Volatility forecasting: Intra-day versus inter-day models

Angelidis, Timotheos and Degiannakis, Stavros

Department of Economics, University of Crete, Gallos Campus, 74100 Rethymno, Greece, Department of Statistics, Athens University of Economics and Business, 76, Patission Street, 10434

2008

Volatility forecasting: intra-day vs. inter-day models

Timotheos Angelidis[a], Stavros Degiannakis[b,*]

[a]*Department of Economics, University of Crete, Gallos Campus, 74100 Rethymno, Greece*

[b]*Department of Statistics, Athens University of Economics and Business, 76, Patission Street, 10434 Athens, Greece*

Abstract

Volatility prediction is the key variable in forecasting the prices of options, value-at-risk and, in general, the risk that investors face. By estimating not only inter-day volatility models that capture the main characteristics of asset returns, but also intra-day models, we were able to investigate their forecasting performance for three European equity indices. A consistent relation is shown between the examined models and the specific purpose of volatility forecasts. Although researchers cannot apply one model for all forecasting purposes, evidence in favor of models that are based on inter-day datasets when their criteria based on daily frequency, such as value-at-risk and forecasts of option prices, are provided.

JEL Classification: C32 ; C52 ; C53 ; G15

Keywords: Arfimax; Arch; Option pricing; Value-at-risk; Volatility forecasting

---

[*] Corresponding author. *Email address*: Stavros.degiannakis@gmail.com (S. Degiannakis).

1. Introduction

One of the most important issues in finance is the choice of an appropriate volatility model for a researcher to forecast the risk that an investor faces. Since Engle's (1982) seminal paper, many researchers have tried to find the most appropriate risk model that predicts future variability of asset returns by employing various specifications of the autoregressive conditional heteroskedasticity (ARCH) model. However, their results are confusing and conflicting, as there is no model that is deemed as adequate for all financial datasets, sample frequencies and applications, e.g., volatility forecasting, risk management and option pricing.

Volatility can be interpreted as the uncertainty that investors face over their investments. A good starting point to judge competitive models is their out-of-sample forecasting performance, as their predictions are used by portfolio managers to measure and reduce risk. The debate on superior volatility forecasting started with the work of Taylor (1986). Since then, many researchers have tried to find the best performing method for different financial markets and time horizons by twisting around versions of the famous ARCH model, but there is still no agreement in the literature on the most adequate volatility specification. For example, in the work of McMillan et al. (2000) no method was unanimously proposed, because volatility techniques have been examined under different frameworks, such as statistical loss functions, sampling schemes, time periods and assets. However, all suggested methods share a common characteristic: They account for volatility asymmetry.

The availability of high frequency datasets rekindled the interest of academics to forecast risk. The volatility estimates based on intra-day returns are more accurate than those of the daily ones, since the squared daily returns, which have been used as a proxy of the true variance, are an unbiased but noisy estimator of volatility. Recently, Koopman et al. (2005) showed that the ARFIMAX (fractionally integrated auto regressive moving average with exogenous variables) specification for the S&P100 index produced more accurate volatility forecasts than the GARCH and stochastic volatility models. However, they did not examine any flexible ARCH models, which account for the fractional integration of the conditional variance and for the skewed and leptokurtic conditional distribution of asset returns.

For risk management purposes and particularly in the value-at-risk (VaR) arena, most of the empirical works are based on daily returns. Although the issue of VaR has been studied extensively, academics have not yet reached any widely accepted conclusion. On the one hand, Giot and Laurent (2003) proposed the APARCH-skT (asymmetric power ARCH with skewed Student-t distributed innovations) model, while Degiannakis (2004) suggested the FIAPARCH (fractionally integrated APARCH) model and stated that the FIAPARCH with skewed Student-t distributed innovations produces the most accurate VaR predictions for CAC40, DAX30 and FTSE100. On the other hand, many authors (see Angelidis et al., 2004 and references therein) proposed different volatility structures to estimate the daily VaR, but yet again, without arriving at a common conclusion, as they argued that the choice of the best performing model depends on the equity index. Finally, González-Rivera et al. (2004) provided evidence in favor of a stochastic volatility model.

