What best predicts realized and implied volatility: GARCH, GJR or FCGARCH?

José Salgado

Prof. Doutor José Dias Curto

April 2011

Thank-you note:

I am very grateful to Prof. Doutor José Dias Curto for all his active participation and support.

ABSTRACT

This thesis focuses on forecasting realized volatility (RV) and implied volatility (IV) on equity markets, a subject of major importance for volatility traders. The accuracy of IV and GARCH-type models to predict RV has been researched extensively. However, little work has been done to model IV.

We test the accuracy of GARCH-type models (GARCH, GJR and FCGARCH) to forecast, one-day ahead, the VIX index (the chosen IV measure) and the S&P500 index's daily realized volatility. While futures on equity's IV are widely available, futures on RV appeared recently on foreign exchange markets. Yet, expansion to equity markets is expectable. Thus, this study is a first step on developing a RV and IV futures trading strategy.

From 2001 to 2010 the models were estimated based on daily data. Forecasts evaluation is based on the mean absolute error criteria and Diebold-Mariano test. We found the GJR/FCGARCH models to have the best performance on both RV and IV. From the results, one can also infer that GARCH-type models are more suitable to foresee IV than RV. A plausible deduction is that past returns and past variance have a higher impact on IV.

Keywords: Forecasting, Realized volatility, Implied volatility, GARCH models, Multiple regimes

JEL: C22, C52, C53

RESUMO

Esta tese centra-se na previsão de volatilidade realizada e volatilidade implícita nos mercados de capitais, um assunto de grande importância para os "traders" de volatilidade. A precisão de modelos GARCH para prever a volatilidade realizada tem sido estudada extensivamente. No entanto, pouco tem sido feito para modelar volatilidade implícita.

Nós testámos a precisão de modelos GARCH (GARCH, GJR e FCGARCH) para prever, com um dia de antecedência, o índice VIX (a medida de volatilidade implícita escolhida) e a volatilidade diária do S&P500. Apesar de futuros sobre volatilidade implícita estarem amplamente disponíveis, os futuros sobre volatilidade realizada só apareceram recentemente nos mercados cambiais. No entanto, a expansão para os mercados de capitais é expectável. Assim, este estudo é um primeiro passo no desenvolvimento de uma estratégia de "trading" de futuros sobre volatilidade realizada e volatilidade implícita.

De 2001 a 2010, os modelos foram estimados com base em dados diários. A avaliação das previsões é baseada no critério do erro médio absoluto e no teste Diebold-Mariano. A conclusão é que os modelos GJR / FCGARCH prevêem melhor quer a volatilidade realizada quer a volatilidade implícita. A partir dos resultados, pode-se, também, inferir que os modelos de tipo GARCH são mais adequados para prever volatilidade implícita do que volatilidade realizada. Uma dedução plausível é que os retornos passados e a variância passada têm um maior impacto sobre volatilidade implícita.

Palavras-chave: Previsão, Volatilidade realizada, Volatilidade implícita, Modelos GARCH, Múltiplos regimes.

JEL: C22, C52, C53

Contents

List of Tables

“ La détermination de ces mouvements (la Bourse) se subordonne à un nombre inifini de facteurs: il est dès lors impossible d'en espérer la prévision mathématique. (...)

Mais il est possible d'étudier mathématiquement l'état statique du marché à un instant donné (...) le coefficient d'instabilité ou de nervosité de la valeur, c'est lui qui mesure son état statique. Sa tension indique un état d'inquiétude; sa faiblesse, au contraire, est l'indice d'un état de calme.”

Louis Bachelier (1990), *Theorie de la Speculation*, probably the first seeking to define volatility

1 INTRODUCTION

Stepping back in time, Louis Bachilier - probably the first one attempting to define volatility in finance - referred in 1990 the impossibility to predict stock returns. In the recent decades, probably due to the severe impact of financial turmoils, both the academy and the industry started to disregard the forecastability of asset returns while the interest on market risk and uncertainty expanded (Curto *et al.*, 2009). Thus, the importance of market risk, i.e., uncertainty about the market price of a financial asset, sharpened the research on modeling and forecasting volatility.

It is widely agreed that financial volatility is critical to finance in areas such as asset pricing, asset allocation and risk analysis. *“The price of essentially every derivative security is affected by swings in volatility”* and “*risk management models used by financial institutions and required by regulators take time-varying volatility as a key input”* (Brownlees *et al.*, 2010, p. 2). The concept of *le coefficient d'instabilité ou de nevorsité* first discussed by Bachelier (1990), now defined as volatility, became a cornerstone of finance.

Using as reference Majmudar and Banerjee (2004), volatility in the field of finance can be grouped into three categories: realized volatility (RV): the standard deviation of asset returns; implied/market volatility: deduced from market prices of derivatives such as options. Assuming that market prices correctly reflect agents' expectations, implied volatility (IV) represents the market prediction about future price fluctuations. The measure of IV used in this work is the VIX index; model volatility: estimated by the theoretical volatility models such as GARCH - Generalized Autoregressive Conditional Heteroskedasticity - and Stochastic volatility. These are latent volatility models because volatility is not directly observable; though they can be estimated with the aim of predicting future volatility.

Each of these types of volatility can be calibrated according to periodicity (daily, monthly, etc.) and to sample period. Both RV and IV can be backwards looking (based on past returns or market prices) or forward looking (the ones to be foreseen).

In the recent decades, a vast literature has been developed in the field of forecasting RV. The pervasive volatilities across and within asset classes during turmoils and the emergence of volatility as an asset class reinforced the need to improve the forecasting tools triggering a vast group of GARCH (discrete-time) and Stochastic (continuous-time) type models.

The forecasting engines that we have decided to apply are GARCH-type models, letting the application of other models for future research. Therefore, GARCH volatility deserves a brief presentation.

Until the eighties of the last century, the variance of the error - the terms error, innovation, disturbance, shock and news will be used interchangeably - in regression financial models was assumed to be constant (homoskedasticity). The idea of financial time series volatility that vary with time, first presented by Mandelbrot (1963), was concretized by the ARCH - Autoregressive Conditional Heteroskedasticity - model proposed in Engle (1982). The ARCH model allowed a time-dependent heteroskedasticity distribution for the asset returns. It tries to capture two stylized facts in finance, i.e., characteristic features of financial series, namely: changing volatility, volatility clustering - the terms volatility clustering, dynamics and persistence will be used interchangeably - and . Volatility persistence is ubiquitous in financial asset returns due to the clear evidence of contiguous periods of high or low volatility. The model proposed by Engle introduced a stationary - assumes that unconditional volatility remains unchanged through time -, parametric and conditional approach to forecast the second moment of the financial asset returns distribution based on the size of past error terms. Bollerslev (1986) introduced the GARCH model, which is a more parsimonious model as it reduces the memory needed.

Although these models are widely used to forecast conditional variances, they are unable to incorporate another stylized fact of returns: the "leverage effect" on volatility; an asymmetric news impact curve that Black (1976) first observed, i.e., a sequence of unexpected bad news or negative innovations is expected to have a higher impact on volatility than a sequence of positive news with the same size. Among many others, GJR - Glosten, Jagannathan, Runkle - (1993) have managed to include asymmetric effects on stock returns' volatility.

Later, model specifications with multiple variance regimes were defined to accurate volatility forecasts, i.e., news of different magnitudes and not only of different signal are expected to have different marginal impacts on future volatility. One of the latest developments was brought by Medeiros and Veiga (2009), that presented the FCGARCH - Flexible Coefficient GARCH - in which a smooth transition variable triggers the change of volatility regimes. In this model each regime the impact of past volatility and past innovations is different, whereas GJR only differentiates the impact of past shocks. The advantages of this model are later explained. To sum up, the GARCH, GJR and FCGARCH are the models employed in this study to forecast volatility.

