Realized Volatility: Assessing the predictive performance of parametric volatility models

Pedro Serrasqueiro

Spring 2011

Resumo

A presente dissertação pretende efectuar uma avaliação da capacidade predictiva de vários modelos GARCH, nomeadamente os modelos GARCH, EGARCH e GJR-GARCH, comparando as suas previsões com duas medidas para a volatilidade. Os resultados são obtidos após um enquadramento teórico da volatilidade realizada e das propriedades das suas distribuições, tanto condicionais como incondicionais, efectuando a mesma análise para os retornos. Em linha com os resultados já existentes na literatura, as distribuições incondicionais são leptocúrticas e positivamente enviesadas, sendo que a volatilidade realizada se afasta mais da normalidade e exibe efeito assimétrico. Por outro lado, os retornos standardizados pelo desvio-padrão realizado aparentam ser aproximadamente normais.

Palavras-Chave: Volatilidade Realizada (RV), Dados de Alta-frequência, Modelos GARCH, Volatilidade Assimétrica.

Abstract

The present dissertation intends to perform an evaluation of the predictive ability of several models of the GARCH family, namely the GARCH, EGARCH and GJR-GARCH models, by comparing their forecasts with two different proxies for volatility. This result is achieved after providing a brief theoretical framework for realized volatility and after assessing its unconditional and conditional distributional properties. Consistently with the results found in previous literature, unconditional distributions for returns and realized volatility are leptokurtic and rightly skewed, with realized volatility departing more from normality. On the other hand, the logarithm of realized volatility and returns standardized by the squared root of realized volatility appear to be close to normal. Furthermore, the logarithm of realized volatility exhibits the already documented asymmetric effect of volatility.

Keywords: Realized Volatility (RV), High Frequency Data, GARCH Models, Asymmetric volatility.

THANK YOU NOTE

Ao professor José Dias Curto, pela orientação e ajuda na escolha do tema.

À minha família.

Contents

List of Figures

List of Tables

Part I
Introduction

With the continuous and growing evolution in financial markets and the development of more complex financial products, which played a preponderant part in the ongoing economic crisis, to understand, both empirically and theoretically, the volatility of returns seems to be a great necessity. The predictability of daily returns is known to be very low, if not null, especially when considering the asset class of stocks. On the other hand, the volatility of these returns seems to be relatively easier to forecast.

The assessment of risk assumes a preponderant role in finance. Therefore, volatility (which, given the univariate dimension of the present work will be considered as standard deviation or variance), being the common measure of risk, has been the subject of extensive and thorough literature. These findings have a wide range of applications, that include derivative pricing (such as options or products that have embedded options), capital allocation, risk management, among many others. It is then understandable why the modeling, estimation and prediction of volatility have been the subject of such detailed investigation.

Traditionally, the analysis of financial time series focused mainly on the conditional mean of the process, disregarding volatility, that is, considering it constant throughout time, although this is never found in empirical evidence. The modeling of the conditional mean was then made using the widely known ARMA or ARIMA models.

Engle (1982) developed the Autoregressive Conditional Heteroskedasticity model, or ARCH, to model the conditional variance directly, in addition to the conditional mean using the models previously mentioned. The estimation of the ARCH model introduced the possibility of estimating volatility as a function of its past realizations. Later, derivations of ARCH appeared, such as the Generalized Autoregressive Conditional Heteroskedasticity, or GARCH, by Bollerslev (1986), where an ARMA model is used to model the conditional variance of the error terms, the ARCH-M by Engle, Lilien and Robinson (1987), where a heteroskedastic term is introduced in the mean equation in order to solve the traditional problem of considering the variance as constant throughout the time series.

On a slight different category fall the EGARCH (Nelson, 1991) and the GJR-GARCH (Glosten, Jagannathan and Runkle (1993)). These models introduced in the modeling of volatility a stylized fact: markets react asymmetrically to good and bad news, where the bad news have a greater impact on volatility. This is often referred to as the *leverage effect* and is rooted on the notion of a News Impact Curve (NIC), motivated by the empirical work of Black (1976), Christie (1982), Nelson (1991) or French et al. (1987). Suffice it to say that a large number of models were developed in addition to those referred, but their thorough explanation falls out of the purpose of this work.

Another class of models that has been the object of study is that of Stochastic Volatility (SV). These models define the volatility not only as a function of its past but also of stochastic fluctuations [see, for instance, the work of Andersen (1994)]. Also, the implied volatility of derivatives (mainly of options) has been thoroughly documented.

In addition, a new procedure has been suggested to construct ex-post measures of volatility and is referred to as Realized Volatility, which will be applied in the present work. Merton (1980) concluded that the volatility over a fixed interval can be estimated by the sum of squared observations, given that there is a sufficiently high sampling frequency. Provided that, for stocks, the Trade and Quote database (TAQ) records every transaction that occurs, the highest possible frequency (tick-by-tick) is now available to researchers, although this brings up the problem of microstructure noise, which shall be explained in more detail later. Measuring volatility in this way allows one to treat it as an observable variable, in contrast to SV and GARCH-type models where it is considered latent and, therefore, gives us the possibility of evaluating the predictive power of volatility models, which is the purpose of the present work.

Using data from Bloomberg for the EURUSD exchange rate, three different volatily models were estimated to produce predictions that were then compared to the realized volatility measure and to daily squared returns by the Mean Squared Error criteria.

Although the term realized volatility actually refers to the realized standard deviation, in this text it will be used interchangeably when references to the variance are made. Clear distinctions will be made when necessary.

The dissertation is organized as follows: in Part II we shall give a literature review, overviewing the properties of volatility and where we explain in detail the GARCH-type models to estimate in this study. Part III contains the empirical study, where we formulate the research aims, describe the data and explain the methodology. Results, conclusions and some limitations of this work, together with suggestions for further work are also included in this part. Part IV is an appendix containing plotted series for residuals.

Part II
Literature Review

1 Properties of volatility

Volatility seems to have several stylized facts that have been thoroughly documented in the literature and over which volatility models are designed. One of these empirical regularities is that standard latent volatility models fail to describe properly the low and slowly decreasing autocorrelations in squared returns, which are associated with the high excess kurtosis of returns, McAleer and Medeiros (2006).

Bollerslev et al (1994) documented several of these empirical regularities:

1. Asset returns are *leptokurtic;*
2. Returns are not independent and identically distributed, a fact that is often denominated as *clustering*;
3. Volatility and returns are negatively correlated. That is to say that for a given price change, volatility will be affected differently according to the sign of that change. In particular, negative price changes have greater impact on volatility than positive ones. This effect was initially documented by Black (1976);
4. Information that is acknowledged when markets are close is reflected in prices when markets reopen;
5. Foreseeable releases of relevant market information are associated with high ex-ante volatility;
6. Volatility and serial correlation are inversely correlated;
7. Measures of macroeconomic uncertainty help to explain changes in market volatility.

Accordingly, the idea that returns follow a normal distribution has been refuted in most studies, although realized and logarithmic realized volatilies seem to be close to normal distributed. Furthermore, the data seems to show that long-range dependence exists in the volatility of financial time series.

In a review paper, McAleer and Medeiros (2006) list several stylized facts of realized volatility. It exhibits a much higher kurtosis than that of a normal distribution at intraday frequencies, which is the same as saying that distributions of realized volatility have fat tails. The higher the sampling interval, the lower the kurtosis. Probability density functions are thus leptokurtic and converge slowly to the normal distribution.

As mentioned in Section 2, the predictive ability of the models under analysis will be assessed using two different proxies of volatility, of which we provide a visual comparison:

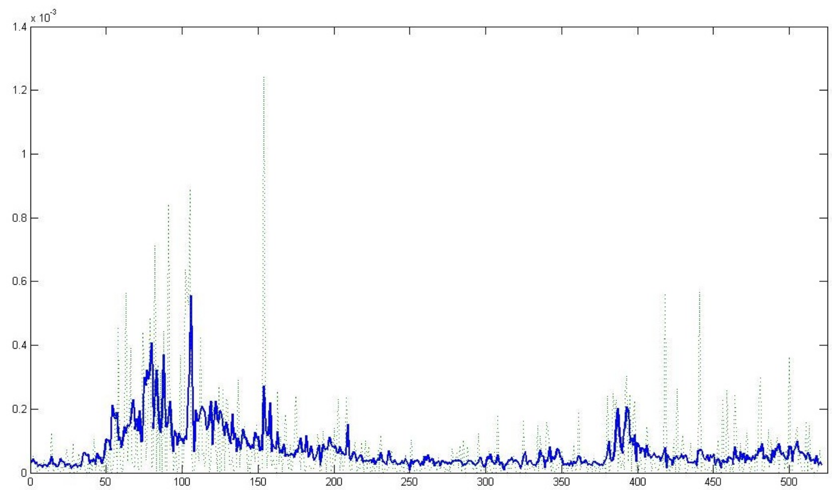


Figure 1: Realized Volatility (solid) and Daily Squared returns (dotted)

By observing Figure 1 one can easily see that realized volatility is a less noisy proxy for actual volatility to a great extent when compared to daily squared returns.

