quad k = 1, ..., m,$$

where $\psi_{\{.\}}$ denotes the indicator function, see Hansen (2005) for details.

The stationary bootstrap of Politis and Romano (1994) is used to estimate $\hat{\mu}_k^c$. This process is based on pseudo-time series of the original data. The pseudo-time series $\{d_{b,t}^*\} \equiv \{d_{\tau b,t}\}$, $b = 1, ..., B$, are samples of d_t, where $\{\tau_{b,1}, ..., \tau_{b,n}\}$ is constructed by combining blocks of $\{1, ..., n\}$ with random lengths. From that, we calculate their sample averages, $\bar{d}_n^* \equiv n^{-1}\sum_{t-1}^{n} d_{b,t}^*$, $b = 1, ..., B$, that can be viewed as independent draws from the distribution of $\bar{d}$, under the bootstrap distribution.

From Goncalves and Jong (2003) its possible to say that the empirical distribution of the pseudo-time series can be used to approximate the distribution of $n^{1/2}(\bar{d}-\mu)$. The test statistic T_n^{SPA} requires estimates of $\hat{\omega}_k^2,\ k=1,...,m$

$$\hat{\omega}_{k,B}^{*2} \equiv B^{-1}\sum_{b=1}^{B}(n^{1/2}\bar{d}_{k,b}^{*}-n^{1/2}\bar{d}_k)^2$$

where $\bar{d}_{k,b}^{*}=n^{-1}\sum_{t=1}^{n}d_{k,\tau_{b,t}}$.

The distribution of the null hypothesis is imposed by recentering the bootstrap variables about $\hat{\mu}^l,\hat{\mu}^c,\text{or}\hat{\mu}^u$. Let assumption 1 hold and $Z_{b,t}^{*}$ be centred about $\hat{\mu}$, for $\hat{\mu}=\hat{\mu}^l,\hat{\mu}^c,\text{or}\hat{\mu}^u$, see Hansen (2005) for details. Then

$$\sup_{z\in\mathbb{R}^m}\left|P^{*}\left(n^{1/2}(\bar{Z}_b^{*}-\mu)\leq z\right)-P\left(n^{1/2}\bar{d}-\mu\right)\leq z\right|\rightarrow^{\mathrm{p}}0,$$

where $\bar{Z}_{k,b}^{*}=n^{-1}\sum_{t=1}^{n}Z_{k,b,t}^{*},\ k=1,...,m.$

The distribution of the test statistic under the null hypothesis can be approximated by the empirical distribution obtained from the bootstrap re-samples $Z_{b,t}^{*},\ t=1,...,n.$ $T_{b,n}^{SPA*}=\max\{0,\max_{k=1,...,m}[n^{1/2}\bar{Z}_{k,b}^{*}/\hat{\omega}_k]\}$ for $b=1,...,B$, and the bootstrap p value is given by

$$\hat{p}\text{SPA}\equiv\sum_{b=1}^{B}\frac{\psi\{T_{b,n}^{SPA*}>T_n^{SPA}\}}{B}$$

where the null hypothesis should be rejected for small p values.

Chapter 4

Data

This research is focused on the US market, more specifically on the NASDAQ-100 stock market index. The data was collected from Datastream. To ensure that results of the research are robust I use a large time period in order to allow a large forecasting period. The period for the analysis is of 30 years from 01/01/1986 to 21/11/2016, totaling 7789 observations. The original data set consists of quotations of NASDAQ-100 index. The sample mean is 1549.5, the median is 1408.5, the minimum is 128.4 and the maximum is 4910.0. The realized volatility was obtained from the Oxford-Man Institute Realized Library[1] in February of 2016. Returns were computed from the downloaded price series, as previously described in Chapter 3. The conditional volatility was estimated using three different estimation windows: 500 days, 1000 days and 2000 days, in order to mitigate the risk of structural changes and to account for non-constant parameters.

Table 4.1: NASDAQ-100 returns

No. observations	7788
Minimum	-0.1508
Maximum	0.1877
Mean	0.0006
Std. Deviation	0.0169
Skewness	0.1447
Kurtosis	7.8716

This table presents the discriptive statistics form the NASDAQ-100 index returns

[1]http://realized.oxford-man.ox.ac.uk/data/download

4.1 Data analysis

Price movements are relatively slow when conditions are calm. Otherwise they move faster in the presence of uncertainty, increase of trading and the arrival of new information to the market. Figure 4.1 represents the variations in price from 1985 to 2016. In the figure it is possible to observe that prices are rising steadily in the first decade. Nevertheless, its not possible to observe the 1987 "Black Monday" crash, due to the graphic scale. In the next five years they suffer an abrupt increase and a sharp drop after that in what became known as the "dot.com" burst. Between 2003 and mid 2008 prices were rising again, however another financial crisis caused a negative impact on the market leading to the second sharp drop in under a decade. After 2010 prices began rising abruptly again until up to mid 2016.

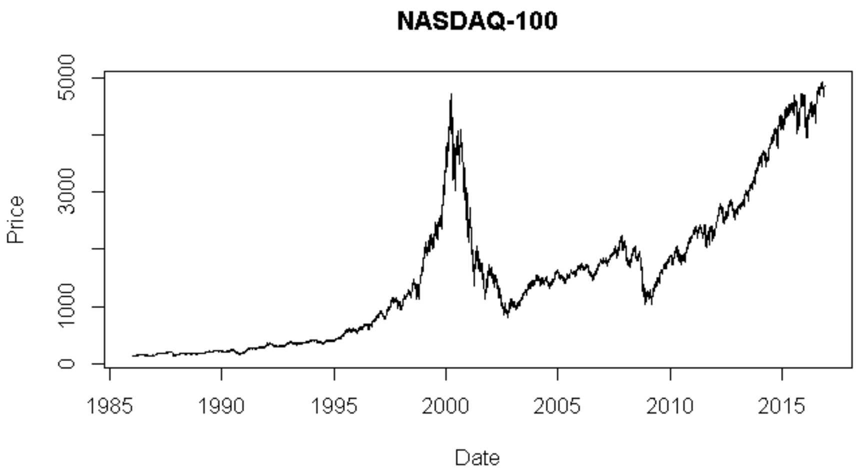


Figure 4.1: NASDAQ price

Figure 4.2 represents the conditional volatility estimated by a GARCH(1,1) model with assumed normal distribution. It is possible to observe that periods of crisis referred previously lead to an increase in the conditional volatility and periods that are relatively normal lead to lower conditional volatility.

Analysing market prices poses more challenges than analysing changes in prices. To investors, returns are the preferred way to measure price changes. Large returns (of any sign) are followed by large returns (of any sign) and vice-versa. In figure 4.3 it is possible to observe that inference. It is possible to see that between 1995 and 2004 there is a cluster