Not only latent volatility models have been used to predict RV. For instance, IV values have been extensively applied on modeling RV. Yet, Cristensen and Prabhala (1998), among others, found these forecasts to be biased.

So far, we have focused on the topic of forecasting RV, which is crucial in asset allocation and risk analysis. RV is also an underlying asset on futures, which are, for the moment, only available on foreign exchange markets. These futures deliver on maturity the squared returns of the period defined ex-ante.

However IV, the one referred as a biased forecast of RV, is of major importance in derivative pricing. Instead of only focusing on GARCH-type models to forecast RV, we should also use these models to predict IV.

Forecasting RV through GARCH-type models and IV values has been subject of a vast study. Whereas this topic has been developed on many studies, forecasting IV has been disregarded. This is a considerable gap in the existing literature given the increasing importance of IV. One should remember that option traders when making investment decisions are interested not only on the underlying asset price movements but also on the IV derived by an option's market price. Moreover, IV is a tradable asset not only by trading vega but, also, through futures and options on IV measures such as the VIX index.

1.1 Objectives

We propose to forecast both RV and IV using the GARCH-type models referred. During a period of ten years, the models are estimated on a daily basis and 1-day ahead volatility forecasts are compared with the daily RV and with the IV by the end of the day.

Consider the set of GARCH-type models

$$M = \{1, 2, 3\} = \{GARCH, \, GJR, \, FCGARCH\}.$$

Let $h^{(m)}_{t|t-1}$ be the volatility forecast for period t, provided by model m and based on all the available information at time $t-1$. We consider these volatility forecasts as IV and RV predictions:

$$\hat{RV}_t = h^{(m)}_{t|t-1}, \tag{1}$$

$$\hat{IV}_t = h^{(m)}_{t|t-1}. \tag{2}$$

Our objective is to assess which model best predicts RV and the foremost model on inferring IV. Our findings are of highly importance for investment and hedging strategies on RV and IV. In fact, we consider this thesis to be a preliminary step on developing a successful trading strategy on futures on RV and IV. Moreover, we appraise if GARCH-type models are more effective on predicting RV or IV. Hence, this thesis is an empirical analysis with fruitful conclusions for both the academy and the industry. We intend to arrive at the theoretical framework that superiorly suits financial markets' volatility mechanisms.

The study is organized in several steps. On section two, the methodology applied is presented, namely the models applied, the use of rolling window on the estimation process and the employment of Diebold-Mariano test to assess the models' predictive accuracy. Section three briefly describes the sample used and section four presents the empirical results. Finally, section five presents our concluding remarks.

2 METHODOLOGY

We first describe the GARCH-type models used, the variables to be forecasted (RV and IV) and, afterwards, the necessary steps to forecast volatility with these models and the process applied to rank them.

2.1 GARCH (1,1)

The discussion about dynamic volatility models must start by recalling the pioneer work of Engle (1982). Let $\mathcal{F}_{t-1}$represent all the available information at time $t-1$ and assume $\mu_{t|t-1}$to be the "true" conditional mean at time t. As aforementioned, before the path breaking ARCH model, volatility was considered to be constant (homoskedasticity), i.e.,

$$r_t = E(r_t|\mathcal{F}_{t-1}) + \varepsilon_t = \mu_{t|t-1} + \sigma z_t,\ \varepsilon_t \sim WN(0,\sigma),\ z_t \sim WN(0,1) \tag{3}$$

Returns' movements are described by an expected conditional mean and an innovation, ε_t. The conditional mean, $\mu_{t|t-1}$, derives from a combination of endogenous or/and exogenous variables, i.e., a set of information included in $\mathcal{F}_{t-1}$. The shock is described as a white noise process[1] with mean zero and scaled by a constant volatility.

Allowing conditional mean dynamics and no conditional variance dynamics was incoherent with the studies of Mandelbrot (1963)[2] and Fama (1965), which showed evidence of changing volatility and volatility clustering. However, heteroskedasticity was viewed as a cross-sectional phenomenon and, on the other hand, cross-sectional heteroskedasticity models were barely suited for dynamics environment (Andersen *et al.*, 2006).

Let's see how did Engle (1982) solved this problem. Consider the discretely sample return process, $\{r_t\}_{t=1}^{T}$, but assume that there is a conditional second moment, i.e., the innovation, ε_t, is scaled by the time-varying conditional volatility. Assume $\sigma_{t|t-1}$ to be the "true" standard deviation at time t. Thus,

$$r_t = E(r_t|\mathcal{F}_{t-1}) + \varepsilon_t = \mu_{t|t-1} + \sigma_{t|t-1} z_t,\ z_t \sim i.i.d.,\ E(z_t) = 0,\ Var(z_t) = 1. \tag{4}$$

To consider conditional heteroskedasticity, Engle (1982) brought the following solution, the ARCH(q) model:

$$\sigma^2_{t|t-1} = f(\varepsilon_{t-1}, ..., \varepsilon_{t-q}) = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} + ... + \alpha_q \varepsilon^2_{t-q} = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon^2_{t-i}. \tag{5}$$

[1] A sequence of independent and identically distributed (iid) random variables with constant mean and variance.

[2] Mandelbrot (1963): "Large changes tend to be followed by large changes of either sign."

To guarantee that both unconditional and conditional variances are positive two conditions are straightforward to observe: $\alpha_0 > 0$ and $\alpha_i \geq 0$, $i\epsilon[1,\, q]$.

With this model, the estimated conditional variance can change in time and it is a linear function of past shocks. Even though both conditional mean and conditional variance are time-varying, the unconditional mean and the unconditional variance must be constant to impose covariance stationary - the probability distribution does not change through time. For this, $\sum_{i=1}^{q} \alpha_i < 1$(see proof in Greene, 2002, p. 239).

As found by Mandelbrot (1963), asset returns' variance is a long memory process. It would be possible to capture this stylized fact with an ARCH(q), where q is high.

Thus, Bollerslev (1986) brought a more parsimonious specification, the GARCH model, which is a generalization of the ARCH model. The GARCH (q, p) is represented by:

$$\sigma_{t|t-1}^2 = f(\varepsilon_t, ..., \varepsilon_{t-q}, \sigma_{t-1}^2, ..., \sigma_{t-p}^2) = \alpha_1 + \sum_{i=1}^{q} \lambda_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-p}^2. \tag{6}$$

First, we refer the inequality constraints. To assure that conditional variance is non-negative, $\alpha_1 > 0$, $\lambda_i \geq 0$, $i\epsilon[1,\, q]$ and $\beta_i \geq 0$, $i\epsilon[1,\, p]$. $\sum_{i=1}^{q} \lambda_i + \sum_{i=1}^{q} \beta_i < 1$ allows the process to be covariance stationary (see Nelson and Cao, 1992). As aforementioned, in financial markets contexts, volatility tends to be highly persistent (see Diebold, 1988 and Bollerslev *et al.*, 1988). This stylized fact asks for an ARCH(∞) model that is infeasible in practice. The GARCH model solves this infeasibility by including past shocks in the past estimated volatility.

Consider a GARCH(1,1):

$$\sigma_{t|t-1}^2 = \alpha_1 + \lambda_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{7}$$

By recursive substitution is straightforward to express it by an ARCH(∞) model,

$$\sigma_{t|t-1}^2 = \alpha_1 (1 - \beta_1)^{-1} + \lambda_1 \sum_{i=1}^{\infty} \beta_1^{i-1} \varepsilon_{t-i}^2. \tag{8}$$

The empirical adequacy and the parsimony of a GARCH(1,1) impelled the academy and the industry to use it as the main tool in volatility forecasting. For this reason, it is one of the models applied in this study.

2.2 GJR (1,1)

One of the key features of the GARCH process is its symmetry - the sign of past shocks does not influence future volatility. However, Black (1976) and Christie (1982), among others, found evidence of the so-called "Leverage Effect", i.e., changes in innovations and changes in volatility are negatively correlated.

