VOLATILIY AND VAR FORECASTING FOR THE IBEX-35 STOCK-RETURN INDEX USING FIGARCH-TYPE PROCESSES AND DIFFERENT EVALUATION CRITERIA

Trino-Manuel Ñiguez

ABSTRACT

In this paper I analyze the relative performance of Gaussian and Student-t GARCH and FIGARCH type models for volatility and Value-at-Risk forecasting of daily stock-returns using data from the Spanish equity index IBEX-35. The in-sample analysis shows that the Student-t FIAPARCH process provides a better fit than the nested models. Regarding the out-of-sample volatility forecasting, both the Gaussian- and the t-FIAPARCH processes show the best performance, although it is not possible to discriminate between them. As for the models' capacity for VaR forecasting, different results are obtained according to the evaluation criteria considered, although if the aim is regulatory VaR it is shown that the Student-t FIAPARCH model would be clearly the most recommendable.

Key Words: APARCH, Fractional Integration, Leverage Effect, Long Memory, Value-at-Risk.

JEL Classification: C32, C52, C53, G15.

1 Introduction

Parametric models for volatility have undergone great development since the seminal ARCH and GARCH models of Engle (1982) and Bollerslev (1986) (see Bollerslev, Engle and Nelson, 1994, for a review of the ARCH literature). In most of the empirical work related to the modelling of financial time series volatility by means of GARCH models, it has been observed that the sum of the parameter estimates remained very near to the stationarity bound of one, which is interpreted as there being long-run temporal dependence in volatility. To account for this fact, Engle and Bollerslev (1986), imposing directly unitary roots, proposed the integrated GARCH (IGARCH) model. This process provides a cumulative impulse response function that is constant over time, which is not very reliable. On the other hand, the covariance-stationary GARCH process generates autocorrelations that decrease excessively quickly in relation to the observed ones of the volatility proxies (absolute value and squared returns), (see Ding and Granger, 1996, or Granger and Ding, 1996). Fractionally integrated GARCH (FIGARCH) models were proposed by Baillie, Bollerslev and Mikkelsen (1996) to capture the hyperbolic decay observed in the autocorrelation function of the volatility proxies (see also Baillie, 1996, or Bollerslev and Mikkelsen, 1996). These processes are an adaptation for the conditional variance of the ARFIMA class of models (see Granger, 1980, Granger and Joyeux, 1980, or Hosking, 1981). The analysis of their performance for the modelling of the volatility of financial time series rapidly attracted the attention of researchers, since this variable is a key input in the calculation of very used financial measures such as, the hedging ratio or the so-called Value-at-Risk (see Jorion, 1997, for a complete overview of this topic).

In this work I aim to analyze the predictive capacity, not only for in- but also for out-of-sample volatility and Value-at-Risk (VaR hereafter) forecasting, of a broad range of Gaussian and Student-t short- and long-memory conditional heteroskedasticity models, including: GARCH, AGARCH, APARCH, EWMA, FIGARCH and FIAPARCH, for the Madrid Stock Market Index IBEX-35. Among these, we pay special attention to the Student-t FIAPARCH method since this seems the candidate to show a better forecasting performance, given that it is capable of reproducing all the stylized features of the volatility proxies, namely, leptokurtosis, clustering, "leverage" effect and hyperbolic decay of the autocorrelations. Apart from presenting greater flexibility, since it consists of regressing a power, δ, of the standard deviation, leaving it free to be determined by the data (see Ding, Granger and Engle, 1993).

There are many papers in which the in- and out-of-sample forecasting ability of short-memory GARCH models has been analyzed (see e.g. Pagan and Schwert, 1990, Brailsford and Faff, 1996, or McKenzie and Mitchell, 2002). On the other hand, the in-sample performance of univariate long-memory models for the conditional volatility has also been widely explored. However, their out-of-sample analysis, in spite of the fact that it has more important implications for risk management, seems to have received much less attention. Thus, Tse (1998) analyzed the in-sample goodness-of-fit of the Gaussian-FIAPARCH model and, Beine et al. (2002) performed an in-sample analysis for the Student-FIGARCH model as well. More recently, Vilasuso (2002) has analyzed the out-of-sample forecasting performance of the FIGARCH model with Gaussian innovations. All of these analyses were performed for exchange rate returns.

On the other hand, several papers concerning the in-sample volatility modelling of the IBEX-35 index returns have been developed. Thus, León and Mora (1999) analyzed the performance of a wide variety of volatility models concluding that the parametric ones exhibited a better performance. Olmeda (1998) and, Marmol and Rebolledo (2000) gave evidence on the existence of long-memory in the volatility of different return indexes from the Madrid Stock Market, the former using non-parametric tests and, the latter focusing on tests by Perron and, Ng, Lobato and Robinson, as well as on the estimation of a stochastic volatility model (see also Pérez and Ruiz, 2000).

The objective of this paper is to calibrate the effect on the forecasting capacity of making the models capable of explaining the different empirical features presented by the volatility of the stock-returns. Furthermore, it provides an analysis on VaR model evaluations. It is worth emphasizing that this topic has recently raised an important debate in the literature due to their implications for risk management and international banking system regulation (see López, 1999). In relation to this, the paper shows that the obtained outcomes change significantly according to the evaluation criteria used.

The remainder of the paper is organized as follows. Section 2 presents the data. Section 3 compiles the theoretical models. Section 4 deals with the models' estimation and the in-sample analysis. Section 5 reports the out-of-sample volatility forecast evaluation results. Section 6 describes the setting for VaR forecasting and discusses the prediction results, and finally, Section 7 summarizes the conclusions.

2 The data

The data consists of daily closing prices of the stock index IBEX-35 from July 1, 1987 to December 30, 2002, for a total of 3,993 observations depicted in Figure 1. The series of continuously compounded daily returns were

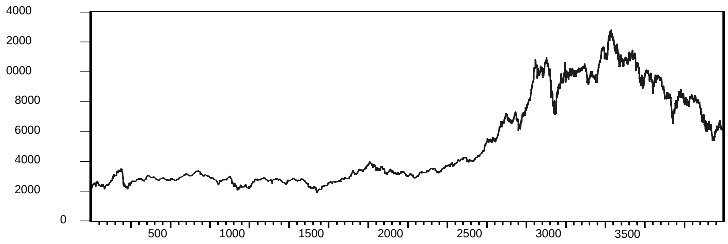

Figure 1: IBEX-35 daily closing prices, 1/07/1987-12/30/2002

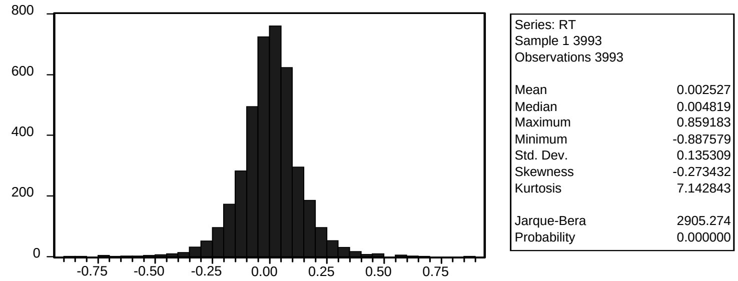

Figure 2: Histogram and descriptive statistics of r_t

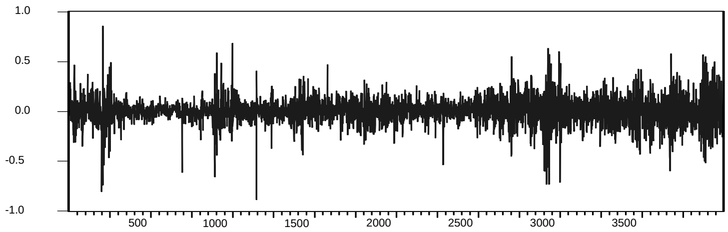

Figure 3: IBEX-35 daily returns, 1/07/1987-12/30/2002

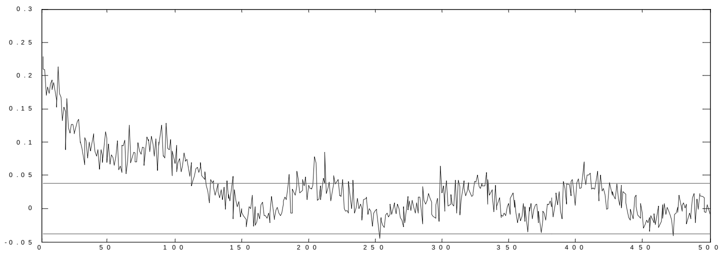

Figure 4: Sample autocorrelations of r_t, 1/07/1987-12/30/2002

calculated as the difference between the logarithm of the closing prices of two consecutive days, $r_t = \log(P_t/P_{t\text{-}1})$.

Figure 2 summarizes the descriptive statistics of the r_t series. The measures of skeweness and kurtosis indicate that the distribution of returns is leptokurtic and slightly negatively skewed in relation to the normal distribution. The Jarque-Bera statistic rejects the null hypothesis of normality at any level of statistical significance, so another distribution rather than the Normal one should be assumed for the returns. Along these lines, many empirical works have shown that the Student-t distribution is able to capture all excess kurtosis presented by the stock returns distribution, since their tail lengths are not fixed but it adjusts to the one of the data (see Bollerslev, 1987).