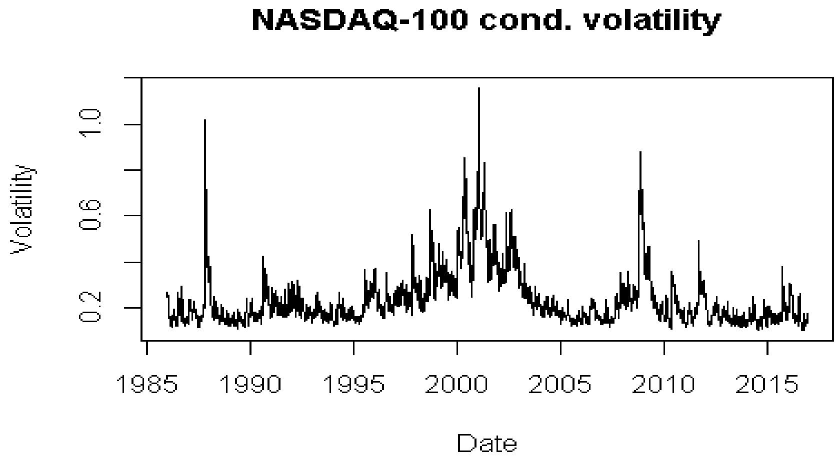


Figure 4.2: NASDAQ conditional volatility

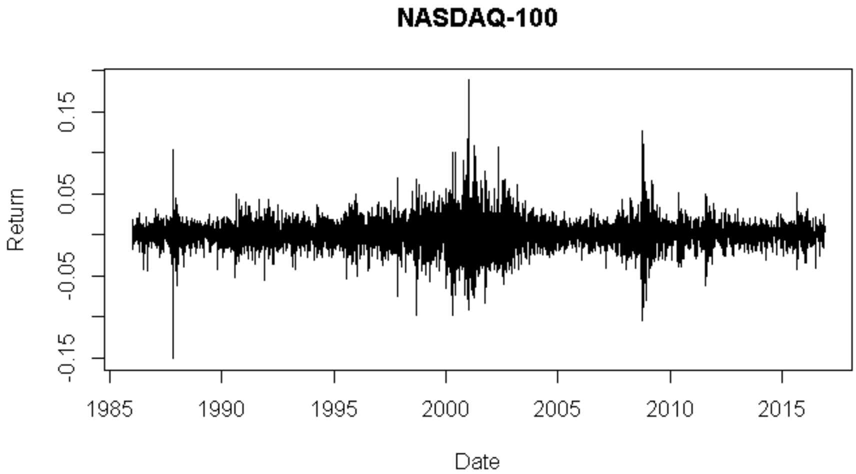
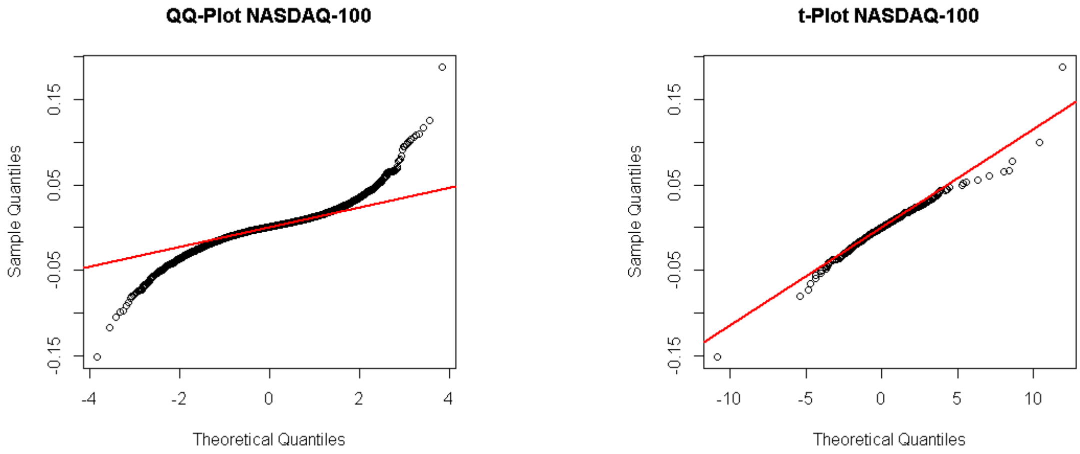


Figure 4.3: NASDAQ-100 daily returns

of large returns and between 2010 and 2015 there is a cluster of smaller returns.

As referred in previous chapters, returns have three stylized facts. Returns do not follow a normal distribution, there is almost no correlation between returns for different days and there is positive dependence between absolute returns on nearby days.

Figure 4.4a represents a quantile-quantile plot, a graphical method for comparing two probability distributions by plotting their quantiles against each other, of NASDAQ-100 returns. It is possible to infer that returns do not follow a normal distribution. In fact we can say that they are heavy tailed or have fat tails as described in section 2.1. Analysing figure



(a) Q-Q plot normal distribution (b) Q-Q plot t-Student distribution

Figure 4.4: Quantile Quantile plot of NASDAQ-100 returns

4.4 it is possible to say that returns have a distribution closer to the t-Student distribution. To ensure that returns do not follow a normal distribution more tests were conducted. The Jarque-Bera test jointly tests if the skewness coefficient is zero and if the kurtosis coefficient is three (matching the normal distribution value for those moments). The test p-vaule is $2.2e^{-16}$, thus rejecting the null hypothesis of normality.

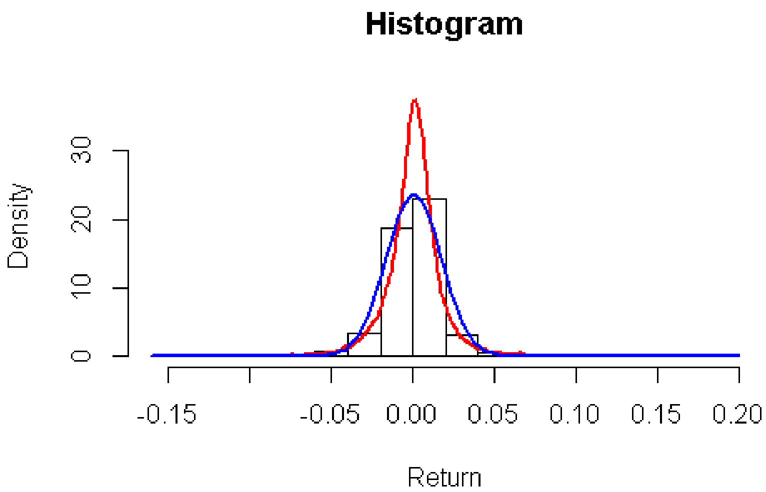
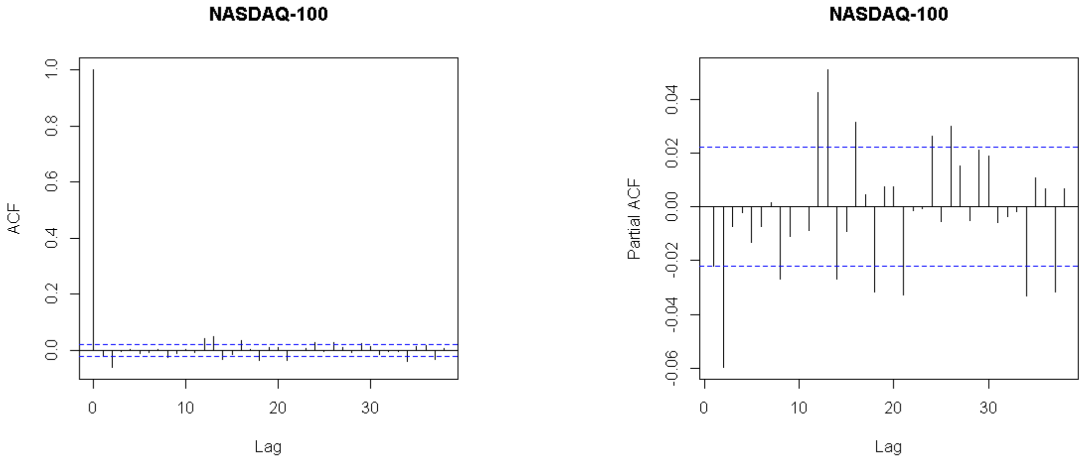


Figure 4.5: NASDAQ Histogram

Returns are expected to be skewed to the left and to have a high peak as explained in section 2.1. Figure 4.5 is an histogram of NASDAQ-100 returns represented in red, whereas the blue line represents the normal distribution. It is possible to see that returns have a high peak, however in this case the skewness is positive and thus they are slightly skewed to the right. It is also known that returns have high kurtosis. This means that the returns are fat tailed in the positive end and the probability of achieving these positive values is higher than the probability assumed by the normal distribution.



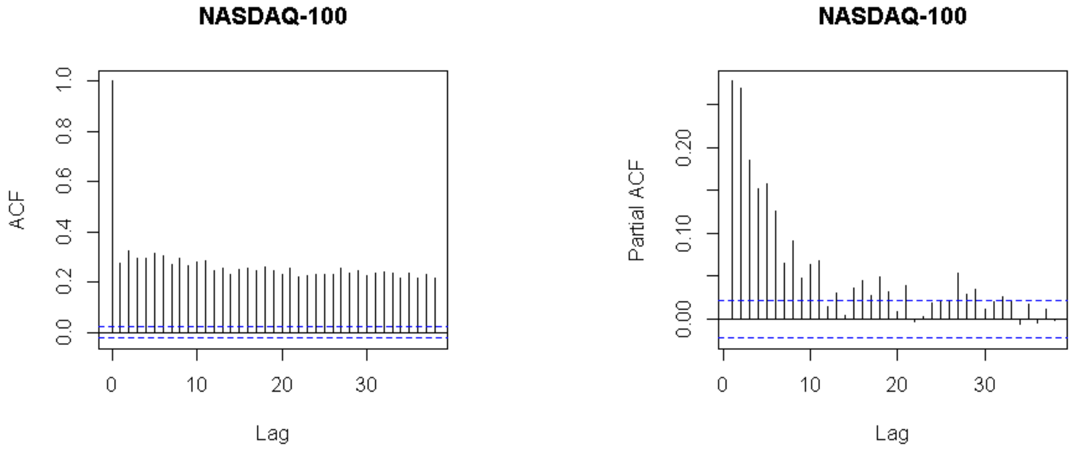
(a) (b)

Figure 4.6: ACF and PACF of NASDAQ-100 returns

The second stylized fact of returns is that there is almost no correlation between returns for different days. In figure 4.6a we can see the autocorrelation function of the NASDAQ-100 returns. Overall the returns do not exhibit strong interdependency, although we can see there are some lags that are above or bellow the significant range, meaning that in those lags the correlation coefficient is statistically significant. This beg the following question: do returns exhibit no serial correlation? In order to answer this question a Box-Pierce test (Box and Pierce, 1970) for the first 40 lags with the null hypothesis of independence of returns was conducted. The X-squared was 170,37, and test p-value was $2,2e^{-16}$. This means that we reject the null hypothesis and therefore returns exhibit serial correlation.