Another stylized fact is long memory. The autocorrelation of returns appears to be insignificant at every scale, but it shows great persistence for a long time interval when the returns are squared. This result holds true for volatilities aggregated at all frequencies (hourly, daily, weekly and monthly).

Also the unconditional distributions of realized volatility are highly rightly skewed and leptokurtic. Realized standard deviation and logarithmic realized variance appear to be much closer to normal distributions, where returns standardized by the realized standard deviation are almost to normal. Furthermore, volatility does not seem to have a unit root, that is, it seems to be stationary, but there is strong evidence of fractional integration. In the following lines a theoretical formulation of realized volatility is given.

Suppose that at day t, the logarithmic prices of some asset follow a standard continuous time process:

$$dp(t+\tau) + \sigma(t+\tau)dW(t+\tau),\ t = 1, 2, 3, ...,\ 0 \leq \tau \leq 1 \tag{1}$$

Barndorff-Nielsen and Shephard (2002) and Andersen et al (2003) proved that daily compound returns, defined as $r_t = log(p_t/p_{t-1})$ are normally distributed conditional on the information set generated by the sample paths of p, such that

$$r_t \mid \Im_t \sim N\left(\int_0^1 \mu(t-1+\tau)d\tau, \int_0^1 \sigma^2(t-e+\tau)d\tau\right) \tag{2}$$

The variance of the distribution, described by the term $\int \sigma^2(t-1+\tau)d\tau$ is called *Integrated Variance*, a measure of ex-post volatility and is the object of interest.

As mentioned above, realized variance is the sum of squared realizations over the period of a trading day. That is, although in practice the data generating process is not continuous (prices are recorded at discrete and irregular intervals), sampling is often made by setting equidistant intervals. So, if a trading day had 8 hours that would, for instance, yield 96 five-minute intervals ($8 * {}^{60}/_{5}$), each one being represented by $p_{i,t}$, the ith observation of day t. The returns are thus $r_{i,t} = p_{t,i} - p_{t,i-1}$. Realized variance is then defined as

$$RV_t = \sum_{t=1}^{n} r_{i,t}^2 \tag{3}$$

As mentioned before, one of the problems that arise when working with volatility is that, contrary to financial asset returns, the actual values of variance of the time series are not directly observable. This problem is usually solved by imposing parametric structures on volatility or using stochastic models with data usually at daily frequency. Another alternative commonly employed is the usage of the volatility implied in derivatives' prices for a future horizon.

Merton (1980) proved that, at a sufficiently high sampling frequency, true volatility – Realized Volatility - could be measured by recording the prices of an asset and summing its intraday squared returns. Given the increasing availability of high quality transaction data, it is only natural that the literature turned to realized volatility for the construction of ex-post measures.

Ideally, a continuous trading environment would yield the aforementioned true volatility [Hansen and Lunde (2006)]. In practice, this ideal environment does not hold. Problems related to bid-ask spreads, infrequent and non-synchronous trading create microstructure noise, making realized volatility, at best, another estimator for the variance of time series, but releases us from the lagging feature of GARCH-type models.

In order to evaluate the economic value of realised volatility, Fleming et al (2003) discovered that an investor that uses a volatility-timing strategy would be willing to pay 50 to 100 basis points per year in order to use realized volatility estimators for the conditional covariance matrix. Note that realized covariances can be constructed in a similar fashion to that of realized variance, summing the cross-products of intraday asset returns. In another study, Bandi et al (2008), found that conditional mean variance investors would pay around 80 basis points per year to switch to realized variance estimates.

Furthermore, working with realized volatility brings up, despite its advantages, the problem of how frequently one should sample the asset prices in order to obtain the most accurate estimate for the variance. When large intervals are used, information about volatility is lost. On the other hand, if short intervals are used (note that, for most stocks, data is now available at the highest possible frequency, i.e., tick-by-tick), the already mentioned microstructure noise diminishes the quality of the estimate. In the present work, the approach followed by Diebold et al (2003) of calculating returns every 5 minutes was used, which yields 288 intraday observations, in the case of the foreign exchange markets, which is open 24 hours per day.

Merton (1980) proved that at a sufficiently high frequency, true volatility of an asset's return can be measured by summing up all intraday squared returns. As we now have tick-by-tick data available, the path to true measures of volatility would then be available. More recently, however, the econometric literature discovered that in real life the situation is not as simple as that, because of market frictions that make the price process a noisy signal. Controlling for this noise that arises especially when sampling at very high frequencies has become an object of thorough study. Initially, the literature adopted the approach of sampling prices every 15 or 30 minutes, in order to eliminate the noise that higher frequencies would contain. Nevertheless, this approach obviously leads to the loss of some information and, until this is quantified, the optimal sampling frequency cannot be obtained. There is then a trade off between the amount of information in the process and the cleanliness of the signal.

A more recent approach to true measures of volatility consisted of incorporating all the data available (tick-by-tick) regardless of how much noise it contained and produce the variance estimates. Afterwards, these estimates are decomposed into components of fundamental signal and microstructure noise, creating a noise-to-signal ratio that can then be an indicator of goodness of the estimates.

In reality, this microstructure noise relates to a variety of frictions contained in the mechanics of the trading process: bid-ask bounces, which are related to the discreteness of price changes, infrequency of trading, strategic component of the order flow, trade size, etc. All this factors make the price process a noisy signal.

Diebold et al (2003) concluded that using a 5-minute sampling scheme optimized the trade off between information and noise. Although this is not the most recent approach to treating realized volatility, the method was applied in the present dissertation, as the decomposition of the estimates into its components was not the main purpose of this work.

1.1 Different sampling schemes

In reality, the price process yields discrete and irregularly spaced observations. Therefore, the sampling scheme that yields the best realized volatility estimates is not straightforward to find. McAleer and Medeiros (2006) list several sampling schemes that can be applied:

1. **Calendar time sampling**, This is the most used sampling scheme in volatility literature. Here, the prices are sampled at fixed intervals, at 5 or 10 minutes, for example. Due to the irregular feature of the price process, data has to be constructed artificially (i.e. one may choose the last or the first observation of a 5 minute interval, a method denominated *previous tick*);

2. **Transaction time sampling**, where prices are gathered prices every nth transaction;

3. **Business time sampling**, where the sampling frequency is chosen such that $IV_{i,t} = \frac{IV_t}{n_t}$;

4. **Tick time sampling**, where prices are recorded as frequently as possible, that is, every price change.

2 ARCH MODELS

Typical structural models for financial time series model the conditional mean by an equation of the form

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_n x_n + u \tag{4}$$

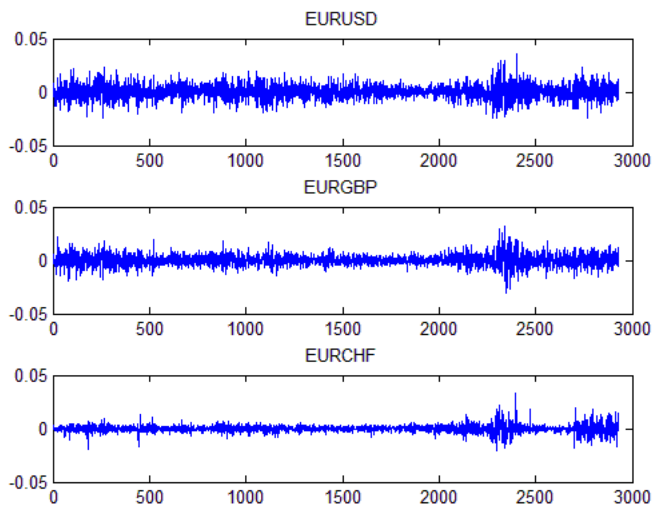
where $u \sim N(0, \sigma^2)$

The assumption is usually made that the variance of the error term, u, is constant, that is, that the errors are homoscedastic. In most financial time series, this assumption is violated, given that it is highly unlikely that the variance remains constant over the whole series, with the errors being heteroskedastic (that is, the variance of the errors is not constant).