From Figure 3 we can see that the variance of the returns presents clusters which change over time, i.e., great changes in returns tend to be followed by great changes, of any sign, and small changes tend to be followed by small changes. It is shown that, although according to the Efficient Market Theory returns are almost unpredictable, they are not independent. This feature can be, fairly easily, noticed analytically as well, by performing the ARCH LM test, (see Engle, 1982), which shows that the null of no high-order ARCH for the residuals in Eq.(15) is rejected at any significance level, so not only is the series ε_t^2 autocorrelated but it also presents long-run temporal dependence. This fact can also be verified graphically by observing the hyperbolic decay of the sample autocorrelations of r_t^2 depicted in Figure 4.

Another important "stylized fact" of financial time series that any valuable model should convincingly explain is the so-called "leverage" effect, first discussed by Black (1976), who observed that volatility tends to increase less in response to "good news" (excess returns higher than expected, $\varepsilon_t > 0$) than in response to "bad news" (excess returns lower than expected, $\varepsilon_t < 0$). The term $(|\varepsilon_t| - \gamma\varepsilon_t)^\delta$ reflects this effect when $\gamma \neq 0$, so when $0 < \gamma < 1$ a positive innovation increases volatility less than a negative one, and vice versa for $-1 < \gamma < 0$. When $\gamma = 0$ a positive innovation has the same effect on volatility as a negative one of the same magnitude.

3 The theoretical models

Given a specification for the conditional mean,

$$r_t = g(\boldsymbol{\theta}; \Omega_{t-1}) + \varepsilon_t \tag{1}$$

where $g(.)$ is a linear function on the parameter vector $\boldsymbol{\theta}$ and the past information set, Ω_{t-1}, and ε_t is the random innovation, the ARCH(p) model was proposed by Engle (1982) to explain the clustering observed in ε_t^2. The process is given by the following equations,

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2, \quad \varepsilon_t = \sigma_t z_t, \quad \varepsilon_t/\Omega_{t-1} \sim \mathbf{D}(0,\sigma_t^2) \tag{2}$$

where $\omega > 0$, $\alpha(L) = \sum_{k=0}^{p} L^k \varepsilon_{t-k}^2$ with $\alpha_k \geq 0$ for all k, L denotes the lag operator and, $z_t \sim_{iid} \mathbf{D}(0,1)$.

The Generalized ARCH (GARCH) model was latter proposed by Bollerslev (1986) to avoid the high-order ARCH specification demanded to capture the conditional variance dynamics. The GARCH(p,q) model is expressed as,

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2 \tag{3}$$

where $\omega > 0$, $\alpha_k \geq 0$ for all k, and $\beta(L) \equiv \sum_{s=1}^{q} \beta_s L^s$ with $\beta_s \geq 0$ for all s. For stability all the roots of $\alpha(L)$ and $[1 - \alpha(L) - \beta(L)]$ have to be outside the unit circle. The process can also be expressed in ARMA(m,p) form for ε_t^2 as follows,

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t \tag{4}$$

where $m = \max\{p,q\}$ and $v_t \equiv \varepsilon_t^2 - \sigma_t^2$ are the innovations, with $E(v_t) = 0$, $E_{t-1}(v_t) = 0$ and, $E_{t-1}(v_t^2) = \sigma^2$. When the polynomial $[1 - \alpha(L) - \beta(L)]$ contains a unitary root we have the Integrated GARCH model, IGARCH(p,q), of Engle and Bollerslev (1986) which, in its ARMA form, is given by,

$$\phi(L)(1 - L)\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t \tag{5}$$

where $\phi(L) \equiv [1 - \alpha(L) - \beta(L)](1 - L)^{-1}$ is of order $m - 1$.

Baillie, Bollerslev and Mikkelsen (1996) proposed the FIGARCH(p,d,q) method to capture the so-called long-memory property observed in the volatility of asset return series. The model is obtained by replacing the differencing operator in Eq.(5) with the fractional polynomial $(1-L)^d$. The process in its ARFIMA form for ε_t^2 is defined as,

$$\phi(L)(1 - L)^d\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t \tag{6}$$

where $0 \leq d \leq 1$. An alternative representation for the FIGARCH(p,d,q) model is the following,

$$\sigma_t^2 = \omega[1 - \beta(1)]^{-1} + [1 - [(1 - \beta(L)]^{-1}\phi(L)(1 - L)^d]\varepsilon_t^2 \tag{7}$$

where the CIRW, which are given by the coefficients in the lag polynomial, $\lambda(L)$,

$$\lambda(L) = 1 - [(1 - \beta(L)]^{-1}\phi(L)(1 - L)^d \tag{8}$$

must be all positive to ensure the non-negativity of σ_t^2. The fractional differencing operator has a binomial expansion which is most conveniently expressed in terms of the hypergeometric function as follows,

$$(1-L)^d = F(-d,1;1;L) = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} L^k \tag{9}$$

The process is not covariance-stationary but is strictly stationary and ergodic, a property that it shares with the IGARCH model, (see Bougerold and Picard, 1992). It is capable of capturing the very slow decay observed in the autocorrelations of the returns volatility proxies, since its cumulative impulse response weights (CIRW hereafter), $\{\lambda_k\}_{k=1}^{\infty}$, decay hyperbolically towards zero. So, innovations affect the optimal forecast of the FIGARCH conditional variance over long lags, unlike short-memory GARCH processes for which the influence of shocks to the process disappears exponentially or, IGARCH processes in which shocks show infinite persistence[1].

On the other hand, Ding, Granger and Engle (1993) generalized the GARCH process Eq.(3) by considering the modelling of a power of the conditional standard deviation, which is made to depend on: A function of past residuals to account for the "leverage" effect, and powers of past conditional standard deviations. The process was called Asymmetric Power ARCH, APARCH(p,q), and is defined as,

$$\sigma_t^\delta = \omega + \sum_{k=1}^{p} \alpha_k (|\varepsilon_{t-k}| - \gamma_k \varepsilon_{t-k})^\delta + \sum_{s=1}^{q} \beta_s \sigma_{t-s}^\delta \tag{10}$$

where $\omega > 0$, $\delta > 0$, $|\gamma_k| < 1$, $\alpha_k \geq 0$ for all k, and $\beta_s \geq 0$ for all s.

The Fractionally Integrated Asymmetric Power GARCH method, FIAPARCH(p,d,q), (Tse, 1998), combines the FIGARCH(p,d,q) model Eq.(7) and the APARCH(p,d,q) model Eq.(10), and can be defined as follows,

$$\sigma_t^\delta = \frac{\omega}{1-\beta(1)} + \{1 - [1-\beta(L)]^{-1}\phi(L)(1-L)^d\}(|\varepsilon_t| - \gamma_k \varepsilon_t)^\delta \tag{11}$$

The process, just as the FIGARCH(p,d,q) and the IGARCH(p,q) processes, is not covariance-stationary, since the hypergeometric function

[1] Recently, Davidson (2003) has proposed a more general class of long-memory GARCH processes, called hyperbolic GARCH (HYGARCH). These processes are able to reproduce, a covariance-stationary sequence of conditional variances with long-run dependences for a determined range of the parameters, since they present a cumulative impulse response function that decreases hyperbolically, quicker than in the FIGARCH case but slower than in the covariance-stationary GARCH case (see Ñíguez and Rubia, 2003, for an extension of these types of processes to a multivariate context).

$F(-d,1,1;L)$ evaluated at $L = 1$ equals zero, so $\boldsymbol{\lambda}(1) = 1$. See the Appendix for a review of the nested models in the FIAPARCH process.

In this study, I consider the one-order version of the following processes, which has proven to be useful in the modelling of financial time series volatility, so the methods to be examined, taking as a reference the FIGARCH(1,d,1) and the GARCH(1,1) models, are: FIAPARCH(1,d,1), FIGARCH(1,d,1), APARCH(1,1), AGARCH(1,1), GARCH(1,1) and the exponential weighted moving average, EWMA(1,1), process, proposed by J.P. Morgan for VaR forecasting (see Riskmetrics, 1996, for further details). The analysis focuses, in particular, on the FIAPARCH(1,d,1) specification, Eq.(11), for which the CIRW, $\{\lambda_k\}_{k=0}^{\infty}$, can be expressed as,

$$\lambda_k = \begin{cases} 0, & k = 0 \\ d - \beta + \phi, & k = 1 \\ \beta\lambda_{k-1} + \left(\frac{k-1-d}{k} - \phi\right)\delta_{k-1}, & k \geq 2 \end{cases} \tag{12}$$

with

$$\delta_k = d\frac{\Gamma(k-d)}{\Gamma(1-d)\Gamma(k+1)} \quad \text{for} \quad k \geq 1 \tag{13}$$

For the model to be well-defined and the conditional standard deviation power to be positive almost surely for all t, all the CIRW must be positive. Bollerslev and Mikkelsen (1996) show that from the recursions $\{\lambda_k\}_{k=1}^{\infty}$ it follows easily that the inequality constraints

$$\beta - d \leq \phi \leq \tfrac{2-d}{3}, \quad d(\phi - \tfrac{1-d}{2}) \leq \beta\,(\phi - \beta + d) \tag{14}$$

are sufficient to ensure that all λ_k are non-negative[2]. Unfortunately, the statistical properties of this model have not been established yet, so that this topic continues to be an interesting area for further research.