The impact of a negative and a positive innovation of the same magnitude tends to impact differently the volatility. A "very bad news" results in a higher volatility than a "very positive news". Many economic reasons have been used the explain the phenomena, some examples are the financial leverage and the volatility feedback (see Andersen *et al.*, 2006). Yet, we are only focused in modeling this feature. Therefore, among the vast group of asymmetric GARCH formulations, we selected the GJR (1993), which is probably one of the most used on the field.

The GJR(1,1) mimics the GARCH(1,1) but includes a dummy variable, which takes the value one whenever the past shock is negative:

$$\sigma^2_{t|t-1} = \alpha_1 + (\lambda_1 + \gamma I_{\varepsilon_{t-1}<0})\varepsilon^2_{t-1} + \beta_1 \sigma^2_{t-1}. \tag{9}$$

A positive and statistically significant estimate for γ indicates a negative asymmetric volatility response to positive and negative shocks.

2.3 FCGARCH (1,1,2)

The GJR model, among other models, introduced a class of threshold-type volatility models. In the case that innovations are positive volatility follows a regime, whereas negative shocks triggers a change of regime - the difference between the regimes concerns the "activation" of γ. However, Engle and Ng (1993) went farther than allowing for a threshold at zero. They allowed for multiple "news impact curves" at different locations, i.e., different volatility regimes for the different ranges of the lagged innovation. Yet, this type of model lacks smoothness on the transition process. Volatility changes abruptly between regimes.

Hargerud (1997) introduced the smooth transition GARCH (STGARCH) model. He discusses both the logistic and the exponential STGARCH specifications. Based on the logistic function, the model allows for a smooth transition between regimes depending on the sign of past returns. With an exponential function, the magnitude of the lagged squared of return smoothly alters the volatility regime. Many other models, which incorporate both sign and size asymmetries based on threshold or smooth transition specifications, have been developed; see Medeiros and Veiga (2009, hereafter MV).

MV also present another nonlinear GARCH model, the FCGARCH. It departs from the logistic STGARCH but allows for both sign and size asymmetries with no fixed values for the location of regimes. Moreover, MV outline the following advantages of the FCGARCH when comparing it with the existent literature on nonlinear GARCH models. First, more than two regimes are possible and are determined by a sequence of tests that avoid the identification problem in nonlinear time series. Secondly, the stationary condition is satisfied with weak restrictions, allowing for rich dynamics that other GARCH-type models fail to describe.

Thus, the FCGARCH(1,1,m), where $m = h + 1$ regimes, is described by:

$$\sigma^2_{t|t-1} = f(\varepsilon_t, \sigma^2_{t-1}, s_t) = \alpha_1 + \beta_1\sigma^2_{t-1} + \lambda_1\varepsilon^2_{t-1} + \sum_{i=2}^{m}[\alpha_i + \beta_i\sigma^2_{t-1} + \lambda_i\varepsilon^2_{t-1}]f_i(s_t), \qquad (10)$$

where $f_i(s_t)$ is a logistic function defined as:

$$f_i(s_t) = \frac{1}{1 + e^{-\gamma_i(s_t - c_i)}}. \qquad (11)$$

$f_i(s_t)$ plays the role of a nonlinear smooth transitional continuous function where the transition variable is s_t. There are different possible choices for the transition variable, which can be exogenous or endogenous to the model, as long as it follows a strictly stationary process. MV uses past innovations. An alternative could be the volume. We will consider only the use of past shocks, s_t is ε_{t-1}. Testing other variables is left for future research.

As aforementioned, MV produced a sequence of tests to determine the number of regimes. However, not only the test is computationally intensive but the model is exponentially less parsimonious when the number of regimes increases - note that each regime implies estimating five other parameters. Given the importance of the principle parsimony in the financial field, it is applied a FCGARCH with only two regimes without the sequence of tests referred before.

It now follows a brief description of the conditions that allow for an FCGARCH(1,1,2),

$$\sigma^2_{t|t-1} = \alpha_1 + \beta_1\sigma^2_{t-1} + \lambda_1\varepsilon^2_{t-1} + [\alpha_2 + \beta_2\sigma^2_{t-1} + \lambda_2\varepsilon^2_{t-1}]f_2(\varepsilon_{t-1}), \qquad (12)$$

to work properly.

"The FCGARCH model is identifiable if there are no two sets of parameters such that the corresponding distributions of the population variable ε are identical.", MV (p. 130). The same input-output map must not be obtained based on two different FCGARCH models, which would imply two equal maximums for the log-likelihood function. Different FCGARCH(1,1,2) models produce the same output - volatility estimates - if one of the models is reducible. Thus, to achieve a feasible estimation process the model has to be irreducible, i.e., there has to be no irrelevant regimes. The inexistence of irrelevant regimes is assured if α_2, β_2, and λ_2 do not equal 0 simultaneously and $\gamma_2 \neq 0$.

The reason for the first condition is straightforward to understand. For the second one should note that when $\gamma_2 = 0$, $f_2(\varepsilon_{t-1}) = 1/2$, i.e., it becomes a constant. Thus, the regime has no reason to exist because it can be incorporated on the plain-vanilla regime.

One also needs to ensure strictly positive conditional variances. Note that $f_2(\varepsilon_{t-1})\epsilon[0,1]$. Thus, if $\sum_{i=1}^{K}\alpha_i > 0$, $\sum_{i=1}^{K}\beta_i \geq 0$ and $\sum_{i=1}^{K}\lambda_i \geq 0$, $K\epsilon\{1, 2\}$, the conditional variance is always positive.

MV proves that the following assumption is a sufficient condition for strict stationary of the model: $\frac{1}{2}(\beta_1+\lambda_1)+\frac{1}{2}\sum_{i=1}^{2}(\beta_i+\lambda_i)<1$.

So, how does the model work? Assume that $\gamma_2>0$:

	$\alpha_1+\beta_1\sigma_{t-1}^2+\lambda_1\varepsilon_{t-1}^2$			$[\alpha_2+\beta_2\sigma_{t-1}^2+\lambda_2\varepsilon_{t-1}^2]f_2(\varepsilon_{t-1})$					
c_2	N/A			highly negative			highly positive		
ε_{t-1}	$\to -\infty$	$\to 0$	$\to +\infty$	$\to -\infty$	$\to 0$	$\to +\infty$	$\to -\infty$	$\to 0$	$\to +\infty$
$f_2(\varepsilon_{t-1})$	N/A			$\to 0$	≈ 1	$\to 1$	$\to 0$	≈ 0	$\to 1$
Regime is	active			inactive	active	active	inactive	inactive	active

Table 1: The engine of a FCGARCH (1,1,2).

Note that $f_2(\varepsilon_{t-1})$ has the opposite behavior when $\gamma_2<0$.

Sign asymmetries are straightforward to observe. Assuming that $\gamma_2>0$, the so called "leveraged effect" is present if the regime has a negative impact on conditional variances. Size asymmetries are defined by the value of c_2. The model also allows other rich dynamics. For instance, with negatives α_2 and β_2 but a positive λ_2, a "very good new" triggers higher marginal volatility per past return than in "tranquil periods".

A final word for γ_2. Its role is to determine the speed of transition between regimes, being that when $\gamma_2\to+\infty$ the model becomes a threshold-type model.