Figure 4.6b represents the partial autocorrelation function of returns. It is possible to observe that many lags are above or below the significance range.

Figure 4.7 represents the ACF and PACF of absolute returns. The PACF shows a significant autocorrelation up to lag 35. Analysing the ACF we can see that there is positive dependence between absolute returns on nearby days thus the third stylized fact of returns



(a) (b)

Figure 4.7: ACF and PACF of absolute NASDAQ-100 returns

is also true.

Daily return time series distribution possesses fat tails, as described above. This may happen if the squared returns are correlated or, in other words, if returns have conditional heteroskedasticity. To test for the ARCH effect I performed an Lagrange Multiplier test for autoregressive conditional heteroskedasticity (ARCH) introduced by Engle (1982). The null hypothesis is: there are no ARCH effects. The Chi-squared is equal to 1344,2 and the test p-value is $2, 2e^{-16}$. This means that we reject the null hypothesis, thus returns are conditionally heteroskedastic. It is not possible to test for the GARCH effect (Bollerslev, 1986).

Chapter 5

Results

5.1 Measures of forecasting accuracy

5.1.1 Linear regression analysis

Correlation is a single number that describes the degree of relation between two variables. It is than used to investigate the degree of dependence between the variables. The diagonal line of a correlation matrix is always one due to the fact that it is the correlation between each variable and itself. The correlation of variable x and variable y is always equal to the correlation of the variable y and variable x, this means that the upper triangle are always a mirror image of the lower triangle. It is possible to see the correlations in the upper triangle of the matrix.

Table 5.1 shows the correlations between the various estimation of the different models used in this research. The range of the correlation coefficient is from -1 to 1. It is possible to see that all GARCH family model estimations are highly and positively correlated for the three estimation windows (500, 1000 and 2000 days). Nevertheless, there is reason to believe that their estimations will not be the same, thus it is still possible to test if any model performs better than the other models.

The main purpose of this research is to be able to identify which model has the best forecasting ability. A common measure of the quality of the forecasts is the Mincer and Zarnowitz (1969) method, as previously described in section 3.2.2.

The linear models were estimated using a univariate regression where the dependent variable was the realized volatility and the independent variable was the GARCH family model estimations for the three different periods.

Table 5.2 shows the results of the univariate linear regression for the GARCH(1,1) model with both specifications of the distribution (normal distribution and t-Student distribution).

Table 5.1: Correlation matrix of GARCH family models

	Panel A													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	1	0.996	0.972	0.970	0.981	0.983	0.999	0.994	0.967	0.970	0.963	0.962	0.970	0.969
(2)		1	0.964	0.966	0.968	0.974	0.998	0.999	0.954	0.959	0.952	0.955	0.959	0.962
(3)			1	0.999	0.984	0.986	0.969	0.962	0.989	0.992	0.998	0.998	0.998	0.999
(4)				1	0.979	0.983	0.968	0.964	0.986	0.990	0.996	0.998	0.995	0.998
(5)					1	0.999	0.976	0.964	0.987	0.987	0.980	0.976	0.989	0.985
(6)						1	0.979	0.971	0.987	0.988	0.981	0.978	0.989	0.987
(7)							1	0.997	0.961	0.964	0.959	0.959	0.966	0.966
(8)								1	0.950	0.955	0.949	0.952	0.955	0.959
(9)									1	0.997	0.989	0.987	0.993	0.991
(10)										1	0.992	0.991	0.995	0.994
(11)											1	0.999	0.998	0.998
(12)												1	0.996	0.999
(13)													1	0.999
(14)														1
	Panel B													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	1	0.997	0.969	0.967	0.977	0.978	0.999	0.996	0.963	0.963	0.960	0.960	0.967	0.967
(2)		1	0.963	0.964	0.969	0.972	0.998	0.999	0.954	0.956	0.950	0.954	0.958	0.961
(3)			1	0.999	0.983	0.984	0.967	0.962	0.983	0.988	0.998	0.998	0.997	0.998
(4)				1	0.980	0.983	0.966	0.963	0.983	0.988	0.995	0.998	0.995	0.998
(5)					1	0.999	0.973	0.966	0.981	0.985	0.979	0.977	0.989	0.987
(6)						1	0.974	0.969	0.981	0.985	0.979	0.979	0.988	0.989
(7)							1	0.999	0.958	0.959	0.956	0.957	0.964	0.964
(8)								1	0.951	0.953	0.948	0.952	0.956	0.959
(9)									1	0.994	0.984	0.984	0.987	0.988
(10)										1	0.988	0.990	0.992	0.993
(11)											1	0.999	0.998	0.997
(12)												1	0.997	0.998
(13)													1	0.999
(14)														1
	Panel C													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	1	0.999	0.981	0.981	0.984	0.986	0.999	0.996	0.976	0.980	0.978	0.978	0.982	0.983
(2)		1	0.975	0.979	0.977	0.982	0.998	0.999	0.970	0.975	0.973	0.977	0.977	0.981
(3)			1	0.998	0.991	0.992	0.980	0.974	0.988	0.993	0.999	0.997	0.998	0.997
(4)				1	0.988	0.991	0.981	0.979	0.987	0.993	0.998	0.999	0.997	0.999
(5)					1	0.999	0.982	0.976	0.989	0.992	0.989	0.985	0.995	0.992
(6)						1	0.985	0.981	0.988	0.993	0.990	0.989	0.995	0.995
(7)							1	0.998	0.975	0.978	0.977	0.978	0.981	0.982
(8)								1	0.968	0.974	0.972	0.977	0.975	0.980
(9)									1	0.994	0.989	0.987	0.992	0.990
(10)										1	0.994	0.993	0.996	0.996
(11)											1	0.998	0.998	0.998
(12)												1	0.996	0.999
(13)													1	0.998
(14)														1

Note: This table shows the correlation between the GARCH family models. Correlation coefficient was estimated using the Pearson method. The table is divided in three parts. Panel A shows the correlation for the models estimated using a 500 day estimation window. Panel B shows the correlation for the models estimated using a 1000 day estimation window. Panel C shows the correlation for the models estimated using a 2000 day estimation window. Correlation coefficients were estimated using the Pearson method.

Model description: (1) GARCH(1,1)-norm; (2) GARCH(1,1)-std; (3) EGARCH-norm; (4) EGARCH-std; (5) GJR-norm; (6) GJR-std; (7) IGARCH-norm; (8) IGARCH-std; (9) NGARCH-norm; (10) NGARCH-std; (11) TGARCH-norm; (12) TGARCH-std; (13) APARCH-norm; (14) APARCH-std.

Table 5.2: Univariate linear regressions

Panel A				
	GARCH(1,1)-norm		GARCH(1,1)-std	
R^2	0.3301		0.3118	
	Intercept	Sigma	Intercept	Sigma
Estimate	-0.00035	0.64829	-0.00022	0.63975
Std. Error	0.00047	0.04126	0.00048	0.04245
t-value	-0.742	15.713	-0.448	15.071
Prob(>(|t|)	0.458	$<2e^{-16}$***	0.655	$<2e^{-16}$***
Panel B				
	GARCH(1,1)-norm		GARCH(1,1)-std	
R^2	0.2981		0.2863	
	Intercept	Sigma	Intercept	Sigma
Estimate	-0.00022	0.63843	$2.507e^{-05}$	$6.221e^{-01}$
Std. Error	0.00032	0.03097	$3.218e^{-04}$	$3.105e^{-02}$
t-value	-0.673	20.612	0.078	20.034
Prob(>(|t|)	0.501	$<2e^{-16}$***	0.938	$<2e^{-16}$***
Panel C				
	GARCH(1,1)-norm		GARCH(1,1)-std	
R^2	0.5127		0.5029	
	Intercept	Sigma	Intercept	Sigma
Estimate	-0.00013	0.65332	0.00031	0.61679
Std. Error	0.00018	0.01425	0.00018	0.01372
t-value	-0.708	45.851	1.733	44.959
Prob(>(|t|)	0.479	$<2e^{-16}$***	0.0832’	$<2e^{-16}$***

Note: This table presents the results of linear regression using realized volatility as the dependent variable. It is divided in two parts. On the left the linear regression results using the GARCH(1,1) model with normal distribution estimations is shown and on the right the GARCH(1,1) model with t-Student distribution estimates is displayed. The table is divided in three panels. Panel A corresponds to the linear regression using the GARCH estimation using a 500 day estimation window, panel B using an estimation window of 1000 and panel C using a 2000 day estimation window.
’ indicates significance at the 10% level.
*** indicates significance at the 0.1% level.