4 Estimation and inference

In our case, as in many applications with high-frequency financial returns, the assumption of conditionally normal distributed innovations is not supported by the data. However, following Bollerslev and Wooldridge (1992) one can estimate GARCH-type models by the Quasi-Maximum Likelihood procedure (QMLE) and obtain robust estimates for density misspecification providing that the model is correctly specified.

[2] See Chung (1999) for a slightly different FIGARCH specification, extendable to the asymmetic-power case, for which the sufficient non-negativity conditions are given by $0 < \beta < d < \phi < 1$.

At this stage, it is worth pointing out that when the QMLE is applied to strongly dependent time series we cannot ensure theoretically that the estimators are consistent and asymptotically normal. Nevertheless, Baillie, Bollerslev and Mikkelsen (1996) showed, by means of simulation techniques, that the procedure performs very well for the typical sample sizes encountered with high-frequency financial data (see also Caporin, 2002). Furthermore, they showed that the "robust" covariance matrix estimator does a good job of adjusting for conditional non-normality in the errors, since the estimation results of the FIGARCH model with conditionally normal and t-distributed errors were very close. They also provided results about the simulated tail rejection frequencies for the estimates significativity, showing that they remained very close to the nominal levels although with a smooth trend to overreject in the right-tail, so in practise standard inference procedures regarding the parameter estimates can be carefully used considering these results. Unfortunately, neither for the FIAPARCH process neither theoretical nor empirical asymptotic properties for the QML estimates have been formally established, remaining this topic as an interesting area for further research (see Hidalgo, 1997, for an analysis of the theoretical properties of non-parametric estimation procedures under long-run temporal dependence).

Being aware of the previously mentioned limitations in relation to the estimation procedures, the different aforementioned conditionally heteroskedastic models are estimated and analyzed according to their forecasting performance. Thus, we proceed as follows: The complete sample, r_t, is split into two parts, the in-sample period takes the first T= $3,093$ observations and the out-of-sample period includes the last N= 900 observations. The in-sample period is then used to calibrate parameters and make one-step-ahead forecast. The process is repeated 900 times keeping a constant-sized window. The estimation procedure is performed in two stages:

First, the small predictable component detected in the r_t series[3] is filtered by using a moving average process, MA(1), which has been chosen following the usual methodology of Box and Jenkins and the information criterion of Akaike,

$$r_t = \mu + \theta\varepsilon_{t-1} + \varepsilon_t, \quad E(\varepsilon_t) = 0, \;\; V(\varepsilon_t) = \sigma^2 \tag{15}$$

This process is estimated by the Ordinary Least Squares method and the MA(1) residuals are obtained. The procedure is repeated N times, taking a rolling window of size T, to obtain the residuals out-of-sample.

Second, the conditional variance of the MA(1) residuals, ε_t, is estimated assuming on the one hand, Gaussian innovations (QMLE) and, on the other,

[3] The existence of small linear dependences in index returns has been attributed to non-synchronous trading of the stocks that make up the index.

t-innovations (MLE). Thus, QML estimates are obtained by maximizing the Gaussian log-likelihood function which is given by,

$$\sum_{t=1}^{T} \log f(r_t/\Omega_{t-1};\boldsymbol{\theta}) = -0.5\sum_{t=1}^{T}\log 2\pi\sigma_t^2 - 0.5\sum_{t=1}^{T}\frac{\varepsilon_t^2}{\sigma_t^2} \tag{16}$$

Alternatively, the Student-t distribution is assumed in order to provide the models with the needed flexibility to capture the observed kurtosis. Note that, since the Student-t approaches the Normal as $\nu \to \infty$ it is more convenient to use the reciprocal of the degrees of freedom parameter, i.e. $\eta = 1/\nu$, as a measure of the tail thickness, which will always remain in the finite range $0 \leq \eta < \frac{1}{2}$ (see Fiorentini, Sentana and Calzolari, 2003). So in this case, the function to be maximized is,

$$\begin{aligned}\sum_{t=1}^{T} \log f(r_t/\Omega_{t-1};\theta) &= T\log\left[\frac{\Gamma[(\eta+1)/2\eta]}{\pi^{1/2}\Gamma(1/2\eta)}(\frac{1-2\eta}{\eta})^{-1/2}\right] \\ &\quad -(1/2)\sum_{t=1}^{T}\log(\sigma_t^2) \\ &\quad -\frac{\eta+1}{2\eta}\sum_{t=1}^{T}\log\left[1+\frac{\eta\varepsilon_t^2}{(1-2\eta)\sigma_t^2}\right]\end{aligned} \tag{17}$$

To apply these procedures to FIGARCH-type models the truncation of the infinite lag polynomial is necessary, so it is fixed at the usual lag of 1000. Furthermore, following standard practise as well, the unconditional standard deviation was chosen for the residual presample values, $\{\varepsilon_t\}_{t=-999}^{0}$.[4].

Tables 1 and 2 report the (Q)ML estimates and the statistics for the in-sample analysis. As we can see, all parameter estimates are statistically significant: The estimated degrees of freedom, $\widehat{\nu}$ is around 7, confirming the existence of leptokurtosis in the returns conditional distribution, $\hat{\delta}$ is statistically different from one or two in APARCH and FIAPARCH models and, $\hat{\gamma}$ and $\widehat{d}$, are both statistically different from zero or one in all long-memory and/or asymmetric models, which confirms that there exists long-run dependence, and asymmetric responses of volatility to negative and positive innovations.

The likelihood ratio (**LR**) test is used to compare the in-sample performance of nested models. The test statistic has an asymptotic distribution with degrees of freedom equal to the number of restrictions, thus if l_0 is the $\log L$ value under the null hypothesis that the true model is, say Student-t FIGARCH(1,d,1), and l is the log-likelihood under the alternative that the true model is Student-t FIAPARCH(1,d,1), we have

[4]All maximum likelihood estimations were carried out using the CML library of GAUSS.

Table 1

(Q)ML estimates of GARCH(1,1)-type models

Gaussian	GARCH	AGARCH	APARCH	EWMA
ω	0.000 (2.55)	0.000 (2.54)	0.000 (1.73)	-
α	0.132 (5.95)	0.127 (5.37)	0.128 (5.56)	0.066 (3.69)
β	0.824 (29.7)	0.824 (28.0)	0.827 (23.0)	0.934 (51.8)
γ	-	0.148 (2.03)	0.154 (1.94)	-
δ	-	-	1.914 (5.65)	-
LogL	2485.7	2493.0	2493.1	2327.8
AIC	-1.6052	-1.6094	-1.6089	-1.5046
Student-t	GARCH	AGARCH	APARCH	EWMA
ω	0.000 (2.14)	0.000 (2.94)	0.001 (2.46)	-
α	0.142 (7.00)	0.140 (6.94)	0.143 (8.08)	0.091 (6.94)
β	0.849 (41.5)	0.849 (40.9)	0.865 (46.5)	0.909 (68.6)
γ	-	0.078 (2.15)	0.098 (2.25)	-
δ	-	-	1.402 (9.28)	-
ν	6.791 (6.93)	6.834 (6.91)	6.732 (6.99)	7.404 (8.43)
LogL	2653.8	2655.8	2661.1	2626.6
AIC	-1.7134	-1.7141	-1.7168	-1.6971

The reported coefficients shown in each row of the table are (Q)ML estimates of Gaussian and Student-t GARCH(1,1)-type models with MA(1) innovations for the IBEX-35 daily stock-return index. Robust (Q)ML t-statistics are reported in parenthesis next to the parameter estimates. AIC denotes the Akaike Information Criterion and LogL the value of the log-likelihood function at the parameter estimates.

Table 2

(Q)ML estimates of FIGARCH(1,d,1)-type models

Gaussian	FIGARCH		FIAPARCH	
ω	0.001	(1.43)	0.015	(1.86)
β	0.636	(2.05)	0.698	(10.7)
ϕ	0.171	(1.63)	0.232	(2.83)
γ	-		0.292	(2.38)
δ	-		1.242	(7.46)
d	0.584	(1.45)	0.576	(5.45)
LogL	2485.7		2512.1	
AIC	-1.6047		-1.6206	
Student-t	**FIGARCH**		**FIAPARCH**	
ω	0.000	(2.59)	0.006	(1.61)
β	0.634	(8.50)	0.681	(9.01)
ϕ	0.142	(2.73)	0.153	(3.48)
γ	-		0.140	(2.87)
δ	-		1.465	(7.30)
d	0.628	(8.05)	0.654	(7.90)
ν	7.018	(9.09)	7.505	(6.58)
LogL	2661.9		2669.8	
AIC	-1.7180		-1.7218	

The reported coefficients shown in each row of the table are (Q)ML estimates of Gaussian and Student-t FIGARCH(1,d,1)-type models with MA(1) innovations for the IBEX-35 daily stock-return index. Robust (Q)ML t-statistics are reported in parenthesis next to the parameter estimates. AIC denotes the Akaike Information Criterion and LogL the value of the log-likelihood function at the parameter estimates.

that $H_0 : (\delta,\gamma) = (2,0)$, so the test statistic $\mathbf{LR} = 2(l - l_0)$ should have a $\boldsymbol{\chi}^2$ distribution with 2 degrees of freedom under the null. For our case, $LR \simeq 76$, so we reject H_0 for any reasonable significance level. Likewise, it can easily be seen that the **LR** test statistic for the Student-t FIAPARCH(1,d,1) model against the rest of the specifications is beyond the acceptance region for any reasonable confidence level.