2.4 Realized Volatility

Let p_t be the asset price at the end of time t. Thus, the logarithm return is:

$$r_t = ln\left[\frac{P_t}{P_{t-1}}\right]. \tag{13}$$

Historical RV is often computed as:

$$\hat{\sigma}_t^2=\frac{1}{N-1}\sum_{t=1}^{N}(r_t-\overline{r})^2, \tag{14}$$

the sample variance of asset returns, which can vary in terms of periodicity. By increasing the frequency to its maximum, reckoning each trade, one would have a measure that considers the complete spectrum of price variations. Notwithstanding, this is a biased estimator of RV. Some authors argue that the use of spot volatility, tick by tick, is constrained due to noise problems introduced by micro-structure effects such as bid-ask spreads, non-synchronous trades and intraday volatility patterns; see Andersen *et al.* (2003) for a comprehensive discussion. On the same paper, five minutes interval was found to be the best periodicity. In fact, the true RV is latent, i.e., it can only be estimated with some error (Andersen *et al.*, 2006) - note that we designated the sample standard deviation as $\hat{\sigma}$, i.e., an estimator of volatility and not the true observer volatility. We ignore this point, and focus, instead, on daily RV defined as the absolute value of daily asset returns,

$$RV_t = |r_t|. \tag{15}$$

Despite being a biased estimator, futures on foreign exchange RV[3] define RV as the squared daily asset return. As this study is a first step on developing a RV futures trading strategy, we shall consider RV as market makers define it. In the case that futures on RV become exchange-traded, it is likely that RV will be designated as the squared daily returns. Although the use of squared returns as a proxy of volatility can be criticised, it is of irrefutable importance to forecast squared returns given its application on exchange-traded futures.

2.5 Implied/Market Volatility

IV is a reverse way of pricing an option. The relation between IV and an option price is similar to the relation between a yield and a bond price. One can value an option(bond) through its price or IV(yield).

IV is of major importance because it reflects the market sentiment on the underlying asset, it is a pricing measure for option traders and it is a tradable asset when hedging all greeks but vega and through futures on IV proxies.

If we were to focus on trading vega, we would likely incur in errors of measurement and misspecification. On a plain vanilla bond the cash-flows are ex-ante determined - ignoring credit risk. But on the other hand, option's cash-flows depend on the underlying asset's spot price at maturity, creating the need for statistical pricing formulas, e.g. Black-Scholes. The use of these type of formulas increases the likelihood of estimation's errors.

However, there are IV proxies that mitigate the problems of measurement errors of models' inputs and model misspecification. One of them is the VIX index, which is presented next. The fact that the VIX index is an underlying asset on futures impel us to view it as a more efficient tool for trading IV. Hence, our focus should be on analysing the accuracy of GARCH-type models to predict the VIX.

[3]http://www.cmegroup.com/trading/fx/realized-fx-volatilities/eur-usd-1-month-realized-volatility_contract_specifications.html

2.5.1 VIX

The VIX is a volatility index calculated by the Chicago Board Options Exchange. It is based on the bid/ask quotes of options on the S&P 500 index and is considered to be the fear index because it reflects the market's expectation of the volatility in the S&P500 index. An increase of the demand for put options triggered by a more prudent attitude rises IV and, consequentially, increases the value of VIX.

The VIX was revised in 2003 incorporating two major changes[4]. First, a wider range of strike prices is included so that the VIX assimilates the volatility skew. Secondly, it is independent of a pricing model.

For the sake of comparison, RV and IV values need to be on the same basis. Thus, given that VIX values are on an annualized basis, we need to adjust these values to a daily basis:

$$IV_t = \frac{VIX_t}{\sqrt{252} * 100} \tag{16}$$

2.6 Forecasting

We proceed by describing the necessary steps to forecast volatility with GARCH-type models.

First, an assumption regarding the error's distribution has to be made. Independently of the Jarque-Bera test result - presented in Data Analysis -, which tests if innovations follow a Gaussian distribution, we assume that the errors follow a Normal distribution to simplify the estimation process. In fact, Bollerslev *et al.* (2010) finds no benefit in assuming a t-Student distribution when forecasting volatility. Thus, the returns' process is defined by

$$r_t = \mu_{t|t-1} + \sigma_{t|t-1} z_t,\ z_t \sim N(0,1). \tag{17}$$

Following, the stationarity of the logarithm returns series should be assessed by the Augmented Dickey-Fuller (ADF) test, that assumes in the null hypothesis the presence of a unit root/nonstationary process. If there is evidence of stationarity, which is usual on logarithm returns, no transformation of the series is needed.

Thereafter, we need to evaluate the existence of statistically significant returns autocorrelations. It is done by performing the Ljung-Box test, which considers no autocorrelation between returns on the null hypothesis. In the case of statistically significant autocorrelations the conditional mean has to be model through ARMA - Autoregressive Moving Average - processes. If the returns are uncorrelated we can assume the series of returns to correspond to the series of shocks, i.e., the conditional mean is zero.

[4]See calculation procedures at http://www.cboe.com/micro/vix/vixwhite.pdf.

After, the Engle's ARCH Lagrange Multiplier test is performed to evaluate the existence of conditional heteroskedasticity. This test confirms the need of conditional variance modeling.

Finally, we apply the models described before on a sample of ten years. We consider the length of the period to be sufficient to support our findings.

Regarding the periodicity of models' estimation, we decided to perform it on a daily basis. We expect this to improve the predictive accuracy of the models. According to Andersen *et al.* (2003), five year of monthly data is the most common rolling window used in volatility forecasting. Thus, on each day from January, 2001 to December, 2010 the previous 5 years of daily observations are used to estimate the GARCH-type models and to forecast the next day volatility. The models are estimated by maximizing the logarithm of the likelihood function, which, as we assume a Gaussian distribution for the conditional distribution - $\varepsilon_t \sim N(0, \sigma_t^2)$ -, is described as:

$$l_t = lnL_t = -\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(h_t) - \frac{\varepsilon_t^2}{2h_t}. \tag{18}$$

For the sake of clarification, we briefly describe the models presented before and the assumptions that allow them to properly work:

$$h_t^{(1)} = \alpha_1 + \lambda_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2, \tag{19}$$

$$Assumption_1: \ \alpha_1 > 0, \ \lambda_1 \geq 0 \, and \, \beta_1 \geq 0,$$

$$Assumption_2: \ \lambda_1 + \beta_1 < 1;$$

$$h_t^{(2)} = \alpha_1 + (\lambda_1 + \gamma I_{\varepsilon_{t-1}<0})\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \tag{20}$$

$$Assumption_1: \ \alpha_1 > 0, \ \lambda_1 \geq 0 \, and \, \beta_1 \geq 0,$$

$$Assumption_2: \ \lambda_1 + \beta_1 < 1;$$

$$h_t^{(3)} = \alpha_1 + \lambda_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 + [\alpha_2 + \beta_2\sigma_{t-1}^2 + \lambda_2\varepsilon_{t-1}^2]\frac{1}{1 + e^{-\gamma_2(\varepsilon_{t-1} - c_2)}}; \tag{21}$$

$$Assumption_1: \ \{\alpha_2, \beta_2, \lambda_2\} \neq \{0, 0, 0\} \ and \, \gamma_2 \neq 0,$$

$$Assumption_2: \sum_{i=1}^{K}\alpha_i > 0,\ \sum_{i=1}^{K}\beta_i \geq 0\, and \sum_{i=1}^{K}\lambda_i \geq 0,\ K = \{1,\, 2\},$$

$$Assumption_3: \frac{1}{2}(\beta_1 + \lambda_1) + \frac{1}{2}\sum_{i=1}^{2}(\beta_i + \lambda_i) < 1.$$

2.7 Evaluation

The measurement of models' capability to forecast 1-day ahead RV and IV is based on two diagnostic tools: Diebold-Mariano (DM) test and mean absolute error (MAE).

Let $\{RV_t\}\, and\, \{IV_t\}$ be the series to be forecasted and $\{h_t^{(m)}\}$, $M = \{1, 2, 3\}$, the forecasts produced by the different GARCH-type models.