When estimating a model for the conditional mean, if we assume that the errors are homoscedastic despite them being heteroskedastic, an implication would be that the estimates for the standard error could come out wrong. Therefore, considering a model that describes the behavior of the variance over time of a financial series seems useful.

In this context, Engle (1982) started a class of models called ARCH: Autoregressive Conditional Heteroscedasticity. These models not only fit the purpose of describing a time-varying variance process but also model an important and extensively documented property of volatility clustering. Clustering is the aggregation of this volatility by periods, that is, large returns in financial assets tend to be followed by large returns, with small returns being tipically followed by small returns. A visual examinantion of the figure below illustrates this "grouping" of volatility within certain time periods.

Figure 2: Foreign Exchange Returns: Volatility Clustering



The ARCH class of models fits the behavior described above. It assumes the form

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + ... + \alpha_q u_{t-q}^2$$

where $u_{t-q}^2$is the qth-lagged squared return. This is to say that the ARCH model, in simple terms, describes the variance of a financial time series as correlated with its previous realizations, in accordance with the observable fact that volatility is clustered.

Despite its apparent usefulness, ARCH models present some limitations: there is no best approach to identify the order of the model, that is, the number of lagged squared residuals to use. Even if there was a standardized procedure for deciding the order of the ARCH model, this value could be very large and that would result in a model that was not parsimonious. Also, the parameters of the model could return negative values, violating the non-negativity constraint imposed by its structure.

In the following lines a description of the models that are evaluated in this work will be given, along with their functional forms. All these models are derived from ARCH and treat volatility as a latent variable, in opposition to an observable one.

2.1 GARCH(p,q)

GARCH stands for Generalised Autoregressive Heteroskedasticity and has the following functional form:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + ... + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_p \sigma_{t-p}^2 \tag{5}$$

It was developed independentely by Bollerslev (1986) and Taylor (1986). As it is the case with the ARCH model described above, the GARCH model describes the conditional variance to depend on its previous own lags. Looking at the equation, it is straightforward to see that under this model the conditional variance is a function of a long-term average, as described by α_o, the lagged squared errors, as described by $\alpha_1 u_{t-1}^2, ..., \alpha_q u_{t-q}^2$and the lagged variance from the model, $\beta \sigma_{t-1}^2, ..., \beta_p \sigma_{t-p}^2$. The values for p and q give the order of the model, that is, the number of lags to be used both for the residuals (q) and for the conditonal variance (p).

The estimation of GARCH models cannot be made with the common Ordinary Least Squares procedure, given that its form is nonlinear. Instead, models of the GARCH family have to be estimated by maximizing a log-likelihood function.

2.2 GJR-GARCH(*p*,*q*)

One of the pitfalls of the standard GARCH models is that they impose a symmetric response of volatility both to positive and negative shocks. It is so because the variance, as specified in these models, is a function of the size of past realizations of squared errors. Therefore, the fact that a shock was negative is never taken into consideraton because the sign is lost. Nevertheless, it has been widely documented in the literature that negative shocks are likely to cause greater volatility than positive ones. This asymmetric effect is often termed as *leverage effect.* That is because it is assumed that when the value of a stock goes down, the firm's debt-to-equity ratio rises, leading shareholders to perceive their future cashflow stream as being more risky. Although this is an effect applicable only to equity markets, the term *leverage* is widely used even when dealing with volatility on other asset classes. The GJR-GARCH model was developed by Glosten, Jagannathan and Runkle (1993), and intended to be able to capture these asymmetries that standard GARCH models fail to describe. It has the functional form:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma u_{t-1} I_{t-1} \tag{6}$$

where $I_{t-1} = 1$if $u_{t-1} < 0$ and 0 otherwise.
This variable $I_{t-1}$is a dummy variable that captures the described *leverage effect*, that is, it assumes the value 1 when the previous error is negative. If γcomes out as being statistically significant, one can say that there is asymmetric response of volatility in that series.

2.3 EGARCH(*p*,*q*)

The EGARCH specification, which stands for Exponential GARCH, was proposed by Nelson (1991). Although there are many ways in which to express its functional form, one of them is:

$$ln(\sigma_t^2) = \omega + \beta ln(\sigma_{t-1}^2) + \gamma\frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha\left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right] \tag{7}$$

Note that with this specification, the natural logarithm of the time series variance is the object of study, instead of the variance itself, as it is modeled in the previous GARCH-type models discussed above. This feature of the EGARCH gives it the advantage of not needing artificially imposed non-negativity constraints. Also, it also accounts for the already mentioned *leverage effect*, given that theγ parameter will be negative when yhe ratio between the past return and the volatility is negative.

In its original specification, the EGARCH model was designed to be used with the Generalised Error Distribution (GED), although most of its applications are done with normal errors, given their computational ease.

Part III

Empirical Study

1 Research purpose and aims

In order to continue the ongoing work in volatility literature of perfecting its estimation, the present dissertation intends to perform an evaluation of three different parametric volatility models, namely the GARCH, EGARCH and GJR-GARCH structures. We aim to evaluate the predictive accuracy of different parametric volatility models, using realized volatility and squared returns as proxies for true volatility. Also, different residual distributions will be applied, namely the Gaussian, t-student and Generalized Error distributions. This assessment is performed using the Mean Squared Error criteria and the Diebold-Mariano test for predictive accuracy.

All in all, we pose the question of which of these three models yields the least noisy estimates of volatility, by comparing it to two measures of realized volatility.

It is expected that the EGARCH and GJR-GARCH specifications perform better at forecasting volatility, given that it is also expected that asymmetric responses of volatility to shocks of the same magnitude are present in the price process of the EURUSD exchange rate. This means lower values for the Mean Squared Error criteria when compared to the original GARCH specification and the rejection of the null hypothesis of the Diebold-Mariano test when testing EGARCH and GJR-GARCH against GARCH. No hypothesis is formulated *a priori* for the predictive accuracy of EGARCH relative to GJR-GARCH, since the two specifications were both designed to capture the asymmetric effect.

2 Data and Methodology

The data used for computing both Realized Volatility and for estimating the GARCH-type models in the previous section was obtained from a Bloomberg terminal. In the case of the former, 5-minute prices were gathered for the EURUSD exchange rate, representing the number of US dollars per Euro. The same procedure was adopted for the latter, only now daily prices were retrieved.

The sample involved data from January 3rd 2005 to December 31st, 2010. GARCH-type models estimation was achieved using data from January 3rd, 2005 to June 8th, 2010. The remaining days were used as an out-of-sample period, for which 9 series were forecasted, given that we are testing 3 models under 3 different residual distributions.

The out-of-sample period, for which 5-minute prices (and therefore returns) were retrieved, despite spanning through 934 days (including non-trading days), the period was substantially reduced to 523 days: for some days data was not fully available (aggregating a different number of returns would not be consistent) and on Fridays and Sundays the exchange rates do not trade for the whole 24-hour period (note that the Foreign Exchange market trades 24 hours per day from Monday to Thursday. On Fridays it closes at GMT 10pm, reopening on Sundays at the same time.). Non-trading days besides weekends, like target holidays, were also excluded.

The sample was adjusted as follows: non-trading days are automatically excluded by Bloomberg by selecting that option. Days that did not have 288 5-minute periods recorded were sorted out by using a PivotTable with Excel^{TM}software.

The figure below shows the arithmetic returns of the original series and the squared returns, as they are the measure of volatility used in GARCH-type models.

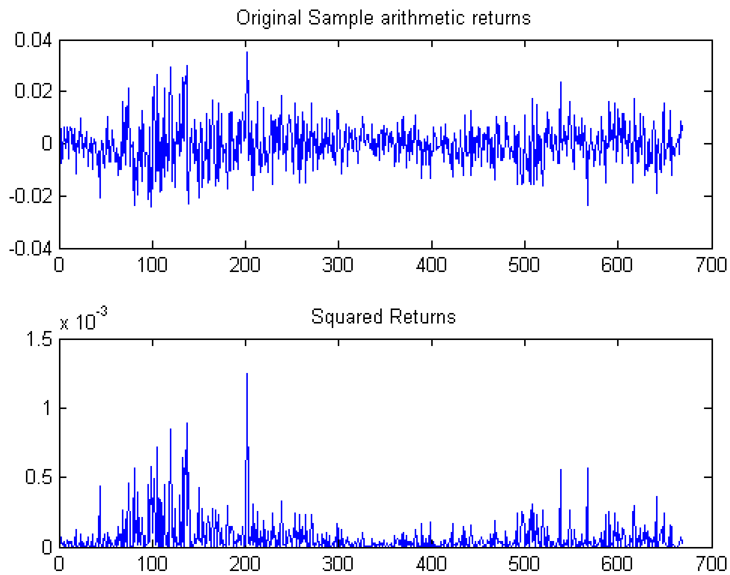


Figure 3: Returns from original series

Estimation of the GARCH, EGARCH and GJR-GARCH volatility models involved modeling the conditional mean in the first place. In order to accomplish this, the correlogram for returns was examined and the appropriate functional form was selected. Given that the series of returns appeared to follow a white noise process, there was no need to enforce an ARMA structure.