According to the **AIC** we can see that the assumption of Student-t errors is the one which provides the models with more flexibility, since the worst Student-t model in adjusting the data, i.e. the Student-EWMA ($AIC = -1.6971$), still provides a better fit than the most flexible Gaussian model, i.e. the Gaussian-FIAPARCH ($AIC = -1.6206$). This means that the greater possibility for the model to achieve extreme values plays a more important role than its ability to account for long-memory and/or asymmetric effects.

In relation to the modelling of the power of the conditional standard deviation, we find that this possibility provides the model with a greater capacity to fit the data only when assuming Student-t errors, since the Student-t APARCH model adjusts better than the Student-t AGARCH process, curiously, the opposite occurs under Gaussianity. Notice also that under this latter distribution, the GARCH, AGARCH and APARCH models perform better than the simpler long-memory process, FIGARCH. On the other hand the Student-FIGARCH provides a much better fit than the short-memory Student-t GARCH models. So the Student-t distribution enhance the in-sample forecasting capacity of the models, especially for the fractional integration case.

In summary, the in-sample volatility forecasting performance of the methods increases with their ability to account for the "stylized" volatility features, with the t-distribution assumption playing the most important role, being the Student-FIAPARCH model the most flexible one which provides the best fit, and the IGARCH-type process, i.e. the EWMA model, being the one which provides the worst goodness-of-fit.

5 Out-of-sample volatility forecasting

In this section, the models are evaluated according to their out-of-sample volatility forecasting capacity. Forecasts are made, estimating the models each time, and keeping a constant-sized rolling scheme. The procedure employed to forecast the conditional variance is outlined as follows:

i) The parameter estimates are used to forecast the conditional variance one-day-ahead from the last day of the in-sample period. The whole procedure (estimation and forecast) is then repeated N times by using a rolling window. At each iteration, the in-sample window is updated one day and the oldest observation is removed, so the window's length, T, always remains constant over the process. Models are completely re-estimated and the forecast is made again from the last observation in the in-sample period through an iterative process. The one-day-ahead forecast error from model m, denoted $e_{m,\mathrm{T}+i+1}$, (i = 0,...,N) is given by the difference between the proxy of 'true' variance, denoted $vol_{\mathrm{T}+i+1}$, which is chosen as the squared MA(1) residuals[5], and the conditional variance forecast, $E_{\mathrm{T}+i}(\varepsilon^2_{\mathrm{T}+i+1}) = \widehat{\sigma}^2_{m,\mathrm{T}+i+1}$, so,

$$vol_{\mathrm{T}+i+1} = \left[r_{\mathrm{T}+i+1} - E_{\mathrm{T}+i}(r_{\mathrm{T}+i+1})\right]^2, \quad E_{\mathrm{T}+i}(r_{\mathrm{T}+i+1}) = \mu + \theta\varepsilon_{\mathrm{T}+i+1}$$

ii) Several criteria are then used to assess the forecasting performance of the models, including: First, the Minzer-Zarnowitz Regression (1969), which consists of estimating the following equation,

$$vol_{\mathrm{T}+i+1} = c + \beta\widehat{\sigma}^2_{m,\mathrm{T}+i+1} + u_{\mathrm{T}+i+1} \tag{18}$$

According to this method, the forecast is optimal, with respect to the available information set, if the null $\mathbf{H}_0 : (c, \beta) = (0, 1)$ is accepted.

Second, the following symmetric and asymmetric loss functions: The mean squared prediction error (MPSE), the mean absolute prediction error (MAPE),

$$MSPE_m = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(e_{m,\mathrm{T}+i+1})^2 \tag{19}$$

$$MAPE_m = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left|e_{m,\mathrm{T}+i+1}\right| \tag{20}$$

the mean mixed error of underprediction (MMEU), and the mean mixed error of overprediction (MMEO) which penalize under- and over-predictions, respectively, more heavily,

$$MMEU_m = \frac{1}{\mathrm{N}}\left(\sum_{i=1}^{\mathrm{N}_U}\sqrt{\left|e_{m,\mathrm{T}+i+1}\right|} + \sum_{t=1}^{\mathrm{N}_O}\left|e_{m,\mathrm{T}+i+1}\right|\right) \tag{21}$$

[5] Andersen and Bollerslev (1998) have shown that for high-frequency assets data a better proxy for their 'true' volatility can be obtained as the sum of the squared intradaily returns, which is called 'realized' volatility. In our case, as in many empirical works, a problem arises concerning data availability, nonetheless this aspect remains as an interesting possibility to be considered in further research.

$$MMEO_m = \frac{1}{\mathrm{N}} \left(\sum_{i=1}^{\mathrm{N}_U} \left| e_{m,\mathrm{T}+i+1} \right| + \sum_{i=1}^{\mathrm{N}_O} \sqrt{\left| e_{m,\mathrm{T}+i+1} \right|} \right) \tag{22}$$

where N_U is the number of times that the predicted conditional variance is smaller than the realized one and N_O its complementary (see Brailsford et al., 1996).

iii) Finally, we verify whether the difference between the loss functions from the different models is statistically significant by performing the test of Diebold and Mariano (1995)[6]. It consists of testing whether the sample mean of the difference series between the forecasting errors through the out-of-sample period, $\overline{\mathbf{d}} = \sum_{i=1}^{\mathrm{N}} \frac{d_i}{\mathrm{N}}$, with $d_i = e_{m,\mathrm{T}+i+1} - e_{m',\mathrm{T}+i+1}$, is different from zero,

$$\mathrm{H}_0 : \mu_d = 0, \qquad \sqrt{\mathrm{N}}(\overline{\mathbf{d}} - \mu_d) \rightarrow N\left(0, \sigma_d^2\right) \tag{23}$$

Thus, the test statistic,

$$t^* = \sqrt{\mathrm{N}} \frac{\overline{\mathbf{d}}}{\widehat{\sigma}_d} \tag{24}$$

converges in distribution to a $\mathbf{N(0,1)}$ under the null. If the forecast is optimal, the K-step ahead errors, K being the prediction horizon, are at most independent of all previous errors, so a consistent estimator has to be used to control for this $K-1$ order dependence. The Newey and West (1987) estimator, $\widehat{\sigma}^2_{NW}$, which is consistent under heteroskedasticity and autocorrelation with a bandwidth parameter $q = K - 1$, is generally used for that purpose,

$$\widehat{\sigma}^2_{NW} = \widehat{\sigma} + \frac{1}{\mathrm{N}} \sum_{s=1}^{q} \sum_{i=s+1}^{\mathrm{N}} \left(1 - \frac{s}{q+1} \right) \left(\mathbf{d}_i \mathbf{d}'_{i-s} + \mathbf{d}_{i-s} \mathbf{d}'_i \right) \tag{25}$$

Note that, in our case, as the forecast horizon is of one step ahead, there is no dependence in our difference series, so $\widehat{\sigma}^2_{NW}$ is given by the unconditional variance.

Table 3 reports the results for the forecasts evaluation according to the Minzer-Zarnowitz Regression. It can be seen that the null of optimal forecast is accepted in most of the cases for a significance level of at least 5 per cent,

[6] Alternatively, boostrap procedures could be used for this purpose. In our case, since the forecast horizon is one day, boostrap procedures for independent data would be useful. However when the forecasting horizon is greater than one period, the prediction errors show autocorrelation, so block boostrap procedures would be needed (see Politis, Romano and Wolf, 1999, or White, 2000).

Table 3

Test for the parameters of the Mincer-Zarnowitz Regression

	Gaussian		Student-t	
	p-value	R^2_m	p-value	R^2_m
GARCH	0.072	0.1145	0.018	0.1156
APARCH	0.053	0.1322	0.142	0.1291
AGARCH	0.055	0.1383	0.244	0.1331
EWMA	0.073	0.1142	0.015	0.1195
FIGARCH	0.665	0.1101	0.032	0.1167
FIAPARCH	0.294	0.1447	0.763	0.1400

P-values for the test of Mincer-Zarnowitz Eq.(18), $H_0 : (c, \beta) = (0, 1)$. R^2_m denotes the determination coefficient of the regression for model m.

and that the percentage of explained volatility, is similar and greater than 10 per cent in all cases. Asymmetric models, and above all the FIAPARCH models, provide sequences of forecasts that adjust better to the volatility proxy. The assumption of t-distributed errors helps, only in some cases, to yield better forecasts. This criterion provides a general view of the models forecasting performance but a more exhaustive analysis is called for, to be able to discriminate among models.