It is possible to see that when the estimation window of the GARCH model increases from 500 days to 1000 days, the R^2 of the regression suffers a small decrease of approximately 3%. However, when the estimation window used in the estimation of the GARCH model is 2000 days it is possible to see that the R^2 of the linear regression model increases to over 50%. All the estimated "Sigma" coefficients are positive and statistically significant at 1% and none of the intercepts is statistically significant for a significance level of 5%. This means that the realized volatility is equal to x times sigma, being x the value of the model conditional volatility.

Tables 5.3, 5.4 and 5.5 exhibit the results of the linear regression for the remaining GARCH family models used in this research estimated using a 500, 1000 and 2000 estimation window respectively. Analysing each table separately it is possible to see that the R^2 for all GARCH family model are alike, meaning that all GARCH models have similar R^2 values, with the exception of the IGARCH model that in all cases has a lower R^2. When the estimation window used to estimate the conditional volatility with GARCH models increases from 500 to 1000 days the R^2 of the linear regressions decreases slightly, nevertheless when the estimation window used in the estimation of conditional volatility with GARCH models increases to 2000 days the R^2 of the regressions increases in accordance to what happened in table 5.2. Overall the R^2 of the linear regression using more complex GARCH model estimations of conditional volatility as the independent variable is higher than when using the GARCH(1,1) estimates of conditional volatility. All estimated "Sigma" coefficients are positive and statistically significant for all the regressions. When using a 500 and 1000 day estimation period to estimate conditional volatility with GARCH models there are some linear regressions that present an intercept statistically significant, although when increasing the estimation window of the GARCH models to 2000 days none of the linear regression intercepts are statistically significant.

It is possible to say that there is a big part of the variation of the realized volatility is not explained by the conditional volatility estimated with GARCH model, since the determination coefficients are not close to 1. These results are expected due to the drawbacks sated previously in the literature review, which include the inefficiency of the regression and misleading standard errors due to the heteroskedasticity of the regression model, and the fact that the model does not penalize biased forecasts.

Table 5.3: Univariate linear regressions (500 day estimation period)

	EGARCH-norm		EGARCH-std			GJR-norm		GJR-std	
R^2	0.403		0.3995		R^2	0.4019		0.3961	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	-0.00062	0.671134	-0.00049	0.66201	Estimate	-0.00069	0.66674	-0.00046	0.64862
Std. Error	0.00042	0.03651	0.00042	0.03628	Std. Error	0.00043	0.03636	0.00042	0.03580
t-value	-1.478	18.382	-1.163	18.249	t-value	-1.617	18.337	-1.097	18.120
Prob(>(|t|)	0.14	$<2e^{-16}$***	0.246	$<2e^{-16}$***	Prob(>(|t|)	0.107	$<2e^{-16}$***	0.273	$<2e^{-16}$***

	IGARCH-norm		IGARCH-std			NGARCH-norm		NGARCH-std	
R^2	0.3176		0.3043		R^2	0.4402		0.4356	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	0.00017	0.59571	0.00011	0.60539	Estimate	-0.00099	0.69866	-0.00087	0.68739
Std. Error	0.00045	0.03900	0.00047	0.04089	Std. Error	0.00041	0.03522	0.00041	0.03498
t-value	0.374	15.273	0.231	14.807	t-value	-2.408	19.835	-2.138	19.649
Prob(>(|t|)	0.708	$<2e^{-16}$***	0.817	$<2e^{-16}$***	Prob(>(|t|)	0.0164*	$<2e^{-16}$***	0.033*	$<2e^{-16}$***

	TGARCH-norm		TGARCH-std			APARCH-norm		APARCH-std	
R^2	0.4153		0.4111		R^2	0.416		0.4119	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	-0.00090	0.69718	-0.00074	0.68614	Estimate	-0.00089	0.69341	-0.00070	0.68036
Std. Error	0.00043	0.03698	0.00042	0.03671	Std. Error	0.00043	0.03673	0.00042	0.03634
t-value	-2.115	18.854	-1.763	18.689	t-value	-2.088	18.880	-1.681	18.723
Prob(>(|t|)	0.0349*	$<2e^{-16}$***	0.0785'	$<2e^{-16}$***	Prob(>(|t|)	0.0373*	$<2e^{-16}$***	0.0934'	$<2e^{-16}$***

Note: This table presents the results of the linear regressions using realized volatility as the dependent variable and the GARCH model estimations (using a 500 day estimation rolling window) as the independent variable. On the left side the linear regression results using the conditional volatility of the GARCH models estimated with a normal distribution and on the right side the linear results using the conditional volatility of the GARCH models estimated with a t-Student distribution.
* indicates significance at the 5% level.

Table 5.4: Univariate linear regressions (1000 day estimation period)

	EGARCH-norm		EGARCH-std			GJR-norm		GJR-std	
R^2	0.3727		0.3741		R^2	0.3755		0.3734	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	-0.00014	0.63696	$6.215e^{-05}$	$6.235e^{-01}$	Estimate	-0.00060	0.66334	-0.00027	0.63801
Std. Error	0.00027	0.02613	$2.649e^{-04}$	$2.551e^{-02}$	Std. Error	0.00029	0.02706	0.00028	0.02614
t-value	-0.505	24.372	0.235	24.444	t-value	-2.054	24.518	-0.964	24.409
Prob(>(|t|)	0.614	$<2e^{-16}$***	0.815	$<2e^{-16}$***	Prob(>(|t|)	0.0402*	$<2e^{-16}$***	0.335	$<2e^{-16}$***

	IGARCH-norm		IGARCH-std			NGARCH-norm		NGARCH-std	
R^2	0.2875		0.2795		R^2	0.3923		0.4091	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	0.00033	0.58432	0.00034	0.59004	Estimate	-0.00057	0.66641	-0.00047	0.66216
Std. Error	0.00031	0.02908	0.00031	0.02995	Std. Error	0.00028	0.02623	0.00027	0.02517
t-value	1.087	20.093	1.088	19.701	t-value	-2.028	25.404	-1.76	26.30
Prob(>(|t|)	0.277	$<2e^{-16}$***	0.277	$<2e^{-16}$***	Prob(>(|t|)	0.0428*	$<2e^{-16}$***	0.0788’	$<2e^{-16}$***

	TGARCH-norm		TGARCH-std			APARCH-norm		APARCH-std	
R^2	0.3873		0.3875		R^2	0.3903		0.39	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	-0.00051	0.66932	-0.00029	0.65403	Estimate	-0.00060	0.67396	-0.00031	0.65336
Std. Error	0.00028	0.02662	0.00027	0.02601	Std. Error	0.00028	0.02664	0.00027	0.02584
t-value	-1.829	25.137	-1.064	25.148	t-value	-2.099	25.296	-1.141	25.280
Prob(>(|t|)	0.0677’	$<2e^{-16}$***	0.288	$<2e^{-16}$***	Prob(>(|t|)	0.0361*	$<2e^{-16}$***	0.254	$<2e^{-16}$***

Note: This table presents the results of the linear regressions using realized volatility as the dependent variable and the GARCH model estimations (using a 1000 day estimation rolling window) as the independent variable. On the left side the linear regression results using the conditional volatility of the GARCH models estimated with a normal distribution and on the right side the linear results using the conditional volatility of the GARCH models estimated with a t-Student distribution.

Table 5.5: Univariate linear regressions (2000 day estimation period)

	EGARCH-norm		EGARCH-std			GJR-norm		GJR-std	
R^2	0.5393		0.5333		R^2	0.5557		0.5502	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	$-6.281e^{-06}$	$6.515e^{-01}$	0.00039	0.61833	Estimate	-0.00046	0.67693	-0.00006	0.64364
Std. Error	$1.707e^{-04}$	$1.347e^{-02}$	0.00017	0.01294	Std. Error	0.00017	0.01354	0.00017	0.01302
t-value	-0.037	48.359	2.336	47.777	t-value	-2.655	49.984	-0.37	49.44
Prob(>(|t|)	0.971	$<2e^{-16}$***	0.0196*	$<2e^{-16}$***	Prob(>(|t|)	0.008**	$<2e^{-16}$***	0.711	$<2e^{-16}$***

	IGARCH-norm		IGARCH-std			NGARCH-norm		NGARCH-std	
R^2	0.5073		0.4999		R^2	0.5549		0.5524	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	0.00038	0.60477	0.00056	0.59160	Estimate	$-8.714e^{-05}$	$6.495e^{-01}$	$2.343e^{-05}$	$6.408e^{-01}$
Std. Error	0.00017	0.01333	0.00017	0.01324	Std. Error	$1.673e^{-04}$	$1.302e^{-02}$	$1.660e^{-04}$	$1.291e^{-02}$
t-value	2.175	45.360	3.228	44.689	t-value	-0.521	49.903	0.141	49.654
Prob(>(|t|)	0.0297*	$<2e^{-16}$***	0.00126**	$<2e^{-16}$***	Prob(>(|t|)	0.603	$<2e^{-16}$***	0.888	$<2e^{-16}$***

	TGARCH-norm		TGARCH-std			APARCH-norm		APARCH-std	
R^2	0.5365		0.53		R^2	0.5479		0.541	
	Intercept	Sigma	Intercept	Sigma		Intercept	Sigma	Intercept	Sigma
Estimate	$-6.789e^{-05}$	$6.551e^{-01}$	0.00034	0.62071	Estimate	-0.00022	0.66466	0.00022	0.62835
Std. Error	$1.728e^{-04}$	$1.362e^{-02}$	0.00017	0.01308	Std. Error	0.00017	0.01351	0.00017	0.01295
t-value	-0.393	48.088	2.021	47.464	t-value	-1.262	49.203	1.296	48.522
Prob(>(|t|)	0.694	$<2e^{-16}$***	0.0434*	$<2e^{-16}$***	Prob(>(|t|)	0.207	$<2e^{-16}$***	0.195	$<2e^{-16}$***

Note: This table presents the results of the linear regressions using realized volatility as the dependent variable and the GARCH model estimations (using a 2000 day estimation rolling window) as the independent variable. On the left side the linear regression results using the conditional volatility of the GARCH models estimated with a normal distribution and on the right side the linear results using the conditional volatility of the GARCH models estimated with a t-Student distribution.
** indicates significance at the 1% level.