By using high frequency data, researchers explore ways to extract more information to enable them to forecast VaR accurately. Giot and Laurent (2004) compared the APARCH-skT model with an ARFIMAX specification in their attempt to compute the VaR for stock indices and exchange rates. They noted that the use of intra-day dataset did not improve the performance of the inter-day VaR model. Giot (2005) estimated the VaR at intra-day time horizons of fifteen and thirty minutes and argued that the GARCH model with Student-t distributed innovations had the best overall performance, and that there were no significant differences between daily and intra-day VaR models once the intra-day seasonality in the volatility was taken into account.

To summarize, although there are indications that the extended models produce the most accurate VaR forecasts, in some cases, a simpler one is preferred. It was also found that the use of the intra-day datasets does not add to the forecasting power of the models. Therefore, the issue of volatility forecasting for risk management purposes is far from being resolved.

For accurate calculation of the price of an option, the volatility forecast of the underlying asset returns is needed. Noh et al. (1994) assessed the performance of the GARCH and implied volatility regression models by conducting a trading game with straddles written on S&P500. The trading strategy based on the GARCH model yielded a daily return of 0.89%, while a daily loss of 1.26% was incurred by employing the implied volatility method. Engle et al. (1993, 1997) evaluated the forecasts of volatility models by using artificial index option prices that do not face the inherent problems of actual option prices, e.g., market depth, wildcard delivery option and non-synchronous coexistence of option and stock prices. ARCH volatility specifications produced the highest profits. Following their work, Xekalaki and Degiannakis (2005) examined the performance of the SPEC (Standardized Prediction Error Criterion) ARCH model selection algorithm in a simulated options market for the S&P500 index. They concluded that from among a set of ARCH specifications, the asymmetric ARCH models exhibit superior forecasting ability over the symmetric ones. Christoffersen and Jacobs (2004) used data on S&P500 call options and argued that one should not look beyond a simple ARCH model that allows for volatility clustering and leverage effect. Under the same framework, González-Rivera et al. (2004) also concluded that for option pricing, simple models perform as effectively as sophisticated specifications.

Financial literature in option pricing area does not provide evidence that there is added forecast gain from complicated volatility specifications in comparison to simple ARCH models that account for asymmetry and volatility clustering. So far, there has been no work that employs an intra-day volatility model in order to forecast option prices, nor the finding that the simple models are as good as the more complex ones in markets outside the U.S. has been examined.

This paper tries to answer the question: “Is there an adequate intra-day or inter-day model for volatility forecasting, risk management, and prediction of option prices in a dataset comprising three equity indices?" Specifically, it investigates whether a forecast method based on intra-day data is able to produce more accurate one-day-ahead volatility forecasts than one based on the inter-day model, as most research has focused only on one issue at a time.

The key argument of this paper is that the choice of a volatility model is a function of the selection criteria implemented. On the one hand, an intra-day model produces statistically more accurate forecasts than an inter-day one when the realized volatility is under investigation. On the other hand, an intra-day specification does not provide any added value in the forecasting arena of inter-day-based financial applications. The results from two financial applications, VaR and option pricing and in a volatility forecasting exercise, are briefly summarized in the following table that shows the best performing model in each case.

	Realized Volatility Forecasts	VaR 95%	VaR 99%	Option Pricing
CAC40	ARFIMAX	TARCH	TARCH	TARCH/FIAPARCH
DAX30	ARFIMAX	FIAPARCH	TARCH	TARCH/FIAPARCH
FTSE100	ARFIMAX	All Models	TARCH	FIAPARCH

The structure of this paper is as follows: Section 2 describes the intra-day and inter-day models, while Section 3 presents the dataset and the estimation procedure. Section 4 investigates the forecasting ability of the models and Section 5 concludes the paper and provides some ideas for further research.