Loss functions values and the results of the test of Diebold and Mariano are displayed in Tables 4 and 5, respectively. Results in Table 4 show that the FIAPARCH model presents the smallest MSPE and MAPE, and these differences are statistically significant in relation to the benchmarking models as we can see from Table 5. In general we can not conclude that the assumption of Student-t innovation helps to achieve more accurate forecats. On the other hand, the asymmetric models provide, in all cases, a significantly better performance than the GARCH, FIGARCH and EWMA processes. In addition, these latter models seem to perform very similarly according to the results emerging from Table 5.

As for the asymmetric loss functions, we observe from Table 5 that their differences are statistically significant, except for some isolated cases among the FIGARCH, GARCH and EWMA models. Thus, from Table 4 it can inferred that Student-t models tend to overpredict volatility since they show a higher error of overprediction (MMEO) and, a smaller error and probability of underprediction (MMEU and PUnd, respectively), with the exception of the t-FIAPARCH model which provides the smallest MMEO and MMEU.

A further analysis in Table 5 revealed that no significant differences existed between the values of the symmetric loss functions of the FIAPARCH models, so both models provide the more accurate volatility forecasts, the t-FIAPARCH process being the most recommendable since it shows a significantly smaller error of overprediction.

6 VaR analysis

6.1 Introduction

VaR measures can have many applications in financial markets, but they are mostly used for risk management and regulatory purposes. In relation to the latter topic, the Capital Accord of 1996 of the Basle Committee on Banking Supervision at the Bank for International Settlements, impelled to financial institutions with significant trading activities to report the VaR of their investment portfolios periodically. According to these risk predictions, the capital that banks must hold to cover their exposure to

Table 4

Out-of-sample volatility forecasting performance

Gaussian	MSPE	MAPE	MMEO	MMEU	PUnd
GARCH	0.1767	0.2524	0.0918	0.0713	0.357
AGARCH	0.1733	0.2487	0.0908	0.0710	0.361
APARCH	0.1721	0.2485	0.0914	0.0705	0.353
EWMA	0.1768	0.2604	0.0985	0.0690	0.326
FIGARCH	0.1768	0.2587	0.0987	0.0687	0.321
FIAPARCH	0.1701	0.2534	0.0978	0.0675	0.318
Student-t	**MSPE**	**MAPE**	**MMEO**	**MMEU**	**PUnd**
GARCH	0.1771	0.2613	0.0976	0.0696	0.338
AGARCH	0.1736	0.2575	0.0966	0.0691	0.342
APARCH	0.1722	0.2570	0.0973	0.0686	0.331
EWMA	0.1764	0.2608	0.0978	0.0693	0.333
FIGARCH	0.1766	0.2628	0.0995	0.0689	0.326
FIAPARCH	0.1708	0.2547	0.0966	0.0685	0.335

Out-of-sample forecasting performance of GARCH(1,1)-type and FIGARCH(1,d,1)-type processes for one-day-ahead conditional variance. The forecasting ability is measured by different loss functions, including: The mean squared prediction error (MSPE) Eq.(19), the mean absolute prediction error (MAPE) Eq.(20), the mean mixed error of underprediction (MMEU) Eq.(21), the mean mixed error of overprediction (MMEO) Eq.(22) and the probability of underprediction (PUnd). MSPE and MAPE values are all scaled by 100 and 10, respectively.

Tabla 5

Statistics of the test of Diebold and Mariano for volatility forecasting performance

Gaussian	MSPE	MAPE	MMEO	MMEU
FIGARCH vs GARCH	0.02	6.63	17.83	-7.92
FIGARCH vs AGARCH	1.37	5.85	13.82	-4.74
FIGARCH vs APARCH	1.95	6.32	13.26	-3.95
FIGARCH vs EWMA	0.97	-0.95	0.40	-0.84
FIGARCH vs FIAPARCH	2.67	2.93	1.28	2.29
GARCH vs AGARCH	1.50	2.76	2.65	0.80
GARCH vs APARCH	2.08	2.97	1.19	2.30
GARCH vs EWMA	-0.02	-3.81	-8.98	4.05
GARCH vs FIAPARCH	2.57	-0.65	-8.76	6.64
Student-t	**MSPE**	**MAPE**	**MMEO**	**MMEU**
FIGARCH vs GARCH	-0.49	2.12	7.03	-3.63
FIGARCH vs AGARCH	1.58	3.80	6.39	-0.68
FIGARCH vs APARCH	2.09	2.62	1.72	1.50
FIGARCH vs EWMA	0.17	1.59	3.91	-1.17
FIGARCH vs FIAPARCH	2.55	4.83	5.34	0.71
GARCH vs AGARCH	1.73	3.12	2.68	1.49
GARCH vs APARCH	1.71	2.21	-2.54	3.98
GARCH vs EWMA	0.52	0.42	-0.41	1.00
GARCH vs FIAPARCH	2.40	3.72	1.69	2.35
G-FIAPARCH vs t-FIAPARCH	-0.92	-0.66	3.52	-3.31

The table presents t-statistics for the test of Diebold and Mariano (1995) for one-day-ahead conditional variance forecasting. The comparison is performed taking as reference the GARCH(1,1) and the FIGARCH(1,d,1) models. The loss functions considered are: The mean squared prediction error (MSPE) Eq.(19), the mean absolute prediction error (MAPE) Eq.(20), the mean mixed error of underprediction (MMEU) Eq.(21) and, the mean mixed error of overprediction (MMEO) Eq.(22).

market risk is stablished. Nowadays, regulators evaluate bank VaR models by observing when the corresponding portfolio returns exceed the reported VaR, determining whether these VaR estimates are "acceptably accurate", by considering the last 250 VaR reported and their corresponding portfolio returns. It is known that evaluation methods currently used by regulators are questionable and can be improved. Those methods (binomial and interval forecast) are based on tests whose statistical power is low for discriminating among reasonable alternative models. López (1999) has proposed a VaR model evaluation test based on the magnitude of the exceptions, which has proven to be more powerful than the former one in classifying VaR models. This test along with different loss functions for the VaR estimates is used here to evaluate the VaR forecasting performance obtained from the different models considered.

6.2 VaR definition

VaR is defined as the maximum expected loss in the value of a portfolio, for a given probability α and a determined time period. The great popularity that this instrument has achieved among financial practitioners is essentially due to its conceptual simplicity, since VaR reduces the market risk associated with any portfolio to just one number. We denote the cumulative assumed conditional distribution function of r_{t+1}, as $F(r_{t+1}/\Omega_t)$, then the $VaR^{\alpha}_{m,t+1}$ of a long position is defined as

$$\alpha = \Pr\left[r_{t+1} \leq VaR^{\alpha}_{m,t+1}\right] = F^{m}(VaR^{\alpha}_{m,t+1}) \tag{26}$$

or equivalently $VaR^{\alpha}_{m,t+1}$ is the solution to

$$\int_{-\infty}^{VaR^{\alpha}_{m,t+1}} f^{m}(r_{t+1}/\Omega_t)dr_t = \alpha \tag{27}$$

where $f(r_{t+1}/\Omega_t)$ is the assumed conditional density function for r_{t+1}. This definition states that the probability of a loss r_{t+1} greater or equal to the one-day-ahead VaR estimated at time t from model m is α. In other words, $VaR^{\alpha}_{m,t+1}$ is the $100 \cdot \alpha$ percentile of $f(r_{t+1}/\Omega_t)$, when the forecasting period is of one day, i.e., $f^{-1}_{r_{t+1}}(\alpha) = VaR^{\alpha}_{m,t+1}$. For example, for $\alpha = .01$ the probability that in just one day the worst return is lower than $VaR^{\alpha}_{m,t+1}$ is one per cent.

6.3 VaR forecasting

Taking one day as the forecast horizon and, the regulatory confidence level as, e.g., 1 per cent, the $\text{VaR}^{.01}$ forecast at time t from model m when a t-distribution is assumed, is the quantile of order 1 of the Student-t conditional distribution $t_{\hat{\nu}}/\Omega_t(\widehat{\mu}_{t+1}, \widehat{\sigma}_{t+1})$, which is given by,

$$\widehat{VaR}^{.01}_{m,\mathrm{T}+i+1} = \widehat{\mu}_{\mathrm{T}+i+1} - \widehat{\sigma}_{m,\mathrm{T}+i+1} \cdot t_{\hat{\nu}_m,.01} \tag{28}$$

where $\widehat{\mu}_{t+1}$ and $\widehat{\sigma}_{m,t+1}$ are the forecast of the conditional mean and the conditional standard deviation, respectively, and, $t_{\hat{\nu}_m,.01}$ is the one percentile of the Student-t conditional distribution with $\hat{\nu}_m$ degrees of freedom, which depends on the estimation of ν from model m.

On the other hand, when a Gaussian distribution is assumed, the VaR$^{.01}$ is given by,

$$\widehat{VaR}^{.01}_{m,\mathrm{T}+i+1} = \widehat{\mu}_{\mathrm{T}+i+1} - \widehat{\sigma}_{m,\mathrm{T}+i+1} \cdot z_{.01} \tag{29}$$

where the one percentile of the standard Gaussian conditional distribution, denoted as $z_{.01}$, is constant for all models and equals 2.325.