Based on the drawbacks already described the coefficient of determination is not a very useful measure of forecasting accuracy (Murphy and Epstein, 1989). Although the R^2 of the regressions that use the GARCH family models as the independent variable is lower when using an estimation period of 1000 days and higher when using a 2000 days estimation period in comparison with conditional volatility estimated using a 500 day estimation window, nothing can be concluded from that, because there is the possibility that the series is not free of outliers which may have a higher impact in the estimations of the conditional volatility of GARCH models that use a different estimation period. The coefficient of determination of the linear regressions seems to be uniform for each estimation window used in the estimation of the GARCH models. This allows us to say that the GARCH models yield similar results in what comes their predictive power of conditional volatility.

5.2 Tests of forecasting accuracy

The test of SPA is used in order to test for the presence of a model with a superior forecasting ability. In this case we test whether there are any alternative models that produce a better forecast than the benchmark. One can say that the null hypothesis is defined as ”the benchmark is as good as any other model in terms of expected loss” (Hansen and Lunde, 2005). The SPA test is based on the test statistic in equation 3.1. It is an hypothesis test where the null is rejected for small p-values. In order to perform the test of SPA I used six different loss functions, the same as in (Hansen and Lunde, 2005), so that it is possible to compare the models in terms of expected loss.

The models were compared using each model as the benchmark, in other words there are fourteen benchmark models and each benchmark model is compared with the other models used in this study in order to understand if there is a model that is not as good as any other model in terms of expected loss. The SPA test provides three p-values: lower, consistent and upper. The p-value that should be used to the evaluation of the forecasting ability of the model is the consistent bound. If the p-value is low one should reject the null hypothesis, therefore there is a better model than the benchmark for predicting conditional volatility in terms of loss.

Table 5.6: Test of SPA p-values

Benchmark: GARCH(1,1)-norm

	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.136	**0.517**	0.876	0.000	**0.298**	0.557	0.052	**0.437**	0.973
MSE_2	0.045	**0.413**	0.876	0.007	**0.466**	0.948	0.100	**0.480**	0.997
QLIKE	0.087	**0.483**	0.685	0.001	**0.271**	0.490	0.000	**0.388**	0.480
R^2Log	0.026	**0.473**	0.675	0.000	**0.172**	0.325	0.000	**0.332**	0.442
MAE_1	0.251	**0.662**	0.910	0.000	**0.196**	0.510	0.018	**0.442**	0.738
MAE_2	0.375	**0.644**	0.937	0.000	**0.299**	0.639	0.122	**0.505**	0.960

Note: This table shows the p-values of the SPA test using GARCH(1,1) with assumed normal distribution of residuals as the benchmark. The table is divided in three parts corresponding to each estimation period used in the estimation of the conditional volatility with the GARCH(1,1) model. On the left the estimation period used was 500 days, 1000 days estimation period in the middle and 2000 days in the right.

Table 5.6 shows the p-values of the SPA test using GARCH(1,1) estimated with normal distribution as the benchmark. It is possible to say that the model is not outperformed by any other model used in the research or in other words this model seems to be able to capture the variation of conditional volatility. Since the p-value for the window period used in the estimation of the GARCH model of 500 days is above 0.413, meaning that we do not reject the null hypothesis, the performance of the GARCH(1,1)-norm is not statistically worse than any of the competing models. When the window period used in the estimation of the GARCH model increases to 1000 days it is possible to see that the p-values decrease. Despite that, it does not mean that the model is outperformed by the competing models since the p-values for all loss functions are greater than 5%, the significance level used in this case. The lowest p-values is 0.172, well above that significance level. Using a larger period of 2000 days in the estimation of the GARCH models does not change the first conclusion that the GARCH(1,1) with assumed normal distribution is not outperformed by the competing models since the p-values for all loss functions as higher than 0.332.

Table 5.7 shows the the p-values of the SPA test using GARCH(1,1) estimated with t-Student distribution as the benchmark. There is evidence that the model is not outperformed by competing models, once the consistent p-value for the window estimation of 500 days used to estimate the GARCH models is higher than 0.422 for all the loss functions. One can say that there is no evidence to say that the GARCH(1,1)-std is worse for forecasting

Table 5.7: Test of SPA *p*-values

Benchmark: GARCH(1,1)-std

	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.183	**0.455**	0.886	0.064	**0.449**	0.938	0.052	**0.406**	0.759
MSE_2	0.101	**0.422**	0.918	0.115	**0.447**	0.999	0.056	**0.389**	0.791
QLIKE	0.142	**0.459**	0.673	0.001	**0.347**	0.846	0.000	**0.315**	0.475
R^2Log	0.062	**0.453**	0.761	0.000	**0.350**	0.841	0.001	**0.348**	0.562
MAE_1	0.408	**0.520**	0.962	0.076	**0.456**	0.982	0.038	**0.395**	0.595
MAE_2	0.533	**0.552**	0.991	0.296	**0.504**	0.996	0.113	**0.456**	0.775

Note: This table shows the p-values of the SPA test using GARCH(1,1) with assumed t-Student distribution of residuals as the benchmark. The table is divided in three parts corresponding to each estimation period used in the estimation of the conditional volatility with the GARCH(1,1) model. On the left the estimation period used was 500 days, 1000 days estimation period in the middle and 2000 days in the right.

conditional volatility in terms of expected loss. The same thing happens when the period used in the estimation of the GARCH model increases to 1000 days. The lowest p-value of all loss functions is 0.347, meaning that the null hypothesis is not rejected. Thus the GARCH(1,1) estimated with t-Student distribution is not outperformed by the competing models. When increasing the time period used in the estimation of the GARCH models to 2000 days there are not many differences in the conclusions. The lowest p-value is 0.315, leading to the conclusion that there is evidence that the GARCH(1,1) model with assumed t-Student distribution is not outperformed by any other model used in the research.

The final conclusions are similar for both GARCH(1,1) models estimated with different distributions. The models present similar p-values for the SPA test for the different loss functions and in all cases the null hypothesis is rejected. This means that there is not much to gain in changing the distribution of the GARCH model since it does not provide improved forecasts.

Table 5.8 shows the p-values of the SPA test using different GARCH models estimated with a 500 day window period as the benchmark. In the case where the EGARCH estimated with normal distribution is the benchmark there is evidence that the model is not outperformed by competing models since the lower p-value for all the loss functions is 0.42. When the EGARCH model estimated with t-Student distribution is the benchmark the same conclusions are made since the p-values of all loss functions are similar to the latter.

When choosing GJR with normal distribution as the benchmark there is reason to believe that the model is not outperformed by any other model that is used in this research since the lowest p-value for all loss functions is 0.311. For the GJR with assumed t-Student distributions conclusions are the same, the smallest p-value is 0.203 thus one do not reject the null hypothesis. The IGARCH model with assumed normal distribution is the only model that presents low p-values. However it does not happen for all the loss functions. In this case there is the possibility that the model is outperformed by other models once the lowest p-value is equal to zero and there are other p-values lower than 5% meaning that, if using those loss functions as a way to evaluate volatility models, there is evidence that there are better models for predicting conditional volatility. This also happens for the IGARCH model with assumed t-Student distribution, although in this case only for one loss function, the MSE_2 where the p-value is equal to 0.034. When choosing NGARCH with both assumed normal distribution or t-Student distribution there is evidence to say that the model is not outperformed by any other model used in the study since the p-values for all loss functions are high in value. As for the TGARCH model the same conclusions can be made. In fact the p-values presented when the TGARCH model is used as the benchmark are the higher of all the SPA test p-values. Regardless this does not mean that the TGARCH model is the best model, among those used in the research, to forecast conditional volatility. Analysing the p-values for when the APARCH with assumed normal distribution is the benchmark there is reason to believe that the model is not outperformed by the competing models. The same thing happens for the APARCH with assumed t-Student distribution where the lower p-value of all loss functions is 0.396, meaning that the null hypothesis that the benchmark model is as good as any other model in terms of loss is not rejected.