In order to evaluate the predictive ability of the models, the 5 minute realized volatility estimator was computed (summing 288 intraday squared returns), as well as the daily squared returns. This evaluation is performed by calculating the forecasted series for the out-of-sample period and then calculating their differences to the realized volatility estimator and daily squared returns.

Then, a straightforward Mean Squared Error is computed and the model/distribution yielding the lowest value comes out as the best forecasting form. Using the error vectors used to calculate the Mean Squared Error, we perform the Diebold-Mariano test as to assess if the differences between the models' predictive accuracies are statistically significant.

2.1 The Diebold-Mariano test

The test was developed by Diebold and Mariano (1995) and tests for the null hypothesis of equal predictive accuracy between the models, which are compared two at a time. Given that we are analyzing three different models, this will yield three Diebold-Mariano tests: GARCH(1,1) against EGARCH(1,1), GARCH(1,1) against GJR-GARCH(1,1) and EGARCH(1,1) against GJR-GARCH(1,1). This analysis will be performed using realized volatility and daily squared returns as proxies.

In this particular case, provided that we have an "observable" measure of volatility, each of these series will produce a vector of errors for each proxy. The vectors consist of the differences between each model's forecasts and the "actual" volatility, here measured by the aforementioned proxies.

The predictive ability of each model is evaluated by a loss function that can take several forms. Here, the loss function to be considered is the MSE, which consists of calculating the average of the square of the vectors ϵ^imentioned above. In notation,

$$MSE_i = \frac{1}{N}\sum_{t=0}^{T}(y^i - proxy^k) \tag{8}$$

where y^i is the forecasted series, $proxy^k$ the proxy for realized volatility with an index $k = 1$, 2 for realized volatility and daily squared returns respectively.N is the number days between t and T, which represent the beginning and the end of the sample period. To calculate the test statistic N consisted only on the number of observations of the adjusted sample.

The test is then based on the difference between the loss function for each model, conventionally denoted d_t.

The hypothesis to be tested are

$$H_0 : E[d_t] = 0$$

$$H_1 : E[d_t] \neq 0$$

and the test statistic, which Diebold and Mariano (1995) proved to follow a Gaussian distribution is

$$S = \frac{\overline{d}}{\sqrt{L\hat{R}V_{\overline{d}}/T}}$$

where

$$\overline{d} = \frac{1}{N}\sum_{t=0}^{T} d_t$$

$$LRV_{\overline{d}} = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j, \gamma_j = cov(d_t, d_{t-j})$$

In case of rejection of the null hypothesis, if the test value is positive that means that the loss function associated with the first model under comparison is greater, making the second model a better predictor. If the test value is negative, the first model under comparison yields better forecasts.

3 Distributional properties of returns and realized volatility

In this section, descriptive statistics of returns and volatility of the series under study will be presented. In particular, the analysis will focus on raw returns, returns standardized by realized standard deviation (the square root of realized volatility), realized volatility itself and the natural logarithm of realized volatility. For each of these series, a table with the mean, the sample standard deviation, skewness, kurtosis, range and the p-value for the Jarque-Bera normality test. Other tests could be performed to test normality. The Jarque-Bera test was chosen on the basis of standardness. Also, *Eviews* histogram plots will be presented.

3.1 Unconditional distributions

3.1.1 Return and standardized return

In Table 1 below are summarized the descriptive statistics mentioned above for returns and returns standardized by the squared root of realized volatility.

	Mean	**Sample STD**	**Kurtosis**	**Skewness**	**Minimum**	**Maximum**	**p-value JB**
r_t	$-0.2E-3$	0.008	4.156	0.270	-0.024	0.035	0.000
r_t/σ_t	-0.039	1.078	3.576	0.202	-4.165	3.834	0.0045

Table 1: Unconditional distribution of return and standardized return

From the results above, conclusions can be drawn regarding the unconditional distributions of the considered series. For r_t, the empirical distribution is *leptokurtic* and *positively skewed*, since kurtosis is above 3 (kurtosis value for the normal distribution), meaning that tails are fatter. Skewness is above 0 (skewness value for the normal distribution), meaning that the right tail is longer. Fatter tails, or *leptokurtosis*, imply that extreme observations that would occur with a certain probability under the normal distribution, now occur with a higher one. In other words, one could say that, empirically, the market attributes probabilities to extreme events higher than one would expect under a normal distribution. Positive skewness indicates that the occurrence of observations to the right of the mean was higher than to the left. Since the mean is close to 0, it is straightforward to infere that, for the sample under consideration, great positive returns occured more frequentely than negative ones. The p-value for the Jarque-Bera test is 0 to the third decimal place, a value that indicates that the empirical distribution of returns is clearly non-normal.

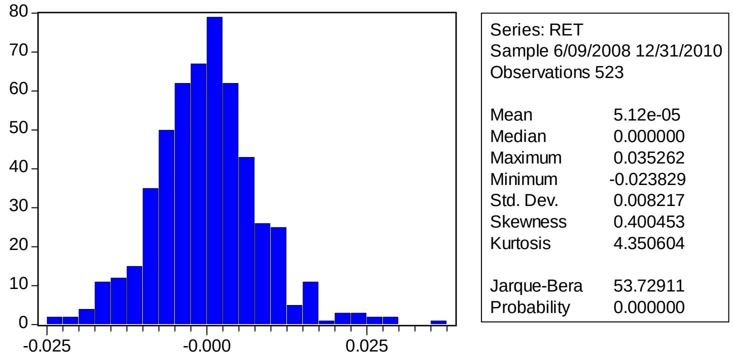


Figure 4: Histogram and summary of statistics for r_t

For r_t/σ_t, the series of standardized returns, fatter tails (*leptokurtosis*) and positive skewness are maintined, indicating results similar to those of raw returns, although it appears that after standardization, returns are less deviated from normality. Nevertheless, the p-value for the Jarque-Bera normality test still indicates that the series is not normal, even for low significance levels (that is, normality would still be rejected at 10%, for example). This result is consistent with the ones obtained by Andersen et al (2001) and Andersen et al (2003). The distributions of returns standardized by the realized standard deviation have been documented to be often less *leptokurtic* than those of returns standardized by the ARCH estimate [see, for example, Bollerslev et al, (1994) and Andersen and Bollerslev (1998)].

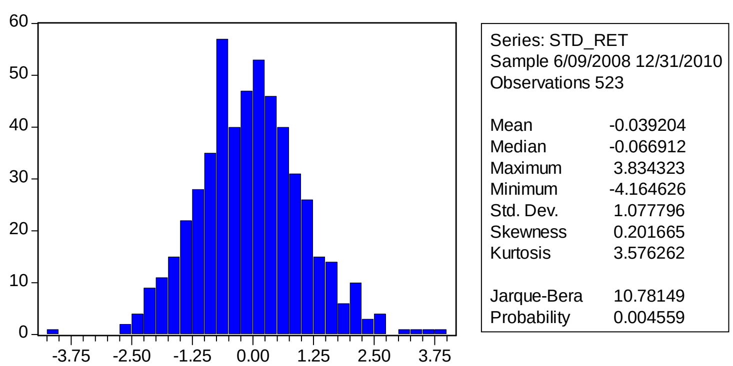


Figure 5: Histogram and summary of statistics for r_t/σ_t

3.1.2 Realized volatility and logarithmic realized volatility

Table 2 presents the same descriptive statistics as Table 1, only now for realized volatility and its natural logarithm.

	Mean	**Sample STD**	**Kurtosis**	**Skewness**	**Minimum**	**Maximum**	**p-value JB**
RV_t	$7.07E-5$	$6.22E-5$	14.507	2.804	$9.53E-6$	$0.56E-3$	0.000
$ln(RV_t)$	-9.813	0.676	3.028	0.593	-11.561	-7.494	0.000

Table 2: Unconditional distribution of realized volatility and its natural logarithm

For RV_t, the series is highly *leptokurtic* and rightly (positively) skewed. The values for kurtosis and skewness [(14.507) and (2.804)] are strongly deviated from those of the normal distribution. The p-value for the Jarque-Bera test clearly indicates the rejection of normality.