6.4 VaR models evaluation

To analyze the ability of the different models for VaR estimation, we use two different criteria. The first is based on the same loss functions previously used for the volatility forecasting analysis and, the second, which reflects more regulatory concerns, is based on loss functions that address the number or the magnitude of the exceptions.

6.4.1 Evaluation of VaR estimates using loss functions

In this section we consider the same methodology used for the volatility forecasting analysis, although in this case the loss functions measure the error made in the VaR estimation over the out-of-sample period. The one-day-ahead VaR prediction error from model m, denoted by $eVaR_{m,\mathrm{T}+i+1}$, is calculated as,

$$eVaR_{m,\mathrm{T}+i+1} = VaR_{m,\mathrm{T}+i+1} - \widehat{VaR}_{m,\mathrm{T}+i+1} \tag{30}$$

where the "realized" VaR([7]), $VaR_{m,\mathrm{T}+i+1}$, is given by,

$$VaR_{\mathrm{T}+i+1} = r_{\mathrm{T}+i+1} + p_{\mathrm{T}+i+1} * \tilde{\sigma}_{\mathrm{T}+i+1} \tag{31}$$

and $r_{\mathrm{T}+i+1}$ is the observed return at time T+i + 1, $\tilde{\sigma}_{\mathrm{T}+i+1}$ is the chosen proxy for the standard deviation, $\tilde{\sigma}_{\mathrm{T}+i+1} = \sqrt{vol_{\mathrm{T}+i+1}}$, and $p_{\mathrm{T}+i+1}$ the

[7] Another possibility for the "realized" VaR could be to take the empirical quantile (see Jorion, 1997, for details).

percentile of the assumed distribution. Notice that while for the Gaussian case, $p_{\mathrm{T}+i+1}$ is constant over time and across models, when the t-distribution is assumed, $p_{\mathrm{T}+i+1}$ depends on the estimated degrees of freedom which, in turn, depends on the estimated model and the data distribution at time T+i + 1. Thist must be taken into account when computing the VaR prediction error under the Student-t assumption. So, the test of Diebold and Mariano for the significance of the difference between the loss functions for VaR prediction from the different models can not be performed in this latter case, since the target VaR is not the same for every model. Therefore, the information obtained from the loss functions in this case gives us a first insight about how the models can perform in predicting the VaR, and it has to be complemented by using another evaluation criteria.

Table 6 presents the values of the different loss functions for one-day-ahead $\mathrm{VaR}^{.01}$ forecasting from every model, and Table 7 reports the Diebold and Mariano test results for the significance of their difference[8]. These tables show that, contrary to what it would be expected, according to the volatility forecasting evaluation results, Gaussian models provide more accurate VaR forecasts than their Student-t counterparts. Among them, the asymmetric GARCH models (AGARCH and APARCH) show the best performance, and the EWMA and the plain FIGARCH models give the highest values of the loss functions. The same ordering is found for the Student-t models, showing, among them, the FIAPARCH model a significantly better performance.

Regarding the asymmetric loss functions we can see that the models with better VaR forecasting performance provide the smallest values of both asymmetric loss functions, showing a very similar probability of underprediction in relation to the Student-t models, which is smaller for the case of FIGARCH-type models.

As we have mentioned above, these results contradict what would have been expected after the volatility analysis, so that an alternative and/or complementary analysis is now performed to obtain more robust conclusions.

6.4.2 Evaluation of VaR estimates using regulatory loss functions

This evaluation method consists of assigning a numerical score to the VaR estimates. This score is calculated following the current regulatory framework, i.e., regulators observe the VaR estimates and portfolio returns, denoted $\left\{r_{\mathrm{T}+i+1}, \widehat{VaR}_{m,\mathrm{T}+i+1}\right\}_{i=1}^{\mathrm{N}}$, for bank (model) m and then construct

[8]In this section, I only present the results for $\alpha = .01$ for the sake of saving space and due to the results for the quantiles 0.025, 0.05 and 0.1 do not provide further information. In any case they are available from the author upon request.

Table 6

VaR Forecasting Performance

Gaussian	MSPE	MAPE	MMEO	MMEU	PUnd
GARCH	0.0830	0.2232	0.3723	0.2876	0.334
AGARCH	0.0822	0.2227	0.3723	0.2866	0.332
APARCH	0.0819	0.2229	0.3736	0.2865	0.330
EWMA	0.0853	0.2292	0.3819	0.2904	0.323
FIGARCH	0.0844	0.2288	0.3836	0.2889	0.316
FIAPARCH	0.0833	0.2291	0.3856	0.2890	0.313
Student-t	**MSPE**	**MAPE**	**MMEO**	**MMEU**	**PUnd**
GARCH	0.1171	0.2694	0.4186	0.3313	0.332
AGARCH	0.1150	0.2677	0.3290	0.4180	0.322
APARCH	0.1141	0.2673	0.4182	0.3286	0.327
EWMA	0.1146	0.2662	0.4158	0.3284	0.331
FIGARCH	0.1169	0.2700	0.4207	0.3299	0.322
FIAPARCH	0.1113	0.2645	0.4162	0.3258	0.325

Out-of-sample forecasting performance for one-day-ahead $VaR^{.01}$. The forecasting ability is measured by different loss functions, including: The mean squared prediction error (MSPE) Eq.(19), the mean absolute prediction error (MAPE) Eq.(20), the mean mixed error of underprediction (MMEU) Eq.(21), the mean mixed error of overprediction (MMEO) Eq.(22) and the probability of underprediction (PUnd).

Tabla 7

Statistics of the test of Diebold and Mariano for VaR forecasting performance

Gaussian	MSPE	MAPE	MMEO	MMEU
FIGARCH vs GARCH	3.60	7.64	11.17	1.30
FIGARCH vs AGARCH	3.00	5.31	7.69	1.59
FIGARCH vs APARCH	3.58	5.37	7.21	1.75
FIGARCH vs EWMA	-1.36	-0.40	1.04	-0.99
FIGARCH vs FIAPARCH	1.68	-0.27	-1.49	-0.05
GARCH vs AGARCH	1.24	0.63	-0.00	0.98
GARCH vs APARCH	1.75	0.36	-1.27	1.07
GARCH vs EWMA	-3.01	-4.59	-5.26	-1.64
GARCH vs FIAPARCH	-0.30	-4.62	-7.92	-0.87

The table reports t-statistics for the test of Diebold and Mariano (1995) for one-day-ahead $VaR^{.01}$. The comparison is performed taking as reference the GARCH(1,1) and the FIGARCH(1,d,1) models. The loss functions considered are: The mean squared prediction error (MSPE) Eq.(19), the mean absolute prediction error (MAPE) Eq.(20), the mean mixed error of underprediction (MMEU) Eq.(21) and, the mean mixed error of overprediction (MMEO) Eq.(22).

a numerical score under a loss function that reflects their concern[9]. The general form of these loss functions is

$$C_{m,\mathrm{T}+i+1} = \begin{cases} f(r_{\mathrm{T}+i+1}, \widehat{VaR}_{m,\mathrm{T}+i+1}) & \text{if} \quad r_{\mathrm{T}+i+1} < \widehat{VaR}_{m,\mathrm{T}+i+1} \\ g(r_{\mathrm{T}+i+1}, \widehat{VaR}_{m,\mathrm{T}+i+1}) & \text{if} \quad r_{\mathrm{T}+i+1} \geq \widehat{VaR}_{m,\mathrm{T}+i+1} \end{cases}$$

where $f(r_{\mathrm{T}+i+1}, \widehat{VaR}_{m,\mathrm{T}+i+1}) \geq g(r_{\mathrm{T}+i+1}, \widehat{VaR}_{m,\mathrm{T}+i+1})$. Numerical scores are generated for individual VaR estimates from every model, so the score for model m, C_m, for the complete regulatory sample is given by,

$$C_m = \sum_{i=1}^{\mathrm{N}} C_{m,\mathrm{T}+i+1} \tag{32}$$

Under very general conditions (see Diebold, Gunther and Tay, 1998), more accurate VaR estimates generate lower numerical scores.

Loss function that addresses the number of exceptions This function only considers the number of exceptions,

$$C_{m,\mathrm{T}+i+1} = \begin{cases} 1 & \text{if} \quad r_{\mathrm{T}+i+1} < \widehat{VaR}_{m,\mathrm{T}+i+1} \\ 0 & \text{if} \quad r_{\mathrm{T}+i+1} \geq \widehat{VaR}_{m,\mathrm{T}+i+1} \end{cases} \tag{33}$$

According to this criterion, an acceptable VaR model for regulatory purposes must produce only 9 exceptions for the full sample of 900 VaR estimates. The number of exceptions resulting from every model is reported in Table 8. It is seen that in this case only the Student-t models are accepted as valid (with the exception of the Gaussian-FIAPARCH model), and long-memory models provide more accurate regulatory VaR than the short-memory and EWMA models, but the method does not allow us to discriminate among Student-t models either, so a more powerful method is needed for this purpose.