From each model, the forecast errors are $\omega_t^{(m)} = RV_t - \sqrt{h_t^{(m)}}$ and $\theta_t^{(m)} = IV_t - \sqrt{h_t^{(m)}}$. $\omega_t^{(1)}$, $\omega_t^{(2)}$ and $\omega_t^{(3)}$ use the same series of RV. As they all use the same data it might be the case that the series of forecast errors are serially correlated - the same applies to IV. The DM test is performed to circumvent this hurdle.

First, consider the following loss functions: $L(\omega_t^{(m)}) = |\omega_t^{(m)}|$ and $L(\theta_t^{(m)}) = |\theta_t^{(m)}|$. The null hypothesis of the DM test assumes an equal forecasting capacity. Thus, for the case of RV it tests the following null hypothesis:

$$H_o:\ E[L(\omega_t^{(i)})] = E[L(\omega_t^{(j)})],\ i\epsilon M,\ j\epsilon M,\ i \neq j. \tag{22}$$

If the absolute value of the statistic test is above 1.96 the null hypothesis is rejected, meaning that one model predicts better than the other. If the statistic is below -1.96 there is evidence that model (i) best predicts, whereas a value higher than +1.96 indicates a better predictive accuracy of model (j).

To assess what do GARCH models best predict - if RV or IV - we use the mean absolute error: $MAE(RV) = \frac{\sum_{t=1}^{T}|\omega_t^{(m)}|}{T}$ and $MAE(IV) = \frac{\sum_{t=1}^{T}|\theta_t^{(m)}|}{T}$.

3 DATA ANALYSIS

This section is intended to analyse the data used to achieve the proposed objectives of this thesis.

The descriptive statistics of the data are presented in table 2. We use the S&P500's daily closing prices from January, 1996 to December, 2010. S&P 500 index values are taken from Bloomberg and VIX index values from CBOE. Despite the fact that VIX calculations procedures were revised in September 2002, past VIX values were re-calculated according to the new rules.

Series	S&P 500's Returns	RV	IV
Start	03/01/96	02/01/2001	02/01/2001
End	31/12/10	31/12/10	31/12/10
Observations	3776	2514	2514
Sample Mean	0,0002	0,0092	0,0139
Median	0,0007	0,0061	0,0128
Sample Std. Deviation	0,0131	0,0103	0,0063
Maximum	0,1096	0,1096	0,0509
Minimum	-0,0947	0,0000	0,0062
Sample Skewness	-0,1856	3,1184	1,8627
Sample Kurtosis	7,4617	16,4257	5,2245

Table 2: Descriptive Statistics of Returns

Logarithm daily returns were computed by the close price on a specific date and the close on the day before. RV corresponds to the absolute value of daily returns, whereas IV is computed by adjusting the VIX index to a daily basis:

$$r_t = ln\left[\frac{P_t}{P_{t-1}}\right],\, RV_t = |r_t|\, and\, IV_t = \frac{VIX_t}{\sqrt{252} * 100} \tag{23}$$

Although S&P 500's returns have a positive mean, the negative skewness points an asymmetry, namely a heavier left tail on the empirical distribution function. It is widely agreed that negative returns - left tail - have more weight on empirical returns distribution than on a Normal distribution. In fact, both tails are heavier than in a Gaussian distribution - a Normal distribution has a kurtosis of three. Yet, the difference between the right tails is lower. These non-Gaussian evidences are later confirmed by the result of the Jarque-Bera test.

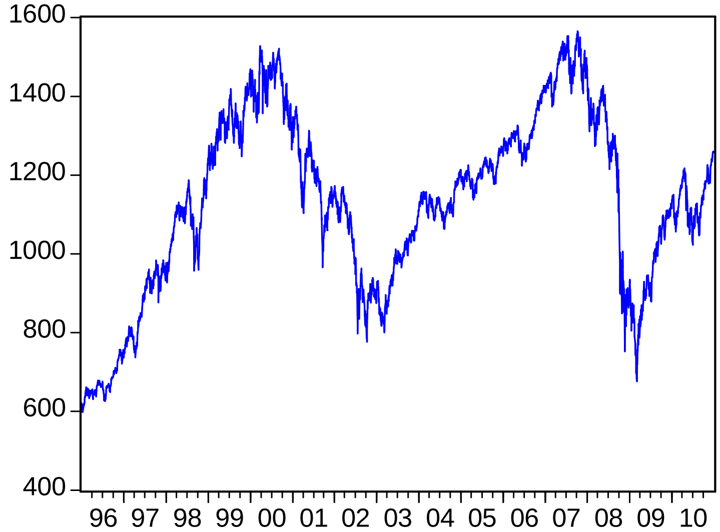

Figure 1: S&P500's daily closing prices

The sample chosen allows to test the models throughout the dotcom and subprime bubbles. One observes bullish markets before 2000 and from 2003 to 2007 and crashes from 2000 to 2002 and 2008 to 2009.

The objective is to test the models under stressed market conditions, when both realized and IV tend to fluctuate more. Assessing forecasts' accuracy on calm and turbulent periods allows to take a better picture of the models' capabilities.

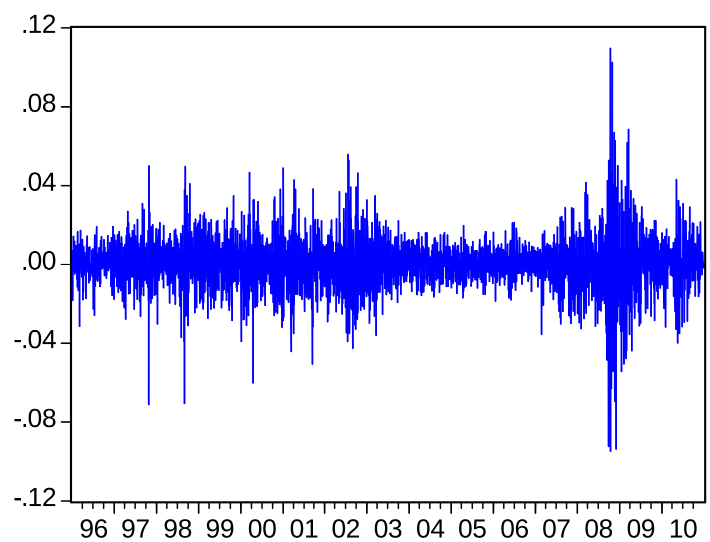

Figure 2: S&P500's daily returns

Figure 2 presents the S&P500's daily returns throughout the sample chosen. It is observable that the subprime crisis presents a more volatile pattern with clear and striking spikes.

The mean stationarity of returns is later statistically proved. Yet, it is easy to observe it in Figure 2 as returns fluctuates around an approximate zero mean. Moreover, volatility clustering is crystal clear on returns movements. 96 to 97 and 03 to 06 calm periods contrast with the turbulent intervals, which take place from 98 to 02 and, especially, from 07 to 10.

Figure 3 presents S&P500 index RV and IV. It is easily observable a higher variation of RV while IV tends to be more stable. The maximum/minimum range and the sample standard deviation is higher on RV. Thus, both graphically and numerically the RV evidences to be more volatile.

We proceed by performing some of the statistical tests referred on Methodology.

As expectable, the Jarque-Bera test result clearly rejects the Gaussian distribution hypothesis - the p-value is below a 5% significance level. Thus, this study may be repeated with the application of other distributions.

The result of the ADF test in Table 3 indicates the presence of stationarity as it is clearly below the critical value at 5% significance level. Therefore, there is no need of transformations on the series used.

The Ljung-Box statistic is above a 5% significance level indicating the returns to be uncorrelated. Hence, we assume a zero conditional mean, i.e., $\mu_{t|t-1} = 0$ and $r_t = \varepsilon_t$.

Finally, on the ARCH(1) LM test, the null hypothesis, which assumes conditional homoskedasticity, is rejected. We conclude that the S&P 500 index returns exhibit ARCH effects, which reinforces the need of GARCH-type models to forecast volatility.