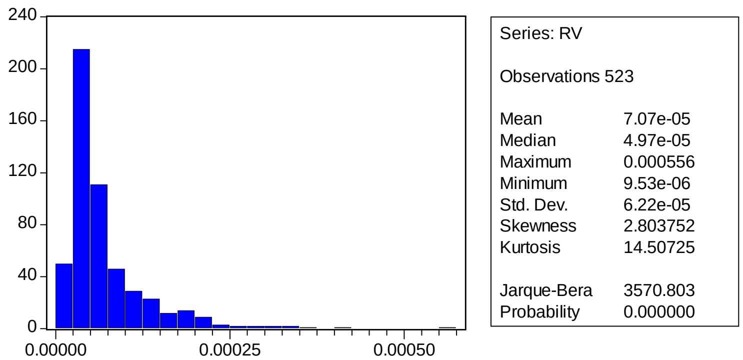


Figure 6: Histogram and summary of statistics for the logarithm ofσ_t^2

For $ln(RV_t)$, the values for kurtosis and skewness are substantially lower and much closer to those of the normal distribution. It is noteworthy that these are much inferior than those of the RV_t series. Notwithstanding, the p-value for the Jarque-Bera also indicates the rejection of normality. This result is consistent with those obtained in volatility literature.

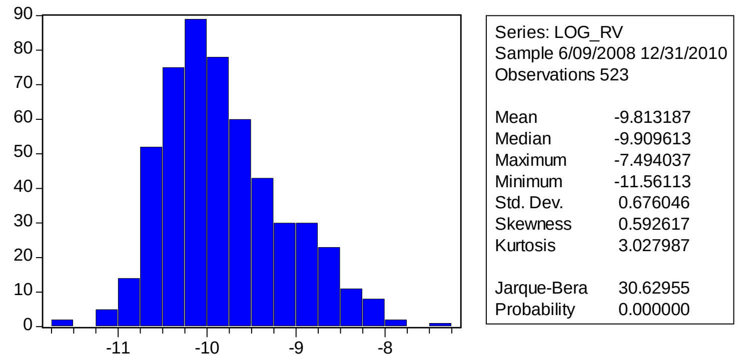


Figure 7: Histogram and summary of statistics for σ_t^2

3.2 Conditional distributions

The conditional distribution of volatility in financial assets has been object of extensive investigation in the last decades. In this section, an analysis on the $ln(\sigma_t^2)$ will be performed, regarding the properties of long memory, temporal dependence and asymmetric effect.

3.2.1 Temporal dependence and long memory

One of the stylized facts of returns is the temporal dependence of volatility. The literature on volatility has extensively documented that this dependence is highly persistent, as it can be seen in the figures below. The first figure illustrates the series of the logarithm of realized volatility. Its visualization leads us to confirm that it is clustered. The second figure is the correlogram for the series until the 25th lag. From the correlogram we can conclude that autocorrelation is statistically significant for these lags and presents a hiperbolic decay to 0.

These charecteristics are coherent with the features of long memory processes.

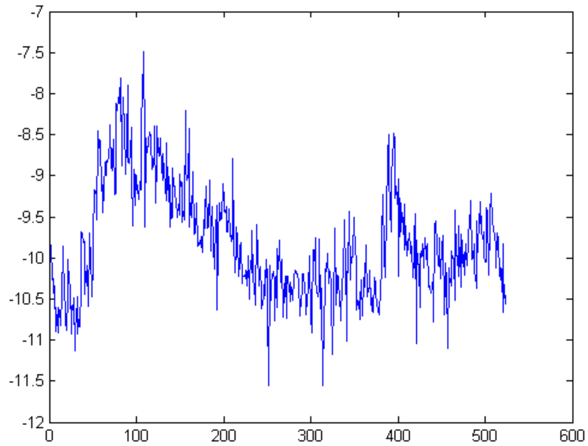
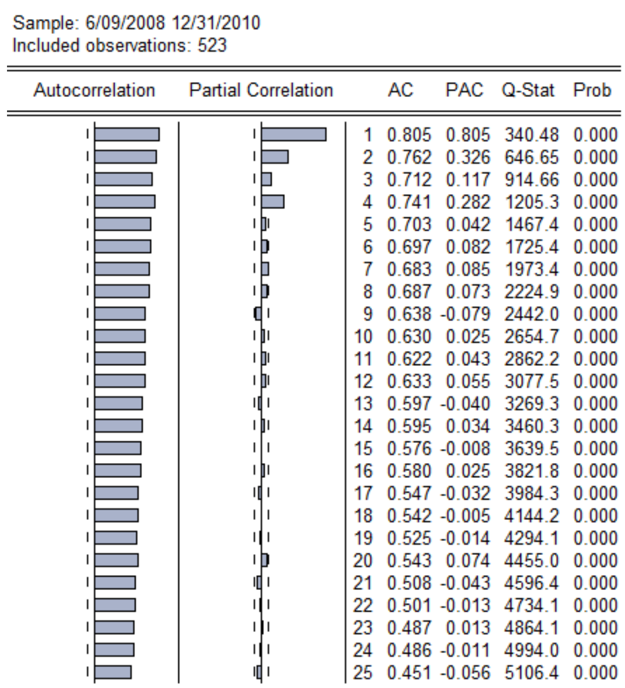


Figure 8: Logarithm of Realized Volatility



Sample: 6/09/2008 12/31/2010
Included observations: 523

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.805	0.805	340.48	0.000
		2	0.762	0.326	646.65	0.000
		3	0.712	0.117	914.66	0.000
		4	0.741	0.282	1205.3	0.000
		5	0.703	0.042	1467.4	0.000
		6	0.697	0.082	1725.4	0.000
		7	0.683	0.085	1973.4	0.000
		8	0.687	0.073	2224.9	0.000
		9	0.638	-0.079	2442.0	0.000
		10	0.630	0.025	2654.7	0.000
		11	0.622	0.043	2862.2	0.000
		12	0.633	0.055	3077.5	0.000
		13	0.597	-0.040	3269.3	0.000
		14	0.595	0.034	3460.3	0.000
		15	0.576	-0.008	3639.5	0.000
		16	0.580	0.025	3821.8	0.000
		17	0.547	-0.032	3984.3	0.000
		18	0.542	-0.005	4144.2	0.000
		19	0.525	-0.014	4294.1	0.000
		20	0.543	0.074	4455.0	0.000
		21	0.508	-0.043	4596.4	0.000
		22	0.501	-0.013	4734.1	0.000
		23	0.487	0.013	4864.1	0.000
		24	0.486	-0.011	4994.0	0.000
		25	0.451	-0.056	5106.4	0.000

Figure 9: Correlogram of $ln(\sigma_t^2)$

To further analyze the series, the Augmented Dickey Fuller test for 18 lags and an intercept was applied, in order to assess the stationarity of the series. The results are shown below.

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.137891	0.0245
Test critical values:	1% level	-3.442722	
	5% level	-2.866889	
	10% level	-2.569680	

*MacKinnon (1996) one-sided p-values.

Figure 10: ADF test for $ln(\sigma_t^2)$

The test rejects the existence of a unit root in the series for a 5% significance level [(-3.138) (0,0245)]. We can thus conclude that the series is stationary.

3.2.2 Asymmetric effect

An asymmetric response of volatility to shocks is often observed in financial markets. Usually, volatility tends to show higher increases when reacting to bad news (negative shocks) than to good news. Regular GARCH models enforce symmetric responses of volatility to positive and negative shocks, as it is a function of lagged observations but not their signs. A more thorough explanation of how this asymmetries are incorporated into functional forms in order to account for this empirical regularity of volatility shall be given later.

A visual examination of the figure below illustrates this effect. Negative observations would fit a steeper line than positive ones, which means that for returns of the same magnitude but of different signs, volatility is greater when observations are negative. This effect was firstly documented by Black (1976), under the name of News Impact Curve. Clearly, asymmetry can be tested for with tests. Engle and Ng (1993) proposed a set of tests to infere if the symmetric GARCH model adequately fits the data to be modeled or if an asymmetric form shall be used. For the purpose of the present work, a visual examination of the data suffices, since the objective here is to assess the predictive power of GARCH-type models.