Loss function that addresses the magnitude of the exceptions These types of loss functions that embody the magnitude of the exception can provide additional information on how the underlying VaR model forecasts the lowest part of the tail of the $f(r_{\mathrm{T}+i+1}/\Omega_{\mathrm{T}+i})$ distribution. Although several are possible, we consider the following,

[9] Here we consider an out-of-sample period of 900 observations while regulators only consider the last 250 reported VaR estimates.

$$C_{m,\mathrm{T}+i+1} = \begin{cases} [abs(r_{\mathrm{T}+i+1}) - abs(\widehat{VaR}_{m,\mathrm{T}+i+1})]^2 & \text{if } r_{\mathrm{T}+i+1} < \widehat{VaR}_{m,\mathrm{T}+i+1} \\ 0 & \text{if } r_{\mathrm{T}+i+1} \geq \widehat{VaR}_{m,\mathrm{T}+i+1} \end{cases}$$

López (1999) has shown the possibility of creating a benchmark based on the distribution of $C_{m,\mathrm{T}+i+1}$, for gauging its magnitude. This way regulators could evaluate the VaR models performance according to the distribution of $C_{m,\mathrm{T}+i+1}$. However, in order to make this benchmark operational for regulators, who need a reference model for evaluating the VaR estimations of banks, it is necessary to assume a process for $\widehat{\sigma}_{\mathrm{T}+i+1}$, so the results will depend strongly on the chosen distribution, $f(r_{\mathrm{T}+i+1}/\Omega_{\mathrm{T}+i})$. In this paper we directly use the magnitude loss function for making comparison across VaR models. So, we focus here on the best process which could be used by regulators to simulate the distribution of $C_{m,\mathrm{T}+i+1}$.

Table 8 reports the number and magnitude of the exceptions from the different models for one-day VaR of different percentile sizes (10, 5, 2.5, 1). We can see that the Student-t models show the best performance according to these criteria, the long-memory models being the one that provide the best adjustment to all quantiles, and the Student-t FIAPARCH model being the most recommendable one for regulatory purposes. On the other hand, in contrast to what we have seen in the previous sections the AGARCH and APARCH models only provide sligtly more accurate forecasts than the worst performer models (GARCH and EWMA). As regards their forecasting ability according to the size of the quantiles, we observe that under the Student-t assumption all models provide the right number of exceptions for all considered quantiles. On the other hand, the Gaussian models tend more to fail, and more so, the greater the size of the percentile is.

7 Conclusion

In this paper, different conditional heteroskedasticity models based not only on the Normal but also on the Student-t distribution have been estimated for IBEX-35 daily returns, and evaluated according to their performance for out-of-sample volatility and VaR forecasting using different criteria. The inference drawn gives evidence of that the goodness-of-fit of the FIAPARCH model with t-innovations is better these of the other models considered in the analysis. This is due to the fact that the Student-FIAPARCH model gathers all the empirical features observed in the stock-return volatility proxies whereas the others do not account for some of them. The assumption of Student-t errors plays an important role since all t-models perform better than their Gaussian counterparts.

Table 8

Number and magnitude of the exception for VaR forecasting

		Gaussian		Student-t	
		Number	Magnitude	Number	Magnitude
GARCH	$VaR^{.1}$	117	1.1537	93	0.7665
	$VaR^{.05}$	63	0.4824	39	0.2430
	$VaR^{.025}$	39	0.2185	14	0.0824
	$VaR^{.01}$	17	0.0897	5	0.0169
AGARCH	$VaR^{.1}$	115	1.1394	93	0.7490
	$VaR^{.05}$	64	0.4504	41	0.2106
	$VaR^{.025}$	38	0.1773	14	0.0626
	$VaR^{.01}$	14	0.0603	3	0.0155
APARCH	$VaR^{.1}$	113	1.1161	90	0.7420
	$VaR^{.05}$	63	0.4394	38	0.2063
	$VaR^{.025}$	34	0.1718	12	0.0603
	$VaR^{.01}$	13	0.0574	3	0.0123
EWMA	$VaR^{.1}$	106	1.0659	92	0.8148
	$VaR^{.05}$	55	0.4626	43	0.2800
	$VaR^{.025}$	30	0.2259	15	0.1063
	$VaR^{.01}$	15	0.1017	5	0.0250
FIGARCH	$VaR^{.1}$	106	1.0438	86	0.7292
	$VaR^{.05}$	51	0.4217	41	0.2066
	$VaR^{.025}$	29	0.1792	12	0.0611
	$VaR^{.01}$	10	0.0700	3	0.0119
FIAPARCH	$VaR^{.1}$	107	0.9814	90	0.7477
	$VaR^{.05}$	55	0.3642	41	0.1901
	$VaR^{.025}$	27	0.1305	11	0.0466
	$VaR^{.01}$	9	0.0390	2	0.0103

The exceptions are computed for one-day-ahead $VaR^{.1}$, $VaR^{.05}$, $VaR^{.025}$ and $VaR^{.01}$. One exception from model m occurs when the out-of-sample return, r_{t+1}, goes beyond its corresponding VaR estimate, $\widehat{VaR}_{m,\mathrm{T}+i+1}$, in that case the magnitude of the exception is computed as $[abs(r_{\mathrm{T}+i+1}) - abs(\widehat{VaR}_{m,\mathrm{T}+i+1})]^2$.

As for the out-of-sample volatility analysis, it is shown that FIAPARCH models provide more accurate forecasts than the rest of the methods, although the accuracy loss in forecasting they yield, under the assumption of Gaussian or t-distributed innovations, is not statistically significant. However, the t-FIAPARCH model would be the most recommendable when demanding a significantly smaller probability of overprediction. The evidence drawn does not allow to conclude that the assumption of Student-t innovation is no a determining factor in achieving more accurate forecats, but it biases most models to overpredict volatility. On the other hand, the worst performer models were found to be the GARCH, FIGARCH and the EWMA under both assumed distributions.

As regards the assessment of models according to their ability for VaR forecasting, we have obtained different results depending on the evaluation criteria considered. Thus, according to the criterion based on the loss functions, the Gaussian AGARCH and APARCH models showed the best performance and the Student-t EWMA and FIGARCH the worst one. On the other hand, according to the criterion based on regulatory loss functions, the Student-t models showed a much better performance than their Gaussian counterparts, the t-FIAPARCH method and the Gaussian-GARCH model being the most and the least recommendable, respectively, for regulatory purposes. Given these results, it seems that a more detailed analysis on how the assumed distribution performs in forecasting the tail of the return distribution would be desirable, but it is beyond the scope of this work.

As a final remark, it would be worth noting that the generalized use of overpredicting models for regulatory VaR, such as the EWMA model implemented in the Riskmetrics software, may lead to sub-optimal allocations of capital resources, lower rates of economic growth, thus affecting global social welfare.

References

[1] Baillie, R. T., 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5-59.

[2] Baillie, R. T., Bollerslev T. and H. Mikkelsen, 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3-30.

[3] Basle Committee on Banking Supervision, 1996. Amendment to the capital accord to incorporate market risk.

[4] Beine, M., Laurent S. and C. Lecourt, 2002. Accounting for conditional leptokurtosis and closing day effects in FIGARCH models of daily exchange rates. *Applied Financial Economics* 12, 589-600.

[5] Black, F., *in Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section* (American Statistical Association, Alexandria, VA, 1976), p.177.

[6] Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[7] Bollerslev, T., 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542-547.

[8] Bollerslev, T., Chou, R. Y. and K. F. Kroner, 1992. ARCH modelling in finance. *Journal of Econometrics* 52, 5-59.

[9] Bollerslev, T., Engle, R. F. and D. B. Nelson, 1994. *ARCH models.* Handbook of econometrics, chapter 49, volume 4. North-Holland.

[10] Bollerslev, T., H. Mikkelsen, 1996. Modelling and pricing long memory in stock market volatility. *Journal of Econometrics* 73, 151-184.

[11] Brailsford, T. J., R. W. Faff, 1996. An evaluation of volatility forecasting techniques. *Journal of Banking and Finance* 20, 419-438.

[12] Bougerold, P., N. Picard, 1992. Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* 52, 5-59.

[13] Caporin, M., 2002. FIGARCH models: stationarity, estimation methods and the identification problem. *GRETA* Working Paper no. 02.02.

[14] Chung, C. F., 1999. Estimating the fractionally integrated GARCH model. *National Taiwan University* Working Paper.

[15] Davidson, J., 2003. Moment and memory properties of linear conditional heteroskedasticity models, and a new model. Forthcoming in *Journal of Business and Economic Statistics*.

[16] Diebold, F. X., R. Mariano, 1995. Comparing predictive accuracy. *Journal of Business and Economics Statistics* 13, 253-264.

[17] Ding, Z., Granger, C. W. J. and R. F. Engle, 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83-106.

[18] Ding, Z., C. W. J. Granger, 1996. Modelling volatility persistence of speculative returns: a new approach. *Journal of Econometrics* 73, 185-215.

[19] Engle, R. F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. inflation. *Econometrica* 50, 987-1008.

[20] Engle, R. F., T. Bollerslev, 1986. Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1-50.