	Statistic	P-value
Jarque-Bera	686.2963	0.001
ADF	-35.5039	0.001
Ljung-Box(15)	21.8911	0.1107
ARCH(1) LM	48.3142	0

Table 3: Statistic tests (1996 to 2000)

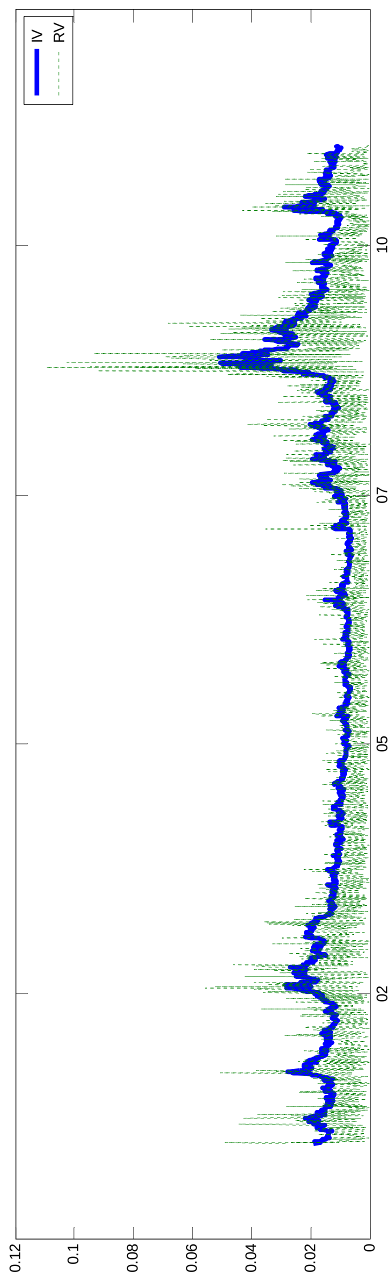

Figure 3: S&P500 index: RV and IV

4 EMPIRICAL RESULTS

Using the methodology aforementioned, we produce daily volatility forecasts assuming the three different GARCH models. To estimate the parameters we use a rolling window, re-estimating the models' parameters. We recorded the series of parameters, model likelihood and volatility forecasts. The descriptive statistics of these series is available on Appendices.

Based on the sample used, Figure 4 relates the past daily innovation with the one-day ahead volatility forecast. This relation is designated as the "News Impact Curve" and was firstly highlighted by Black (1976). It captures the impact of a shock on the next day volatility. The importance of this relation concerns the different inclination of the curve for positive vs. negative innovations.

The graph is slightly asymmetric and seems to confirm the presence of a higher marginal impact on volatility when negative shocks are registered, i.e., a higher slope of the curve in the case of negative shocks.

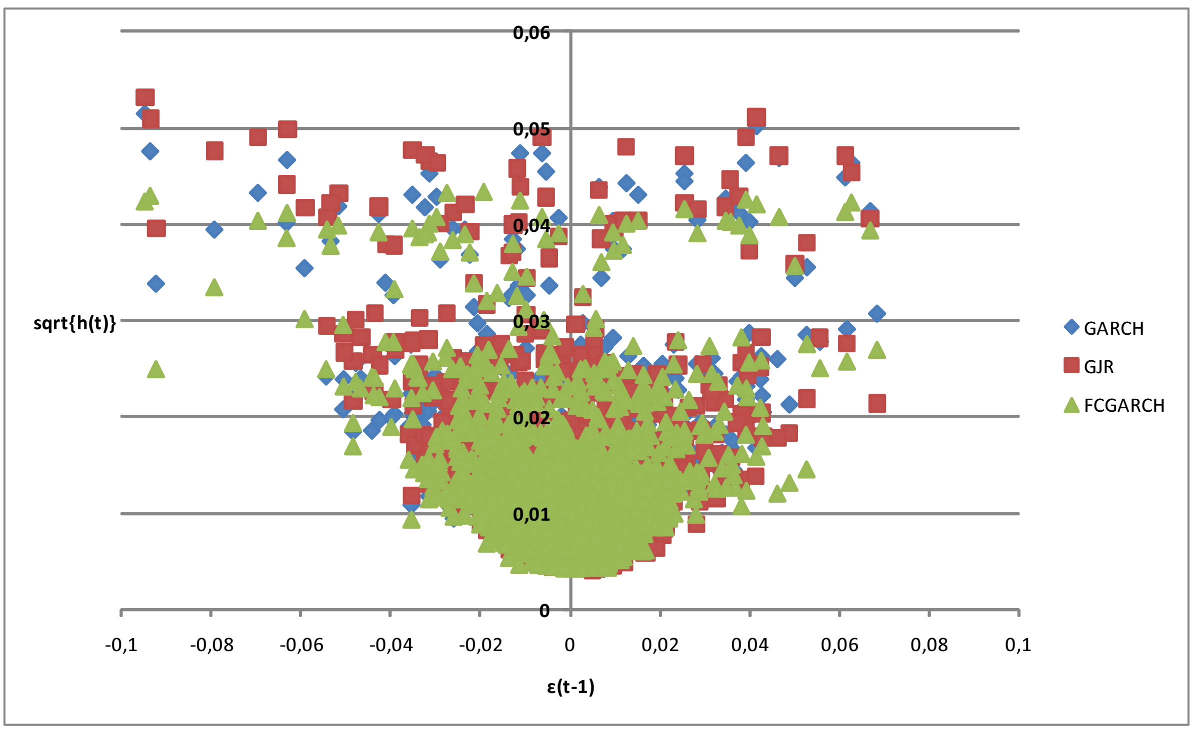

Figure 4: News Impact Curve

Figure 5 presents the GARCH, GJR and FCGARCH volatility forecasts, which are, on a visual analysis, synchronized. The only exceptions are the high FCGARCH spikes during the dot-com crisis.

Figure 6 and figure 7 present the models' forecasts, as well as, RV and IV. Despite the models' dynamics in responding to recent shocks, they are still unable to follow the noisy behavior of RV. Yet, it is visible that models' forecasts pursue the RV's directional tendency. The pattern of IV is clearly more stable and convergence of forecasts with IV seems to be far more frequent than with RV.

Finally, we present the statistic measures used to rank the models and achieve the objectives of this thesis.

	GARCH-GJR	GARCH-FCGARCH	GJR-FCGARCH
DM (RV)	4.889	4.459	0.183
DM (IV)	5.730	3.394	1.306

Table 4: Diebold-Mariano test

According to the results, GJR model beats the GARCH and the FCGARCH has a better predictive accuracy than GARCH on both RV and IV because the DM statistics are above +1.96. However, one cannot reject the hypothesis of equal forecasting capacity of GJR and FCGARCH on both RV and IV as the absolute value of the statistics are in-between -1.96 and +1.96. Thus, we conclude that GJR and FCGARCH are the best models to predict RV and IV.

	GARCH	GJR	FCGARCH
MAE (RV)	0.0068	0.0067	0.0067
MAE (IV)	0.0025	0.0027	0.0027

Table 5: Mean Absolute Errors

By analysing the MAE results, we achieve the interesting conclusion that the capacity of GARCH-type models to predict IV overperforms the accuracy of these models to forecast RV. Thus, IV is better predicted by GARCH-type models than RV - lower MAE on all models. We can deduce that market's expectation about future volatility is best inferred by GARCH models than future RV.