2. Volatility models

Although, there is no unique model that produces the most accurate volatility forecasts for all financial areas of volatility forecasting and datasets, the general conclusions of the volatility forecasting literature can be summarized in the following lines. In a majority of studies, models that are based on high frequency data achieve the best risk predictions. However, a flexible ARCH inter-day specification that accounts for recent developments in financial modeling (i.e., leptokurtic and asymmetric conditional distribution of returns, fractional integration, and Box-Cox power transformation of the conditional volatility), also appears to produce accurate risk forecasts. Nevertheless, some studies provide evidence in favor of the simple ARCH specifications, which account only for volatility clustering and leverage effect.

We analyzed three methods of volatility estimation to compare our results with the main findings of financial literature, instead of estimating all volatility models that have trivial or crucial differences in specifications: a simple inter-day model (TARCH under the normal distribution), a complex inter-day model (FIAPARCH under the skewed Student-t distribution), and an intra-day model (ARFIMAX under the skewed Student-t distribution). These models are representative of the research that has been done on the subject.

The ARCH framework is usually presented in the following equations:

$$\begin{gathered} y_t = E(y_t \mid I_{t-1}) + \varepsilon_t \\ \varepsilon_t = z_t \sigma_t \\ z_t \overset{i.i.d.}{\sim} f(0,1) \\ \sigma_t = g\left(\{\varepsilon_{t-i}\}; \{\sigma_{t-j}\}; \{\upsilon_{t-1}\} \forall i \geq 1, \forall j \geq 1\right). \end{gathered} \tag{1}$$

The return series from time $t-1$ to t, $y_t = 100\ln(P_t/P_{t-1})$, where P_t, the asset price at time t, is decomposed into two parts: The conditional mean of return at period t, which depends on the information set that is available at time $t-1$, $E(y_t \mid I_{t-1})$, and the innovation process, ε_t. $\{z_t\}$ is a sequence of independently and identically distributed random variables, while $f(.)$ is their probability density function. The conditional standard deviation of innovations, σ_t, is a functional form, $g(.)$, of the past innovations, their conditional standard deviation, and a vector of predetermined variables, υ_t, that are included in the information set I at time t.

The first specification, named AR(k)TARCH(p,q) with normally distributed standardized innovations, represents a simple ARCH model that accounts for the asymmetry response of innovations to volatility:

$$y_t = c_0 + (1 - c(L))^{-1} z_t \sigma_t$$
$$z_t \overset{i.i.d.}{\sim} N(0,1) \tag{2}$$
$$\sigma_t^2 = a_0 + a(L)(z_t\sigma_t)^2 + \gamma(L) d_t (z_t\sigma_t)^2 + b(L)\sigma_t^2,$$

where $c(L) = \sum_{i=1}^{k} c_i L^i$, $a(L) = \sum_{i=1}^{q} a_i L^i$, $b(L) = \sum_{i=1}^{p} b_i L^i$, $\gamma(L) = \sum_{i=1}^{q} \gamma_i L^i$, and $d_t = 1$ if $z_t > 0$ and $d_t = 0$ otherwise. The predictable component of the conditional mean is considered as a k^{th} order autoregressive process to account for the non-synchronous trading effect. Although the standardized innovations are normally distributed, the innovation process, ε_t, has fatter tails than the normal distribution. The TARCH specification, which was introduced by Glosten et al. (1993), allows good news, $(y_{t-i} - E(y_{t-i} \mid I_{t-i-1}) \equiv z_{t-i}\sigma_{t-i} > 0)$, and bad news, $(z_{t-i}\sigma_{t-i} < 0)$, having a different effect on the conditional variance. Hence, the AR(k)TARCH(p,q) model with normally distributed standardized innovations accounts for (i) non-synchronous trading in the stocks making up an index, (ii) volatility clustering, and (iii) asymmetric (symmetric), unconditional (conditional) distribution of returns.