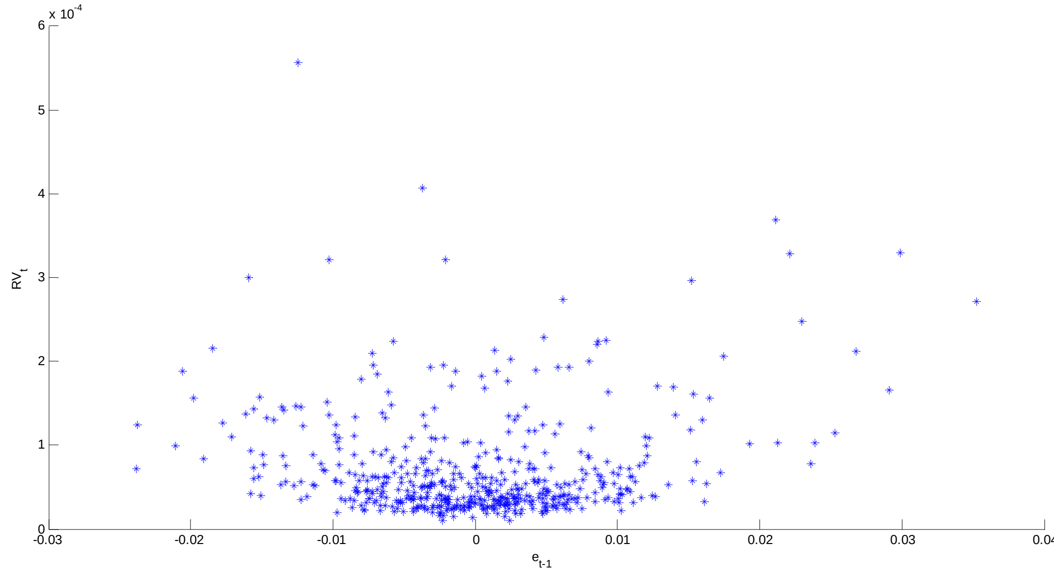


Figure 11: Asymmetry in EURUSD volatility

4 Results

In this section, we present the results for accuracy and comparisons between models. We evaluate the Mean Squared Error for each model, error distribution and volatility proxy. Also, Table 4 lists the Diebold-Mariano test values and p-values for the aforementioned comparisons. Here we will confront the results with the hypothesis in Section 2 of the present work, where we stated that it would be expected that models aiming at capturing asymmetric responses of volatility yield a better performance at forecasting. Also, we stated that applying heavier tailed distributions would allow us to account for the *leptokurtosis* usually present in financial assets' returns and therefore gain in predicting volatility.

DISTRIBUTION/PROXY	MODEL	MSE
Gaussian/RV	GARCH(1,1)	1,15E-09
Gaussian/RV	EGARCH(1,1)	5,50E-10
Gaussian/RV	GJR-GARCH(1,1)	1,18E-09
t-student/RV	GARCH(1,1)	2,20E-09
t-student/RV	EGARCH(1,1)	2,90E-09
t-student/RV	GJR-GARCH(1,1)	2,22E-09
GED/RV	GARCH(1,1)	2,21E-09
GED/RV	EGARCH(1,1)	2,91E-09
GED/RV	GJR-GARCH(1,1)	2,22E-09
Gaussian/RET SQ	GARCH(1,1)	1,40E-08
Gaussian/RET SQ	EGARCH(1,1)	1,44E-08
Gaussian/RET SQ	GJR-GARCH(1,1)	1,41E-08
t-student/RET SQ	GARCH(1,1)	1,41E-08
t-student/RET SQ	EGARCH(1,1)	1,45E-08
t-student/RET SQ	GJR-GARCH(1,1)	1,41E-08
GED/RET SQ	GARCH(1,1)	1,41E-08
GED/RET SQ	EGARCH(1,1)	1,45E-08
GED/RET SQ	GJR-GARCH(1,1)	1,41E-08

Table 3: Mean Squared Error

Contrary to the intuition stated in Section 2, it is not clear that models designed to capture the asymmetric effect that is usually present in financial assets' returns. Looking at Table 3, the conclusion is that, according to the Mean Squared Error criteria, the simple GARCH(1,1) specification dominates all other specifications regarding predictive ability, even when considering different distributions (Gaussian, t-student and GED) for residuals and the two proxies (realized volatility and daily squared returns) for volatility. An exception is made when using the t-student and the GED distributions and squared returns as a proxy, where GARCH and GJR-GARCH have the same value for the loss function

This result is in line with the work of Hansen and Lunde (2005), where the authors perform a comparison between 330 ARCH-type models and conclude that, for exchange rates, a GARCH(1,1) specification overperforms all the other ones. The only exception seems to be when models are evaluated under a Gaussian error distribution and using realized volatility as the proxy for actual volatility. Here, the EGARCH specification seems to perform better when predicting future volatility.

Note that we are analyzing data for the EURUSD exchange rate and, although Figure 8 suggests asymmetric volatility in this exchange rate, it is obvious that this asymmetry cannot be attributed to the leverage effect that relates sotck prices' responses to firms debt-to-equity ratios. So, even though we have some evidence of unequal responses for the variance of the exchange rate, these do not seem to be significant enough to justify switching to a more sophisticated GARCH specification, disregarding the already mentioned exception.

It is also noteworthy that Mean Squared Error values are much more divergent when switching from gaussian error distributions to one of the remaining two (t-student or GED) than between t-student and GED themselves. This result seems natural, given that these two distributions allow for heavier tails than normal. Nevertheless, although one would expect that residual distributions that account for fatter tails would ultimately result in better forecasting structures, using Mean Squared Error as the loss function leads us to conclude that Gaussian errors yield better forecasts.

The table below summarizes the results for the Diebold-Mariano tests. In total, 18 tests were performed in order to account for all possible combinations between parametric models, error distributions and volatility proxies.

DISTRIBUTION/PROXY	**TEST**	**TEST VALUE**	**P-VALUE**
Gaussian/RV	GARCH vs. EGARCH	9,961	0,000
Gaussian/RV	GARCH vs. GJR	-3,288	0,001
Gaussian/RV	EGARCH vs. GJR	-9,773	1,47E-22
t-student/RV	GARCH vs. EGARCH	-6,126	9,02E-10
t-student/RV	GARCH vs. GJR	-1,987	0,047
t-student/RV	EGARCH vs. GJR	6,018	1,77E-09
GED/RV	GARCH vs. EGARCH	-6,125	9,08E-10
GED/RV	GARCH vs. GJR	-2,038	0,042
GED/RV	EGARCH vs. GJR	6,02	1,75E-09
Gaussian/SQ RET	GARCH vs. EGARCH	-1,984	0,047
Gaussian/SQ RET	GARCH vs. GJR	-1,238	0,216
Gaussian/SQ RET	EGARCH vs. GJR	1,694	0,090
t-student/SQ RET	GARCH vs. EGARCH	-2,129	0,033
t-student/SQ RET	GARCH vs. GJR	-0,8671	0,385
t-student/SQ RET	EGARCH vs. GJR	2,074	0,038
GED/SQ RET	GARCH vs. EGARCH	-2,143	0,032
GED/SQ RET	GARCH vs. GJR	-0,8887	0,374
GED/SQ RET	EGARCH vs. GJR	2,089	0,036

Table 4: Diebold-Mariano test results

Here, test results corroborate those stated before (see Appendix for plotted series). Grouping the models by distribution and volatility proxy and, considering only the cases where the null hypothesis is rejected under the 5% significance level, we conclude that the GARCH specification dominates all the other ones, followed by GJR-GARCH, with the EGARCH structure yielding the worst forecasts. The only exception seems to be when considering a Gaussian error distribution and realized volatility as a proxy, where the EGARCH performs better than all the other ones, followed by the GARCH specification and leaving GJR-GARCH as the noisiest model.

Regarding the cases where the null hypothesis is not rejected, that is, where the Diebold-Mariano test suggests that both models have the same predictive capacity, this seems to happen only when comparing GARCH with GJR and using daily squared returns as a proxy for volatility, regardless of the distribution.

For all other specifications, the results dictate which model performs better under different conditions.

5 Conclusions

The present dissertation aimed to perform an evaluation of three different parametric structures of volatility by adopting the relatively recent non-parametric approach of treating volatility as an observable variable, obtained by the sum of high frequency intraday financial returns.