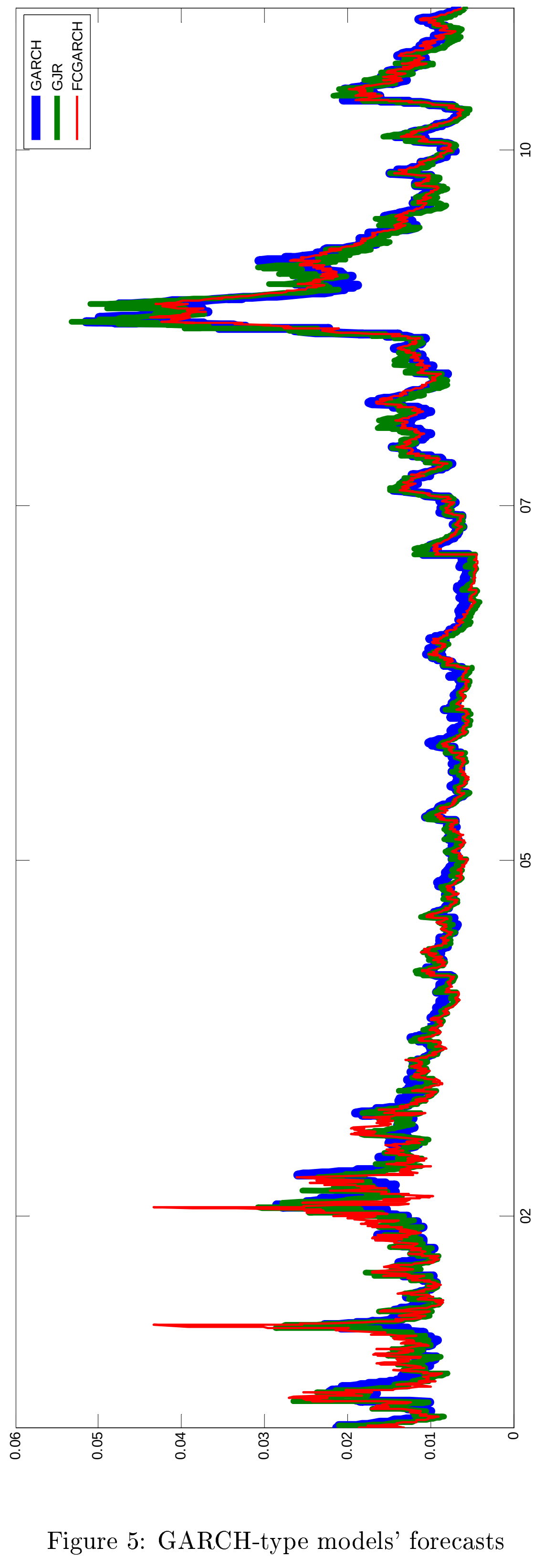

Figure 5: GARCH-type models' forecasts

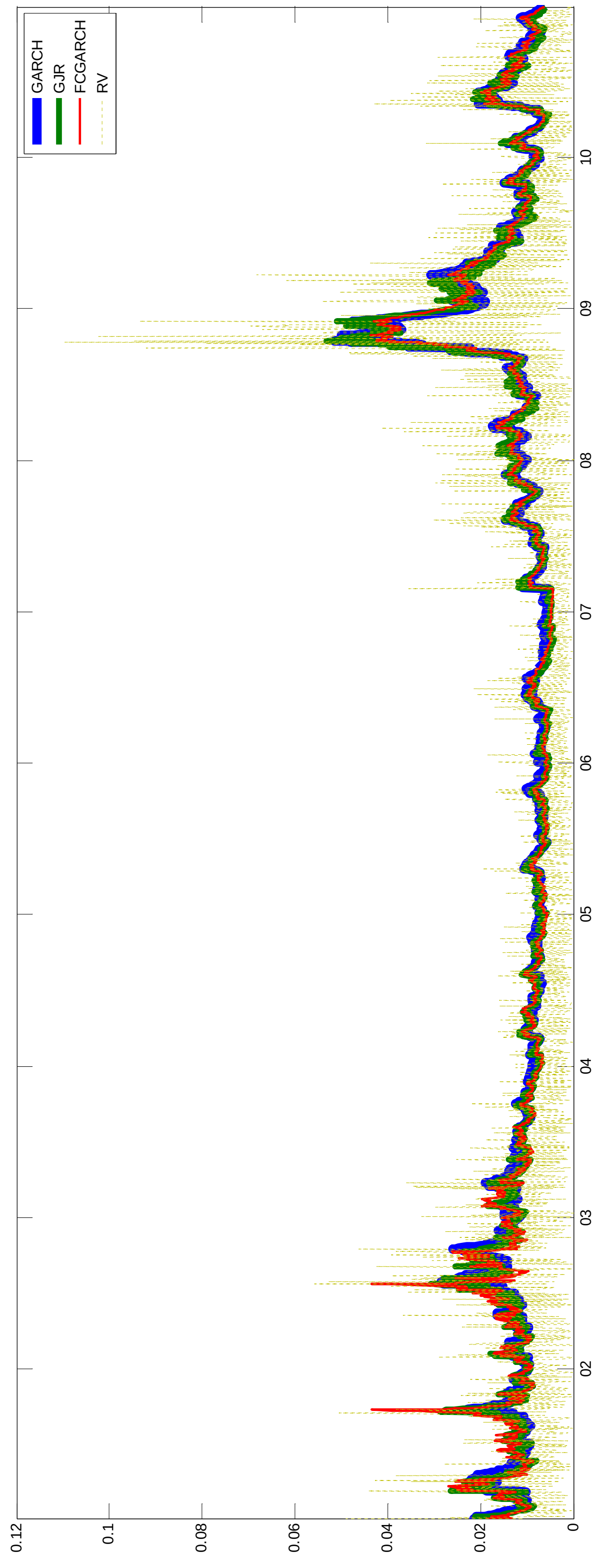

Figure 6: Forecasts and RV

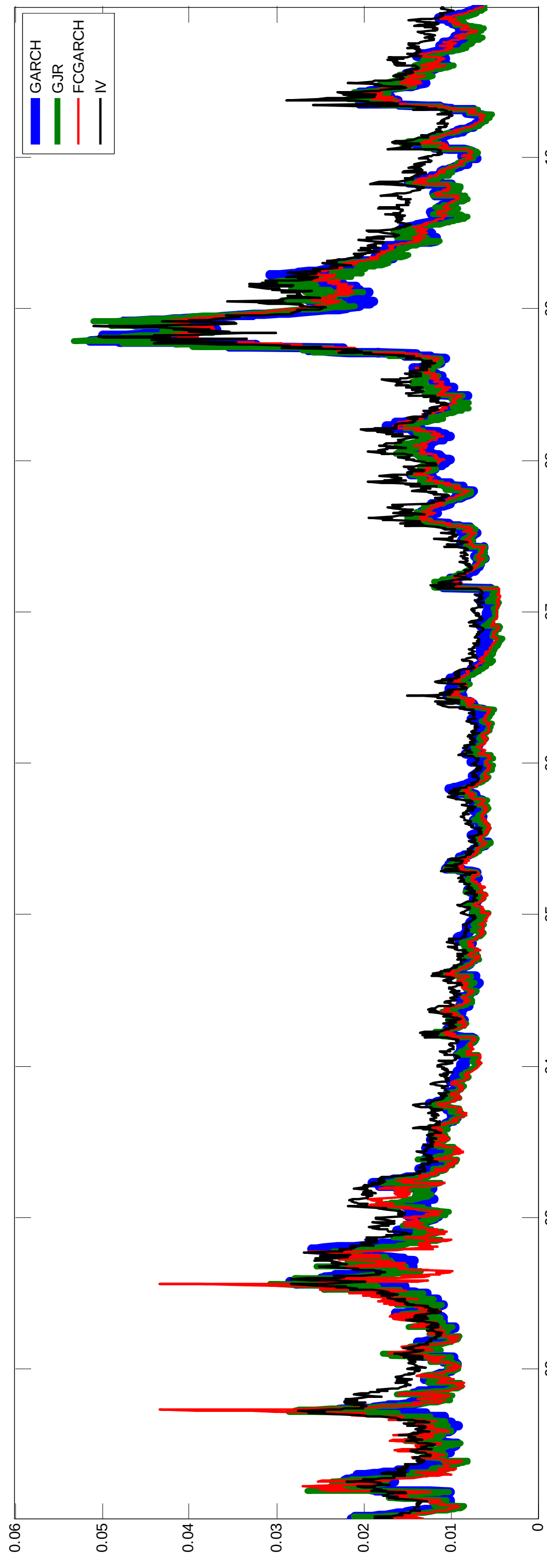

Figure 7: Forecasts and IV

5 CONCLUSIONS

We performed an analysis of the accuracy of GARCH-type models to forecast, from 2001 to 2010, the VIX and daily RV of the S&P500 index. We apply the plain-vanilla GARCH model, incorporate the leverage effect by employing the GJR model and introduce a smooth transition process between volatility regimes by fitting the FCGARCH model. This is a preparatory step in developing a trading strategy on futures on which the underlying assets are daily squared returns or the VIX index - a measure of IV.