Table 5.9 shows the p-values of the SPA test using different GARCH models estimated with 1000 day window period as the benchmark. When the EGARCH with assumed normal distribution is the benchmark model there is evidence to say that the benchmark is not outperformed by the competing models due to the fact that the lowest p-value for all the loss functions is 0.421, meaning that the null hypothesis is not rejected thus the benchmark is as good as any other model in terms of loss. For the EGARCH with assumed t-Student distribution the conclusions are the same once the p-values of all the loss functions are higher than 0.461. GJR estimated with normal distribution as the benchmark brings us some intriguing results. There is evidence that the model is outperformed by other models since the p-values of the R^2LOG and MAE_1 are bellow 5%, being that the p-values for the other loss functions are quite low as well. This is surprising because the GJR model is an asymmetric model. Franses and Dijk (1996) had the same results when the sample contained extreme observations. Different results are obtained when the GJR with assumed t-Student distribution is the benchmark. The p-values for all the loss functions are higher than 0.213, meaning that there is no evidence that the model is outperformed by the competing models. This findings may lead to the conclusion that when the GJR is estimated with t-Student distribution is more robust. Liu and Morley (2009) found that GARCH models with non-normal distributions are more robust in what comes to volatility forecast. When analysing the IGARCH with assumed normal distribution as the benchmark there is some evidence that the model is outperformed by the competing models as happened when using a smaller rolling window. For the MSE_2 the p-value is bellow 5% meaning that for this loss function the null hypothesis is rejected thus there are better models among the competing models in terms of loss. Other loss functions present a low p-value, however it is not bellow the significance level. When the IGARCH with t-Student distribution is used as the benchmark there is no evidence that the model is not good for predicting volatility in terms of expected loss, since the lower p-value for all the loss functions is 0.15. For when the NGARCH, TGARCH and APARCH with both distributions are the benchmark the conclusion is always the same. All the SPA test p-values are high for those models when they are used as the benchmark indicating that the benchmark model is not outperformed by any other model used in the research.

Table 5.10 shows the p-values of the SPA test using different GARCH models estimated with 2000 day window period as the benchmark. When the EGARCH model estimated with normal distribution is the benchmark model there is no evidence that the model is outperformed by competing models. In fact the p-values are quite high for all the loss functions, therefore the null hypothesis is not rejected. For the case of the EGARCH with t-Student distribution, although the p-values are not as high the same conclusion is made once the lowest p-value for all the loss functions is 0.382 not near the rejection significance level of 5%. For the GJR with normal distribution there is no evidence that the model is not as good as any other model used in the research in terms of loss once all p-values are above 0.398. The same happens for the GJR with t-Student distribution where the lowest p-value of all the loss functions is 0.378 thus there is no evidence that the model is outperformed by the contending models. In this case the GJR is not outperformed by other models as happened when the rolling window estimation period was 1000 days. This probably happens due to the fact that the window is larger and this leads to a dilution of the extreme values or in other words the extreme values do not have such a high impact as in the previous case. When the IGARCH with normal distribution is the benchmark model there is some evidence that the model does not perform as good as the other models in terms of expected loss. With the exception of the p-value for the MSE_2 the p-values for all the other loss functions are quite low. Despite that only the p-value for the MAE_1 is bellow 5%, meaning that for that loss function we reject the null hypothesis that the model is as good as any other model for predicting volatility. For the IGARCH with t-Student distribution the p-values continue to be low, however none of them bellow the significance level of 5%, implying that the model is not outperformed by competing models. Looking at the NGARCH with normal distribution it is possible to say that there is evidence that the model is a good model for predicting volatility in terms of loss in comparison with the models used in the research since the lower p-value for all the loss functions is 0.431. The same conclusions are made for the NGARCH with t-Student distribution where the lowest p-value of all loss functions is 0.415 meaning that we do not reject the null hypothesis thus this model is as good as the competing models. For the TGARCH estimated with both distributions there is no reason to believe that the benchmark model is outperformed by any of the competing models once the p-values are high in both cases. As for the APARCH the conclusions are the same once the p-values present high values thus we do not reject the null hypothesis that the benchmark model is as good as any other model in terms of expected loss.

Table 5.8: Test of SPA p-values (500 day estimation period)

	EGARCH-norm			EGARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.021	**0.482**	0.990	0.128	**0.452**	0.986
MSE_2	0.000	**0.440**	0.998	0.000	**0.471**	0.996
QLIKE	0.028	**0.513**	1.000	0.382	**0.543**	0.999
R^2Log	0.013	**0.582**	1.000	0.282	**0.484**	1.000
MAE_1	0.022	**0.516**	0.924	0.428	**0.651**	0.985
MAE_2	0.037	**0.420**	0.856	0.309	**0.559**	0.952

	GJR-norm			GJR-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.047	**0.441**	0.629	0.051	**0.468**	0.644
MSE_2	0.043	**0.388**	0.838	0.000	**0.203**	0.203
QLIKE	0.060	**0.393**	0.393	0.084	**0.447**	0.577
R^2Log	0.019	**0.311**	0.311	0.035	**0.444**	0.568
MAE_1	0.049	**0.340**	0.340	0.095	**0.477**	0.529
MAE_2	0.077	**0.448**	0.448	0.089	**0.515**	0.535

	IGARCH-norm			IGARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.007	**0.087**	0.087	0.008	**0.167**	0.238
MSE_2	0.000	**0.000**	0.000	0.000	**0.034**	0.078
QLIKE	0.010	**0.347**	0.347	0.024	**0.322**	0.335
R^2Log	0.012	**0.349**	0.349	0.001	**0.207**	0.290
MAE_1	0.000	**0.142**	0.190	0.004	**0.232**	0.367
MAE_2	0.003	**0.169**	0.169	0.012	**0.230**	0.412

	NGARCH-norm			NGARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.092	**0.513**	0.984	0.030	**0.465**	0.957
MSE_2	0.192	**0.505**	0.986	0.002	**0.295**	0.887
QLIKE	0.109	**0.501**	0.984	0.045	**0.519**	0.970
R^2Log	0.028	**0.458**	0.998	0.012	**0.453**	0.994
MAE_1	0.039	**0.455**	0.772	0.015	**0.359**	0.635
MAE_2	0.077	**0.520**	0.804	0.039	**0.427**	0.664

	TGARCH-norm			TGARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.233	**0.537**	1.000	0.827	**0.827**	1.000
MSE_2	0.777	**0.777**	1.000	0.395	**0.458**	1.000
QLIKE	0.085	**0.471**	0.997	0.706	**0.706**	1.000
R^2Log	0.009	**0.414**	1.000	0.647	**0.647**	1.000
MAE_1	0.156	**0.502**	0.923	0.702	**0.702**	0.998
MAE_2	0.410	**0.607**	0.967	0.648	**0.651**	0.998

	APARCH-norm			APARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.053	**0.563**	1.000	0.091	**0.505**	1.000
MSE_2	0.187	**0.527**	1.000	0.002	**0.396**	1.000
QLIKE	0.099	**0.556**	0.995	0.194	**0.515**	1.000
R^2Log	0.006	**0.436**	0.999	0.085	**0.494**	1.000
MAE_1	0.094	**0.536**	0.897	0.222	**0.598**	1.000
MAE_2	0.178	**0.599**	0.940	0.191	**0.570**	0.999

Note: This table shows the p-values of the SPA test using different GARCH family models that were estimated using a 500 day estimation window as a benchmark. On the left the p-value of the SPA test of GARCH-type (benchmark) model that was assumed to have normal distribution and on the right the GARCH-type model that was assumed to have t-Student distribution.