After an introduction to provide context, realized volatility is explained in more detail and its conditional and unconditional distributional properties are documented, as well as for returns. The empirical distribution of realized volatility is clearly non-normal, given its high kurtosis and positive skewness. Notwithstanding, the distributions for logarithmic realized volatility and returns standardized by realized standard deviation seem to be close to normal. Raw returns are also clearly non-normal. Furthermore, realized volatility seems to have an asymmetric behavior. These results are consistent with the ones found in the literature.

As for the predictive ability of the GARCH, EGARCH and GJR-GARCH models, the GARCH specification seems to dominate the others for the majority of the cases . Although the asymmetric effect of volatility is visible in Figure 11, the only specification that was not designed to capture this leverage specifically seems to perform better than the others. This could mean that the *leverage* effect, traditionally attributed to the debt-to-equity ratio of firms, is not as meaningful when considering exchange rates. Furthermore, between the EGARCH and GJR-GARCH models, the latter appears to be a better predictor of volatility except for the case where we use a Gaussian error distribution and 5 minute return aggregation as a proxy for volatility.

Further work can be done by extending this analysis to other exchange rates, either one at a time or simultaneously, performing a multivariate analysis. An analysis considering different sampling schemes and intervals would also be of interest, as well as using different distributions in the estimation of the GARCH models.

6 Limitations and further work

The present dissertation tries to assess the predictive accuracy of three models of the GARCH family by using different proxies for realized volatility. Although realized volatility (computed as the sum of 5 minute intraday squared returns) would ideally yield the true value of volatility, due to market frictions that arise from its structure it is only another estimator for volatility. Nevertheless, results in the literature suggest that it seems to deliver more reliable estimates than the traditional parametric volatility models.

We compare the forecasts from parametric models to an approximation of what would be the actual volatility, following the apporach of Diebold et al (2003). These authors suggest that extracting 5 minute prices yield the best trade off between noise and information in the series. This could be seen as a limitation of the present study, since other conclusions could be drawn about optimal sampling schemes. In fact, there are several authors who developed these alternative sampling techniques that necessarily yield different estimates for realized volatility. For instance, Hansen and Lunde (2006b) proved that the previous tick method, which involves selecting the price for the first transaction in a given interval, is an adequate approach to sample prices.

Although some adjustments could be made to the methodology, the main results of this study are not expected to be substantially different, as results in previous literature already document that the GARCH specification seems to perform better than other models when considering exchange rate volatility.

For further work, this assessment could be extended to other exchange rates or even other asset classes as stocks or bonds, as it would be interesting to evaluate the predictive ability of parametric models for interest rate volatility.

References

[1] Andersen, T.G. (1994), Stochastic autoregressive volatility: a framework for volatility modeling, *Mathematical Finance* 42, 75-102.

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

[3] Andersen, T. G.,Bollerslev, T. and Diebold, F. X. (2003), Parametric and Nonparametric Volatility Measurement, *Handbook of Financial Econometrics*, North-Holland Press.

[4] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001a), The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43–76.

[5] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), Modeling and forecasting realized volatility, *Econometrica* 71, 529–626.

[6] Bandi, F., Russell, J. R. and Zhu, Y. (2008), Using high-frequency data in dynamic portfolio choice, *Econometric Reviews* 27, 163–198.

[7] Barndorff-Nielsen, O. E. and Shephard (2002), Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal Of The Royal Statistical Society Series B* 64(2), 253-280.

[8] Black, F. (1976), Noise, *Journal of Finance* 41, 529–543.

[9] Black, F., (1976), Studies of stock market volatility changes. Proceedings of the American Statistical Association, *Business and Economic Statistics Section*, 177–181.

[10] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31, 307-27.

[11] Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994), ARCH Models, *Handbook of Econometrics* IV, 2959-3038, North-Holland Press.

[12] Christie, A.A., (1982), The stochastic behavior of common stodck variances, *Journal of Financial Economics* 10, 407–432.

[13] Diebold, F. X. and Mariano, R.S. (1995), Comparing Predictive Accuracy, *Journal of Business & Economic Statistics* 13, 253-263.

[14] Engle, R. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50, 987-1007.

[15] Engle, R., Lilien, D. and Robins, R. (1987), Estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M model, *Econometrica* 55, 251-27.

[16] Engle, R. F. and Ng, V. K. (1993), Measuring and Testing the Impact of News on Volatility, *Journal of Finance* 48(5), 1749-78.

[17] Fleming, J., Kirby, C., Ostdiek, B. (2003), The economic value of volatility timing using "realized volatility", *Journal of Financial Economics* 67, 473–509.

[18] French, K. R., William Schwert, G. and Stambaugh R. F. (1987), Expected Stock Returns and Volatility, *Journal of Financial Economics* 19, 3-29.

[19] 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, 1779-180.

[20] Hansen, P. R., Lunde, A. (2005a), A forecast comparison of volatility models: Does anything beat a GARCH(1, 1)?, *Journal of Applied Econometrics* 20, 873–889.

[21] Hansen, P. R., Lunde, A. (2006b), Realized variance and market microstructure noise (with discussion), *Journal of Business and Economic Statistics* 24, 127–218.

[22] Merton, R. C. (1980), On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

[23] McAleer, M. and Medeiros, M. C. (2006), Realized volatility: a review, *Textos para discussão* 531, Department of Economics PUC-Rio.

[24] Nelson, D.B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59, 347-370.

[25] Taylor, S. J. (1986), *Modelling Financial Time Series*. Chichester: Wiley.

Part IV
APPENDIX

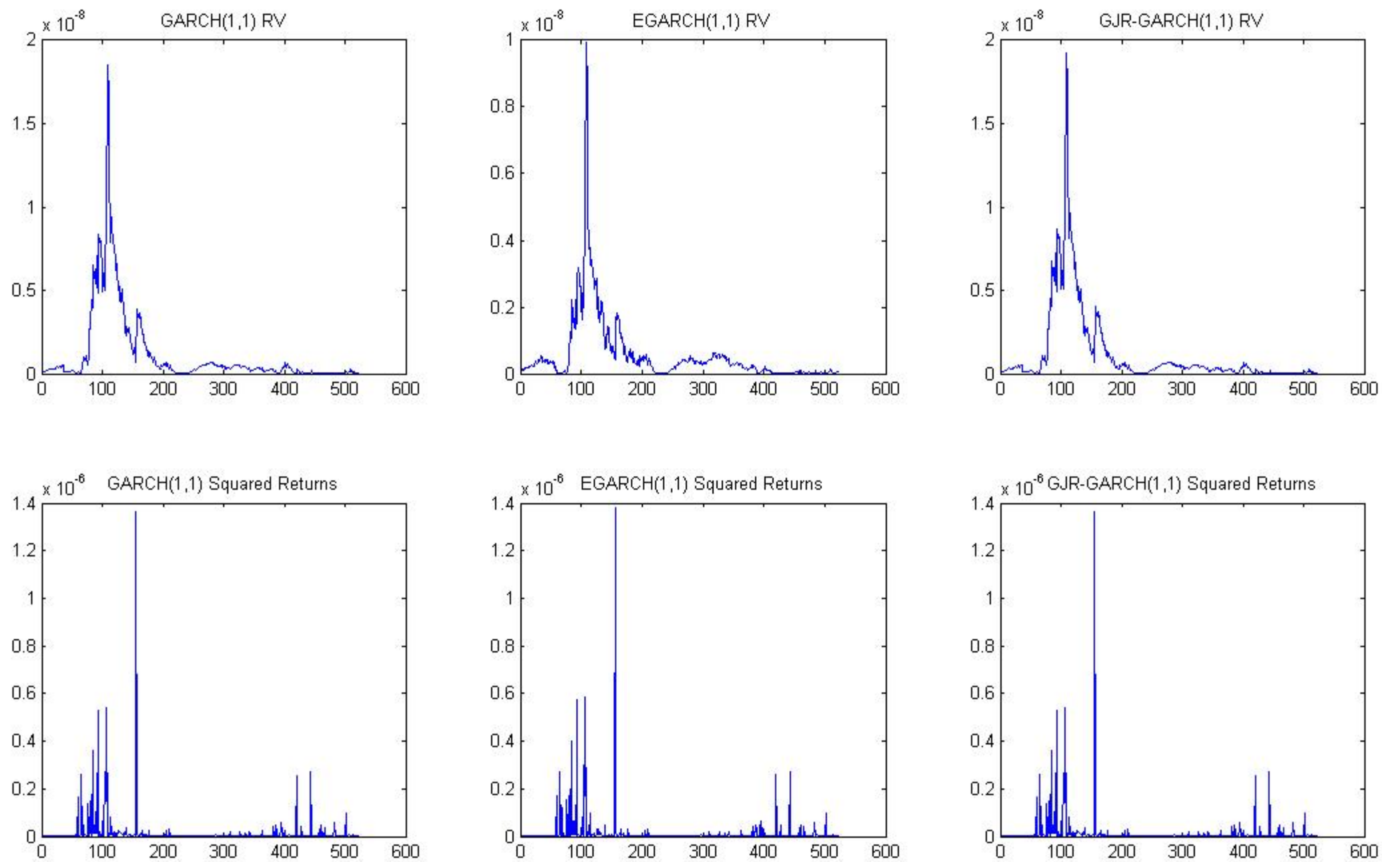


Figure 12: Squared errors towards RV and daily squared returns (Gaussian residuals)