[21] Fiorentini, G., E. Sentana and G. Calzolari, 2003. Maximum likelihood estimation and inference in multivariate conditional heteroskedastic dynamic regression models with Student-t innovations. *CEMFI* Working Paper 0007, forthcoming in *Journal of Business and Economic Statistics*.

[22] Glosten, L. R., Jaganathan R., D. Runkle, 1989. Relationship between the expected value and the volatility of the nominal excess return on stocks. Unpublished manuscript. *Northwestern University*.

[23] Geweke, J., 1986. Modelling the persistence of conditional variance: A comment. *Econometric Reviews* 5, 57-61.

[24] Granger, C. W. J., 1980. Some properties of time series data and their use in econometric model specification. *Journal of Time Series Analysis* 1, 15-39.

[25] Granger, C. W. J., R. Joyeux, 1980. An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15-29.

[26] Granger, C. W. J., Z. Ding, 1996. Varieties of long memory models. *Journal of Econometrics* 73, 61-77.

[27] Hidalgo, J., 1997. Non-parametric Estimation with Strongly Dependent Multivariate Time-Series. *Journal of Time Series Analysis* 18, 95-122.

[28] Higgins, M., A. Bera, 1992. A class of non-linear ARCH models. International Economic Review 33, 137-158.

[29] Hosking, J. R. M., 1981. Fractional differencing. *Biometrica* 68, 165-176.

[30] Jorion, P., 1997. *Value at Risk: the new benchmark for controlling market derivatives risk*. Irwin.

[31] León, A., J. Mora, 1999. Modelling conditional heteroskedasticity: application to the IBEX 35 stock-return index. *Spanish Economic Review* 1, 215-238.

[32] López, J. A., 1999. Methods for evaluating Value-at-Risk estimates. *FRBSF Economic Policy Review* 2, 2-17.

[33] Marmol, F., J. C. Rebolledo, 2000. Dependencia temporal en la volatilidad estocástica. Un aplicación al índice IBEX 35. *Fundación Caixa Galicia. CIEF.*

[34] McKenzie, M., H. Mitchell, 2002. Generalized asymmetric power ARCH modelling of exchange rate volatility. *Applied Financial Economics* 12, 555-564.

[35] Mincer, J., V. Zarnowitz, 1969. The evaluation of economic forecasts. J. Mincer, ed., *Economic Forecasts and Expectations*, (New York: National Bureau of Economic Research).

[36] Ñíguez, T. M., A. Rubia, 2003. Forecasting the multivariate volatility of an exchange rates portfolio under long-run dependence. Forthcoming *IVIE* Working Paper.

[37] Olmeda, I., 1998. Algunos resultados sobre memoria de largo plazo en series bursatiles. *Moneda y Crédito* 207, 145-203.

[38] Pagan, A., G. Schwert, 1990. Alternative models for conditional stock volatility. *Journal of Econometrics* 45, 267-290.

[39] Pantula, S. G., 1986. Modelling the persistence of conditional variance: A comment. *Econometric Reviews* 5, 71-74.

[40] Pérez, A., E. Ruiz, 2001. Finite sample properties of a QML estimator of stochastic volatility models with long memory. *Economics Letters* 70, 2, 157-164.

[41] Politis D. N., Romano J. P. and M. Wolf, 1999. *Subsampling*. Springer-Verlag.

[42] RiskMetrics, 1996. Technical document. *J.P. Morgan Global Research*. New York.

[43] Schwert, G. W., 1990. Stock volatility and the crash of '87. *Review of Financial Studies* 3, 1, 77-102.

[44] Taylor, S. J., 1986. *Modelling Financial Time Series*. John Wiley & Sons.

[45] Tse, Y. K., 1998. The conditional heteroskedasticity of the Yen-Dollar exchange rate. *Journal of Applied Econometrics*, vol. 13, 49-55.

[46] Vilasuso, J., 2002. Forecasting exchange rate volatility. *Economic Letters* 76, 59-64.

[47] White, H, 2000. A reality check for data snooping. *Econometrica* 68, 1097-1126.

[48] Zakoian, J., 1991. Threshold heteroskedasticity model. Unpublish manuscript. *CREST. INSEE*.

APPENDIX

The FIAPARCH(p,d,q) model (Tse, 1998) includes the following specifications as special cases:

1) The FIPARCH(p,d,q) model when $\gamma_k = 0$ for all $k = 1,...,p$,

$$\sigma_t^\delta = \frac{\omega}{1-\beta(1)} + [1 - [(1-\beta(L)]^{-1}\phi(L)(1-L)^d] \left|\varepsilon_t\right|^\delta \tag{A.1}$$

2) The FIAGARCH(p,d,q) model when $\delta = 2$,

$$\sigma_t^2 = \frac{\omega}{1-\beta(1)} + [1 - [(1-\beta(L)]^{-1}\phi(L)(1-L)^d](\left|\varepsilon_t\right| - \gamma_k\varepsilon_t)^2 \tag{A.2}$$

3) The FIGARCH(p,d,q) model of Baillie, Bollerslev and Mikkelsen (1996) Eq.(7), when $\delta = 2$ and $\gamma_k = 0$ for all k.

4) The IGARCH(p,q) model of Engle and Bollerslev (1986) Eq.(5), when $d = 1$, $\delta = 2$ and $\gamma_k = 0$ for all k, the EWMA(1,1) model used in the RiskMetrics data set when $d = 1$, $\delta = 2$, $\gamma_k = 0$ and $\omega = 0$ for all k,

$$\sigma_t^2 = (1-\beta)\varepsilon_{t-1} + \beta\sigma_{t-1}^2 \tag{A.3}$$

5) The APARCH(p,q) model of Ding, Granger and Engle (1993) Eq.(10), when $d = 0$, and the other seven specifications included in it, namely, the ARCH(p) model of Engle (1982) Eq.(2), when $d = 0$, $\delta = 2$, $\gamma_k = 0$ for all k, and $\beta_s = 0$ for all $s = 1,...,q$; the GARCH(p,q) model of Bollerslev (1986) Eq.(3), when $d = 0$, $\delta = 2$ and $\gamma_k = 0$ for all k, the PARCH(p,q) model when $d = 0$ and $\gamma_k = 0$ for all k,

$$\sigma_t^\delta = \omega + \sum_{k=1}^{p} \alpha_k \left|\varepsilon_{t-k}\right|^\delta + \sum_{s=1}^{q} \beta_s \sigma_{t-s}^\delta \tag{A.4}$$

the AGARCH(p,q) model of Glosten et al. (1989), when $d = 0$ and $\delta = 2$,

$$\sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_k(\left|\varepsilon_{t-k}\right| - \gamma_k\varepsilon_{t-k})^2 + \sum_{s=1}^{q} \beta_s \sigma_{t-s}^2 \tag{A.5}$$

the AVGARCH(p,q) model of Taylor and Schwert (1989), when $d = 0$, $\delta = 1$ and $\gamma_k = 0$ for all k,

$$\sigma_t = \omega + \sum_{k=1}^{p} \alpha_k \left|\varepsilon_{t-k}\right| + \sum_{s=1}^{q} \beta_s \sigma_{t-s} \tag{A.6}$$

the TARCH(p) model of Zakoian (1991), when $d = 0$, $\delta = 1$ and $\beta_s = 0$ for all $s = 1, ..., q$,

$$\sigma_t = \omega + \sum_{k=1}^{p} \alpha_k^+ \varepsilon_{t-k}^+ - \sum_{k=1}^{p} \alpha_k^- \varepsilon_{t-k}^- \tag{A.7}$$

where $\alpha_k^+ = \alpha_k(1 - \gamma_k)$, $\gamma_k^- = \alpha_k(1 + \gamma_k)$, $\varepsilon_{t-k}^- = \varepsilon_{t-k} - \varepsilon_{t-k}^+$ and

$$\varepsilon_{t-k}^+ = \left\{ \begin{array}{cc} \varepsilon_{t-k} & \text{if } \; \varepsilon_{t-k} > 0 \\ 0 & \text{otherwise} \end{array} \right\} \tag{A.8}$$

the NARCH(p) model of Higgins and Bera (1992), when $d = 0$, $\gamma_k = 0$ for all k, and $\beta_s = 0$ for all $s = 1, ..., q$,

$$\sigma_t^\delta = \omega + \sum_{k=1}^{p} \alpha_k \left|\varepsilon_{t-k}\right|^\delta \tag{A.9}$$

and, the log-GARCH(p,q) model of Geweke (1986) and Pantula (1986), when $d = 0$ and $\delta \to 0$,

$$\begin{aligned} \log \sigma_t &= \omega^* \log \omega - \sum_{k=1}^{p} \alpha_k \log \sqrt{2/\pi} \\ &\quad + \sum_{k=1}^{p} \alpha_k \log(|\varepsilon_{t-k}| - \gamma_k \varepsilon_{t-k}) + \sum_{s=1}^{q} \beta_s \log \sigma_{t-s}, \end{aligned} \tag{A.10}$$

where $\omega^* = 1 - \sum_{k=1}^{p} \alpha_k - \sum_{s=1}^{q} \beta_s$.