Table 5.9: Test of SPA p-values (1000 day estimation period)

	EGARCH-norm			EGARCH-std				GJR-norm			GJR-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u		SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.000	**0.605**	1.000	0.552	**0.552**	1.000	MSE_1	0.000	**0.142**	0.142	0.004	**0.241**	0.412
MSE_2	0.000	**0.421**	1.000	0.008	**0.461**	1.000	MSE_2	0.005	**0.456**	0.729	0.002	**0.462**	0.710
QLIKE	0.000	**0.481**	1.000	0.568	**0.568**	1.000	QLIKE	0.000	**0.074**	0.074	0.000	**0.315**	0.579
R^2Log	0.000	**0.397**	1.000	0.607	**0.607**	1.000	R^2Log	0.000	**0.033**	0.033	0.000	**0.251**	0.517
MAE_1	0.000	**0.470**	0.999	0.664	**0.664**	1.000	MAE_1	0.000	**0.037**	0.037	0.000	**0.252**	0.467
MAE_2	0.001	**0.573**	0.985	0.665	**0.665**	1.000	MAE_2	0.001	**0.085**	0.085	0.001	**0.213**	0.372

	IGARCH-norm			IGARCH-std				NGARCH-norm			NGARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u		SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.000	**0.080**	0.149	0.028	**0.420**	0.726	MSE_1	0.000	**0.422**	0.759	0.003	**0.451**	0.977
MSE_2	0.000	**0.001**	0.001	0.000	**0.150**	0.308	MSE_2	0.000	**0.172**	0.310	0.037	**0.488**	0.914
QLIKE	0.000	**0.175**	0.419	0.000	**0.323**	0.733	QLIKE	0.000	**0.198**	0.494	0.001	**0.445**	0.993
R^2Log	0.000	**0.265**	0.457	0.002	**0.309**	0.719	R^2Log	0.000	**0.256**	0.677	0.000	**0.414**	0.999
MAE_1	0.000	**0.201**	0.501	0.028	**0.498**	0.922	MAE_1	0.000	**0.367**	0.585	0.000	**0.342**	0.888
MAE_2	0.000	**0.205**	0.450	0.078	**0.518**	0.940	MAE_2	0.000	**0.278**	0.554	0.000	**0.288**	0.723

	TGARCH-norm			TGARCH-std				APARCH-norm			APARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u		SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.000	**0.419**	0.960	0.294	**0.494**	1.000	MSE_1	0.000	**0.350**	0.945	0.021	**0.657**	1.000
MSE_2	0.000	**0.445**	0.957	0.750	**0.750**	1.000	MSE_2	0.006	**0.608**	0.995	0.029	**0.525**	0.998
QLIKE	0.000	**0.419**	0.989	0.024	**0.509**	1.000	QLIKE	0.000	**0.426**	0.965	0.006	**0.530**	1.000
R^2Log	0.000	**0.371**	0.999	0.010	**0.456**	1.000	R^2Log	0.000	**0.352**	0.994	0.000	**0.442**	1.000
MAE_1	0.000	**0.367**	0.922	0.028	**0.459**	1.000	MAE_1	0.000	**0.372**	0.873	0.001	**0.479**	1.000
MAE_2	0.001	**0.403**	0.881	0.281	**0.550**	1.000	MAE_2	0.000	**0.327**	0.758	0.011	**0.615**	1.000

Note: This table shows the p-values of the SPA test using different GARCH family models that were estimated using a 1000 day estimation window as a benchmark. On the left the p-value of the SPA test of GARCH-type (benchmark) model that was assumed to have normal distribution and on the right the GARCH-type model that was assumed to have t-Student distribution.

Table 5.10: Test of SPA p-values (2000 day estimation period)

	EGARCH-norm			EGARCH-std				GJR-norm			GJR-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u		SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.633	**0.633**	1.000	0.176	**0.546**	1.000	MSE_1	0.029	**0.528**	0.956	0.014	**0.378**	0.927
MSE_2	0.766	**0.766**	1.000	0.068	**0.382**	0.971	MSE_2	0.243	**0.464**	0.994	0.052	**0.424**	0.949
QLIKE	0.077	**0.489**	1.000	0.668	**0.668**	1.000	QLIKE	0.001	**0.394**	0.574	0.003	**0.452**	0.844
R^2Log	0.009	**0.479**	1.000	0.657	**0.657**	1.000	R^2Log	0.000	**0.420**	0.579	0.001	**0.416**	0.837
MAE_1	0.710	**0.710**	1.000	0.450	**0.566**	1.000	MAE_1	0.001	**0.398**	0.647	0.001	**0.418**	0.714
MAE_2	0.639	**0.639**	1.000	0.182	**0.528**	0.995	MAE_2	0.014	**0.500**	0.903	0.009	**0.470**	0.854

	IGARCH-norm			IGARCH-std				NGARCH-norm			NGARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u		SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.001	**0.194**	0.242	0.018	**0.213**	0.213	MSE_1	0.090	**0.541**	0.981	0.067	**0.513**	0.993
MSE_2	0.043	**0.409**	0.667	0.062	**0.222**	0.222	MSE_2	0.116	**0.490**	0.962	0.057	**0.420**	0.975
QLIKE	0.000	**0.136**	0.136	0.000	**0.173**	0.196	QLIKE	0.021	**0.537**	0.987	0.020	**0.463**	0.999
R^2Log	0.000	**0.146**	0.146	0.000	**0.281**	0.358	R^2Log	0.006	**0.431**	1.000	0.006	**0.415**	1.000
MAE_1	0.000	**0.034**	0.034	0.017	**0.217**	0.252	MAE_1	0.043	**0.533**	0.936	0.028	**0.546**	0.967
MAE_2	0.004	**0.083**	0.105	0.015	**0.162**	0.162	MAE_2	0.064	**0.466**	0.931	0.036	**0.506**	0.948

	TGARCH-norm			TGARCH-std				APARCH-norm			APARCH-std		
	SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u		SPA_l	$\mathbf{SPA_c}$	SPA_u	SPA_l	$\mathbf{SPA_c}$	SPA_u
MSE_1	0.341	**0.535**	1.000	0.190	**0.500**	0.996	MSE_1	0.267	**0.617**	0.999	0.131	**0.586**	0.998
MSE_2	0.236	**0.482**	0.999	0.090	**0.380**	0.904	MSE_2	0.238	**0.502**	1.000	0.080	**0.377**	0.935
QLIKE	0.092	**0.558**	1.000	0.096	**0.517**	1.000	QLIKE	0.032	**0.599**	1.000	0.048	**0.528**	1.000
R^2Log	0.007	**0.497**	1.000	0.046	**0.468**	1.000	R^2Log	0.008	**0.521**	1.000	0.008	**0.460**	1.000
MAE_1	0.306	**0.545**	1.000	0.277	**0.580**	1.000	MAE_1	0.037	**0.536**	0.974	0.070	**0.595**	0.982
MAE_2	0.436	**0.584**	0.997	0.186	**0.476**	0.975	MAE_2	0.142	**0.623**	0.998	0.127	**0.617**	0.974

Note: This table shows the p-values of the SPA test using different GARCH family models that were estimated using a 2000 day estimation window as a benchmark. On the left the p-value of the SPA test of GARCH-type (benchmark) model that was assumed to have normal distribution and on the right the GARCH-type model that was assumed to have t-Student distribution.

Pagan and Schwert (1990) found that GARCH and EGARCH models produce moderate results, Franses and Dijk (1996) found that non-linear GARCH models do not outperform the normal GARCH model if the sample contains extreme observations, Brailsford and Faff (1996) found that GJR and GARCH models are superior to other models, Hansen and Lunde (2005) found that the GARCH(1,1) model is not outperformed by other models for the US daily return, Chong et al. (1999) found that EGARCH model outperforms the other variations of the GARCH models and that IGARCH model can not be recommended for forecasting volatility and Awartani and Corradi (2005) had similar results in accordance with Kişinbay (2010) that found that asymmetric models, including EGARCH provide improvements in forecasting volatility.

As it is possible to see in tables 5.8 to 5.10 the IGARCH model with assumed normal distribution is the one that presents evidence that the model is outperformed in all the periods used to estimate the GARCH models. However this evidence is not totally consistent since this does not happen for all the loss functions, meaning that for some loss functions the model is not outperformed by other competing models. There are some cases for the IGARCH with assumed t-Student distribution where the p-value is bellow the significance level although not for all the loss functions leading to the same conclusion as the latter model. The GJR with assumed normal distribution provides intriguing results since the model is as good as any other model used in the research when the estimation windows are 500 days and 2000 days and when the window estimation is 1000 days the model does not perform better than the competing models.

Chapter 6

Conclusion

This research is focused on the predicting ability of volatility of GARCH family models. I compared seven GARCH models in order to study which model performs the best in forecasting volatility for the NASDAQ-100 stock exchange.

Firstly, I performed a series of Mincer-Zarnowitz regressions. I concluded that there was a big part of the realized volatility that was not explained by this method due to the fact that the determination coefficient was not close to one. It is possible to say that by this method the GARCH models that have better performance are the ones that are more sophisticated, although one can not say which model is best in terms of predicting ability, since the method does not allow for GARCH model comparison.

In order to assess which of the models had better performance in terms of volatility forecast I used the SPA test proposed by Hansen and Lunde (2005).