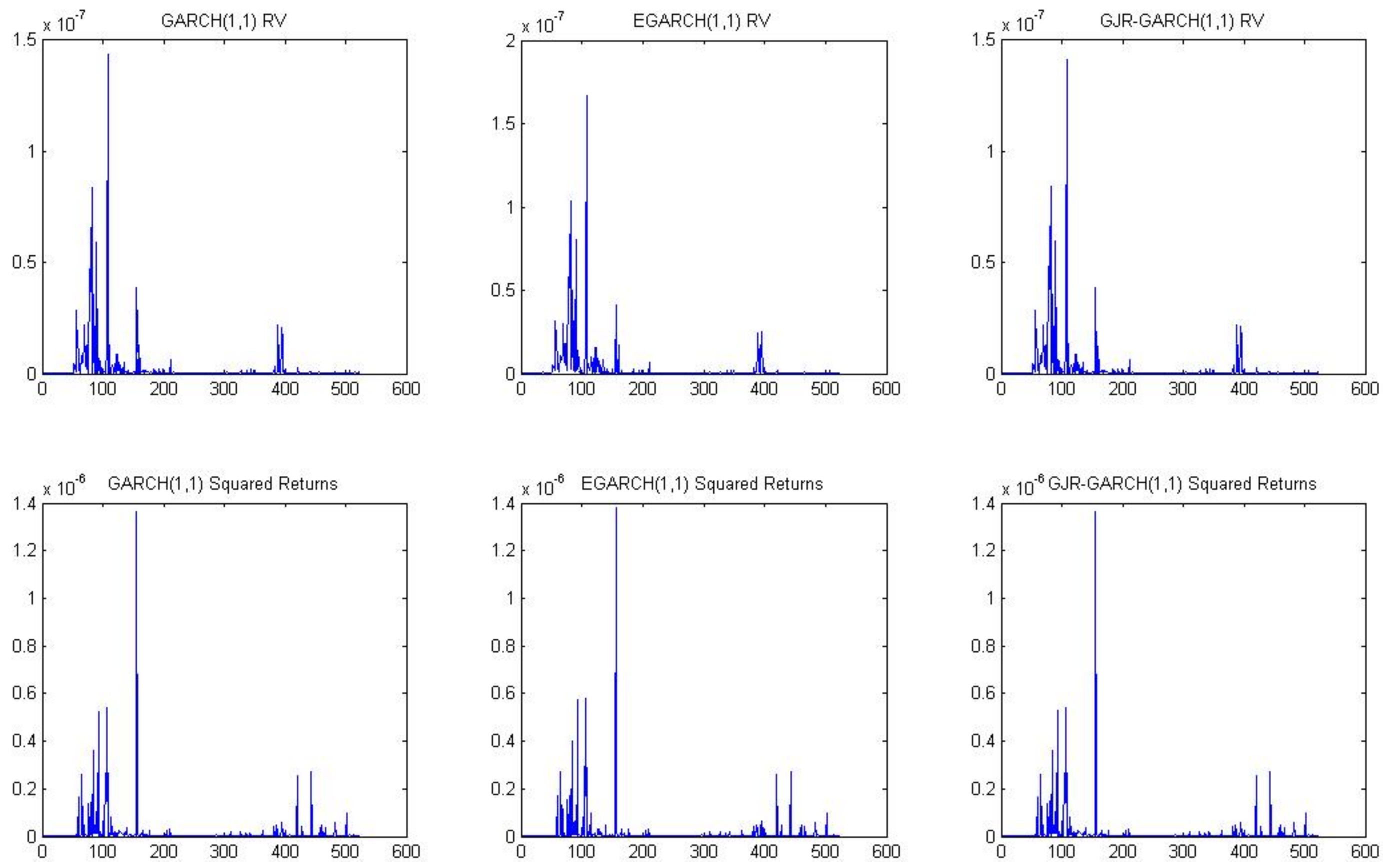


Figure 13: Squared errors towards RV and daily squared returns (t-student residuals)

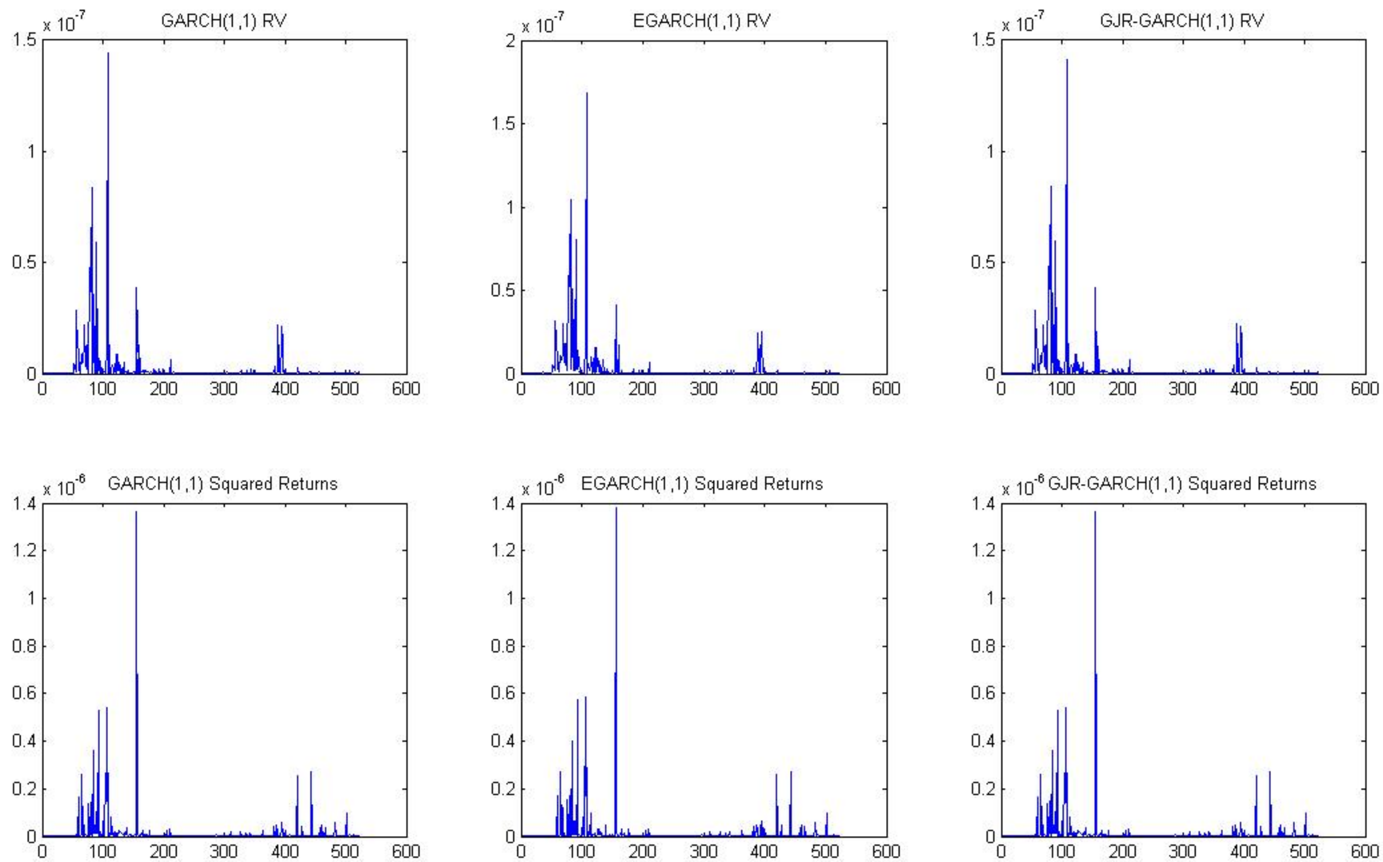


Figure 14: Squared errors towards RV and daily squared returns (GED residuals)