Based on Diebold-Mariano test, the GJR(1,1) and FCGARCH(1,1,2) reveal to best predict both RV and IV. The results can be justified by the embedded dynamics on GJR and FCGARCH allowing for different volatility regimes. Thus, trading strategies on RV and IV futures shall adopt these models to produce forecasts. This is left for future research.

Moreover, by comparing mean absolute errors one can also deduct that GARCH models are more suitable to foresee IV than RV. Thus, a plausible corollary is that past returns and past variance affect more IV, market's expectation of future volatility, than future RV.

6 REFERENCES

Ahoniemi, K. (2006), Modeling and Forecasting Implied Volatility - an Econometric Analysis of the VIX Index, Helsinki Center of Economic Research.

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2003), Modeling and Forecasting Realized Volatility, Econometrica 71(2), 579-625.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen and F. X. Diebold (2006), Volatility and Correlation Forecasting, Handbook of Economic Forecasting, 778-878.

Bachilier, L. (1990), Théorie de la Spéculation, Annales Scientifiques de l'É.N.S. 3e série, tome 17, 21-86.

Black, F. (1976), Studies in stock price volatility changes, ASA Proceedings, Business and Economic Statistics, 177–181.

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

Bollerslev, T., R. F. Engle and J. M. Wooldridge (1988), A Capital Asset Pricing Model with Time Varying Covariances, Journal of Political Economy 95, 116–131.

Brownlees, C., R. Engle and B. Kelly (2010), A Practical Guide to Volatility Forecasting through Calm and Storm, New York University and University of Chicago.

Christie, A. (1982), The stochastic behavior of common stock variances: Value, leverage and interest rate effects, Journal of Financial Economics 10, 407–432.

Curto, J. D., J. C. Pinto and J. A. Tomaz (2009), A new approach to bad news effects on volatility: the multiple-sign-volume sensitive regime EGARCH model (MSV-EGARCH), Portuguese Econonmetric Journal 8:23-36.

Christensen, B.J. and N.R. Prabhala (1998), The relation between implied and realized volatility, Journal of Financial Economics, 50.

Diebold, F. X. (1988), Empirical Modeling of Exchange Rate Dynamics, Springer-Verlag, New York.

Diebold, F. X. (2004), The Nobel Memorial Prize for Robert F. Engle, Scandinavian Journal of Economics 106(2), 165-185.

Engle, R. F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation, Econometrica 50, 987–1008.

Engle, R. F. and V. K. Ng (1993), Measuring and Testing the Impact of News on Volatility, Journal of Finance 48, 1749–1778.

Fama, E. F. (1965), The Behavior of Stock Market Prices, Journal of Business 38, 34-105.

Glosten, L., R. Jagannanthan, and R. Runkle (1993), On the relationship between the expected value and the volatility of the nominal excess returns on stocks, Journal of Finance

48, 1779–1801.

Greene, W. H. (2002), Econometric Analysis, New York University, Pretince Hall.

Hagerud, G. E. (1997), A new non-Linear GARCH model, Ph.D. Dissertation, Stockholm School of Economics.

Majmudar, U. and A. Banerjee (2004), VIX Forecasting, 40th Annual Conference of The Idian Econometrics Socitey Institute for Social and Economic Change, Bangalore, India.

Mandelbrot, B. (1963), The Variation of Certain Speculative Prices, Journal of Business 36, 394-419.

Nelson, D. B. and C. Q. Cao (1992), Inequality Constraints in the Univariate GARCH Model, Journal of Business & Economic Statistics, Vol. 10, No. 2, pp. 229-235.

Medeiros, M. C. and A. L. Veiga (2009), Modeling Multiple Regimes in Financial Volatility with a Flexible Coefficient GARCH(1,1) Model, Econometric Theory, 25, 117-161.

7 APPENDICES

	α_1	λ_1	β_1	Likelihood	$\sqrt{h_t}$
Sample Mean	3.929e-006	0.0856	0.8871	3.9608e+003	0.0118
Median	2.000e-006	0.0878	0.8944	3.9304e+003	0.0103
Sample Std. Deviation	3.254e-006	0.0131	0.0238	213.8664	0.0063
Maximum	1.3279e-005	0.1194	0.9145	4.3429e+003	0.0514
Minimum	2.000e-006	0.0563	0.8153	3.6429e+003	0.0053
Sample Skewness	1.5062	-0.1452	-1.4432	0.3681	2.8246
Sample Kurtosis	3.8245	2.4257	4.1470	1.9001	13.6716

Table 6: GARCH(1,1) model statistics.

	α_1	λ_1	γ	β_1	Likelihood	$\sqrt{h_t}$
Sample Mean	3.269e-006	3.4125e-008	0.1416	0.9079	3.9927e+003	0.0117
Median	1.366e-006	1.5006e-008	0.1307	0.9226	3.9619e+003	0.0101
Sample Std. Deviation	3.273e-006	5.8142e-008	0.0498	0.0392	210.2855	0.0068
Maximum	1.179e-005	7.9649e-007	0.2514	0.9577	4.3670e+003	0.0532
Minimum	4.559e-007	5.3629e-018	0.0604	0.8157	3.6734e+003	0.0041
Sample Skewness	1.1387	4.4252	0.8133	-1.0182	0.3560	2.7230
Sample Kurtosis	2.8491	32.1206	2.5547	2.7112	1.8988	12.8254

Table 7: GJR(1,1) model statistics.

	α_1	λ_1	β_1	α_2	λ_2	β_2	γ_2	c_2	Likelihood	$\sqrt{h_t}$
Sample Mean	2.196e-006	0.0118	1.1879	-3.5897e-006	0.0227	-0.8141	1.5258	-3.3667	-3.9878e+003	0.0116
Median	7.418e-007	0.0315	1.5204	2.8749e-007	-0.0039	-1.4093	0.4783	0.6485	-3.9503e+003	0.0103
Sample Std. Deviation	8.059e-006	0.0967	0.5541	1.5119e-005	0.1232	0.9973	1.4529	5.4465	199.5477	0.0063
Maximum	2.898e-005	0.1997	1.7925	2.1164e-005	0.3322	0.4755	3.6746	1.4830	-3.6836e+003	0.0438
Minimum	-2.025e-005	-0.2506	0.4343	-7.1501e-005	-0.2832	-1.8806	0.2381	-10.4998	-4.3438e+003	0.0043
Sample Skewness	0.7454	-0.8148	-0.3905	-2.5360	0.4952	0.3998	0.4770	-0.4669	-0.3830	2.3288
Sample Kurtosis	7.7222	3.9123	1.2638	10.6353	3.9363	1.2465	1.2651	1.2254	1.9128	10.0433

Table 8: FCGARCH(1,1,2) model statistics.