In the course of this research I find that the simple GARCH model estimated with both normal and t-Student distribution is not outperformed in what comes to forecasting next-day volatility by any other model used in the research for all the three estimation periods. This results are in accordance with the results in Hansen and Lunde (2005). Several authors such as Kişinbay (2010), Pagan and Schwert (1990) and Awartani and Corradi (2005) found that the EGARCH model provides better improvements in forecasting volatility. The same thing happens in this research where there is no reason to believe that the EGARCH model estimated with both normal and t-Student distribution is not as good as any other model to predict conditional volatility in terms of loss. In terms of forecasting next-day volatility there are some possible exceptions. The IGARCH model presents results for some loss functions that are inferior to the results produced by competing models in agreement with Chong et al. (1999) that does not recommend the use of this model to forecast volatility. The GJR provides intriguing results. It is also possible to say that the GJR model estimated with normal distribution is outperformed by the competing models when the window estimation used is 1000 days. Franses and Dijk (1996) found that GJR is outperformed by the competing models when the sample contains extreme observations, however that is not the case here. The same conclusions are not made when using the GJR model estimated with t-Student distribution. Liu and Morley (2009) found that GARCH models with non-normal distributions are more robust in what comes to volatility forecasting. The NGARCH, TGARCH and APARCH models are not outperformed by any competing models. Thus these models also provide good conditional volatility forecasts for the next day. I found that increasing the number of days used in the estimation of the rolling window may lead to different results, however in the case of the IGARCH model there is no difference in changing the window estimation since the model is allways outperformed by the competing models. Given that the models are highly correlated as presented earlier in section 5.1.1 this results are expected in the sense that the models used in the research are similar in terms of their characteristics although there is the possibility of outliers that would have impact in the final results.

The realized volatility is not the most accurate estimator of σ_t^2 when the periodic pattern is known for each day t (Taylor, 2005). Hansen and Lunde (2006) show that realized volatility is affected by market microstructure noise under the general specification for the noise that allows for various forms of stochastic dependencies. With the increase of high-frequency data availability there have been a number of volatility including realized kernel and bipower variation (Barndorff-Nielsen and Shephard, 2002). Kernel based estimators can unearth important characteristics of market microstructure noise and thus a simple kernel-based estimator dominates the realized volatility (Hansen and Lunde, 2006). The use of a different measure of volatility such as the realized kernel may yield different results.

It is not possible to create a perfect risk model. The risk forecaster needs to evaluate the pros and cons of the various models (in these case to forecast conditional volatility) and data choices to create what inevitably can only be an imperfect model (Danielsson, 2002). In the case of predicting volatility one is trying to predict what will happen in the future, thus no model can achieve that to a degree of 100%. However one can use different models to achieve the better possible volatility forecast. The use of new GARCH models or other models can generate better forecasting results. Realized measures, namely realized kernel and bipower variation, are more informative about the current level of volatility than is the squared return, making them very useful in modeling and forecasting future volatility alluring to the use of a realized measure in the GARCH equation (GARCH-X model) (Hansen et al., 2012).

In order to understand if there is a better model among the ones used in the research I used the SPA test that gives us the chance to compare models in terms of expected loss. The use of utility functions instead of the loss functions used in this research could yield different results on which models perform better than the benchmark.

Bibliography

Akgiray, V. (1989). Conditional heteroskedasticity in time series of stock returns: Evidence. *Journal of Business*, 62(1):55–80.

Allen, D. M. (1971). Mean square error of prediction as a criterion for selecting variables. *Technometrics*, 13(3):469–475.

Andersen, T. G. and Benzoni, L. (2008). *Financial Time Series: Realized Volatility*. Springer Verlag.

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

Andersen, T. G., Bollerslev, T., Diebold, F., and Labys, P. (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55.

Awartani, B. M. and Corradi, V. (2005). Predicting the volatility of the s&p-500 stock index via garch models: the role of asymmetries. *International Journal of Forecasting*, 21:167–183.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, 64:253–280.

Black, F. (1976). Studies of stock price volatility changes. *American statistical association, Business and Economic statistics section*, pages 177–181.

Black, F. and Scholes, M. (1974). The valuation of option contracts and a test of market efficiency. *Journal of Finance*, 27:399–417.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31:27–307.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52((1-2)):5–59.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). *ARCH Models Handbook of Econometrics*, volume 4. R. F. Engle and D. McFadden, Amsterdam: North-Holland.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65:1509–1526.

Brailsford, T. J. and Faff, R. W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking and Finance*, 20:38–419.

Brooks, C. (2014). *Introductory econometrics for finance*, volume 3. Cambridge University press.

Chong, C. W., Ahmad, M. I., and Abdullah, M. Y. (1999). Performance of garch models in forecasting stock market volatility. *Journal of Forecasting*, 18:333–343.

Christensen, B. and Prabhala, N. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, 50:125–150.

Corrado, C. J. and Miller, J. T. W. (2005). The forecast quality of cboe implied volatility indexes. *Journal of Futures Markets*, 25:339–373.

Danielsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance*, 26(7):1273–1296.

Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journla of empirical finance*, 1(1):83–106.

Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *Journal of Finance*, 53:2059–2106.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, 50:987–1008.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5:1–50.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica*, 55(2):391–407.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48:1749–1778.

Engle, R. F. and Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance*, 1:237–245.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., and Arshanapalli, B. G. (2014). *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*. Wiley.

Fama, E. (1965). The behaviour of stock market prices. *Journal of Business*, 38:34–105.

Flannery, M. J. and Protopapadakis, A. A. (2002). Macroeconomic factors do influence aggregate stock returns. *The review of financial studies*, 15(3):751–782.

Franses, P. H. and Dijk, D. V. (1996). Forecasting stock market volatility using (non-linear) garch models. *Journal of forecasting*, 15:229–235.

Gaunersdorfer, A. and Hommes, C. (2007). *Long Memory in Economics: A Nonlinear Structural Model for Volatility Clustering*. Springer, Berlin Heidelberg.

Glosten, L. R., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of finance*, 48:1779–1801.

Goncalves, S. and Jong, R. (2003). Consistency of the stationary bootstrap under weak moment conditions. *Economics Letters*, 81:273–278.

Granger, C. W. J. (1996). Can we improve the perceived quality of economic forecasts? *Journal of Applied Econometrics*, 11:455–473.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a garch(1,1)? *Journal of applied econometrics*, 20:873–889.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Hunter, J. S. (1986). The exponentially weighted moving average. *Journal of Quality Technology*, 18(4).

Jiang, G. J. and Tian, Y. S. (2005). The model-free implied volatility and its information content. *The Review of Financial Studies*, 18:1305–1342.

Kişinbay, T. (2010). Predictive ability of asymmetric volatility models at medium-term horizons. *Applied Economics*, 42:3813–3829.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, 65(3):361–393.

Liu, W. and Morley, B. (2009). Volatility forecasting in the hang seng index using the garch approach. *Asia-Pacific Financial Markets*, 16:51–63.

Lorde, T. and Moore, W. (2008). Modeling and forecasting the volatility of long-stay tourist arrivals. *Tourism Analysis*, 13(1):43–51.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of business*, 36:394–419.

Marcucci, J. (2005). Forecasting stock market volatility with regime-switching garch models. *Nonlinear Dynamics & Econometrics*, 9.

McKenzie, M. and Mitchell, H. (2002). Generalized asymmetric power arch modeling of exchange rate volatility. *Applied Financial Economics*, 12(8):555–564.

Mincer, J. and Zarnowitz, V. (1969). *The Evaluation of Economic Forecasts*. Cambridge: National Bureau of Economic Research.

Murphy, A. H. and Epstein, E. S. (1989). Skill scores and correlation coefficients in model verification. *Monthly Weather Review*, 177(3):572–582.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692.

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

Pagan, A. R. and Schwert, G. W. (1990). Alternative models for conditional stock volatilities. *Journal of Econometrics*, 45:90–267.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89:1303–1313.

Poterba, J. M. and Summers, L. H. (1986). The persistence of volatility and stock market fluctuations. *American Economic Review*, 76:1142–1151.

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of finance*, 44:1115–1153.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19:425–442.

Taylor, S. J. (1986). Forecasting the volatility of currency exchange rates. *International Journal of Forecasting*, 3:70–159.

Taylor, S. J. (2005). *Asset Price Dynamics, Volatility, and Prediction*, volume 1. Princeton University press, 41 William Street, Princeton, New Jersey 08540.

Teräsvirta, T. (2009). An introduction to univariate garch models. *Handbook of Financial time series*, pages 17–42.

West, K. D. and Cho, D. (1995). The predictive ability of several models of exchange rate volatility. *Journal of econometrics*, 69(2):367–391.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68:1097–1126.

Wu, J. (2010). Threshold garch model: Theory and application. *The University of Western Ontario*.

Xiao-lin, L., Balcilar, M., Gupta, R., and Chang, T. (2016). The causal relationship between economic policy uncertainty and stock returns in china and india: Evidence from a bootstrap rolling window approach. *Emerging Markets Finance and Trade*, 52(3):674–689.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics*, 18(5):931–955.

Zivot, E. and Wang, J. (2006). *Modeling Financial Time Series: Rolling Analysis of Time Series*, volume 22. Springer.