The second model, named AR(k)FIAPARCH(p,q) with skewed Student-t distributed conditional innovations, represents an ARCH model that accounts for recent developments in the area of inter-day volatility modeling:

$$y_t = c_0 + (1 - c(L))^{-1} z_t \sigma_t$$
$$z_t \overset{i.i.d.}{\sim} skT(0,1;v,g)$$
$$\sigma_t^\delta = a_0 + \left(1 - (1 - b(L))^{-1} a(L)(1-L)^{-d}\right)(|z_t\sigma_t| - \gamma z_t\sigma_t)^\delta \tag{3}$$
$$skT(z_t;v,g) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}\left(\frac{2s}{g+g^{-1}}\right)\left(1 + \frac{sz_t + m}{v-2} g^{-d_t}\right)^{-\frac{v+1}{2}},$$

where g is the asymmetry parameter, $v > 2$ denotes the degrees of freedom of the distribution, $\Gamma(.)$ is the gamma function, $d_t = 1$ if $z_t \geq -m/s$, and $d_t = -1$ otherwise, $m = \Gamma((v-1)/2)\sqrt{(v-2)}\left(\Gamma(v/2)\sqrt{\pi}\right)^{-1}\left(g - g^{-1}\right)$ and $s = \sqrt{g^2 + g^{-2} - m^2 - 1}$.

Tse (1998) built the fractional integration form of the APARCH model, while Giot and Laurent (2004) and Degiannakis (2004) applied the APARCH-skT and FIAPARCH-skT specifications, respectively.

Furthermore, δ imposes a Box-Cox asymmetric power transformation in the conditional standard deviation process. The fractional integration parameter d accounts for the response of the conditional variance to past shocks, which decay at a slow hyperbolic rate. Finally, γ captures the asymmetric relation between the conditional variance and the innovations.

The AR(k)FIAPARCH(p,q) model with skewed Student-t distributed standardized innovations accounts for (i) non-synchronous trading, (ii) volatility clustering, (iii) power transformation and fractional integration of the conditional variance, and (iv) asymmetric and leptokurtic conditional and unconditional distribution of returns.

The third model, named AR(k)ARFIMAX(p,q) under the skewed Student-t distribution, represents a long memory specification that accounts for recent developments in the ultra-high frequency financial modeling:

$$y_t = c'_0 + (1 - c'(L))^{-1} z_t \sigma \tilde{h}_{t|t-1} \tag{4.a}$$

$$z_t \overset{i.i.d.}{\sim} skT(0,1; v', g') \tag{4.b}$$

$$\tilde{h}^2_{t|t-1} = \exp\left(\ln h^2_{t|t-1} + 0.5\sigma^2_u\right) \tag{4.c}$$

$$(1 - a'(L))(1 - L)^{d'}\left(\ln h_t^2 - w'_0 - w'_1 y_{t-1} - \gamma' d'_{t-1} y_{t-1}\right) = (1 + b'(L))u_t \tag{4.d}$$

$$u_t \overset{i.i.d.}{\sim} N\left(0, \sigma^2_u\right), \tag{4.e}$$

where $c'(L)=\sum_{i=1}^{k}c'_i L^i$, $a'(L)=\sum_{i=1}^{q}a'_i L^i$, $b'(L)=\sum_{i=1}^{p}b'_i L^i$, $d'_t=1$ when $y_t>0$ and $d'_t=0$ otherwise. The AFRIMAX specification in (4.d)–(4.e) was applied in intra-day volatility datasets by Andersen et al. (2003), and Koopman et al. (2005).

The framework in (4), which was proposed by Giot and Laurent (2004), is estimated in two steps. First, the ARFIMAX presentation in Equations (4.d)-(4.e) is estimated. The leverage effect parameter, γ', reveals whether large past negative returns increase intra-day volatility, h_t, more than past positive outcomes. Since it is assumed that $u_t \sim N\left(0,\sigma_u^2\right)$, the $\exp\left(u_t\right)$ is log-normally distributed and hence, the unbiased one-day-ahead realized volatility is estimated according to (4.c). Equations (4.a)-(4.b) are an ARCH specification with autoregressive conditional mean and skewed Student-t distributed standardized innovations, where the conditional variance $\sigma_{t|t-1}^2=\sigma'^2\tilde{h}_{t|t-1}^2$ is a fraction of the realized volatility.

The AR(k)ARFIMAX(p,q)-skT specification accounts for (i) non-synchronous trading, (ii) fractional integration of the intra-day volatility, (iii) asymmetric relation of intra-day volatility with past negative returns, and (iv) asymmetric and leptokurtic conditional and unconditional distribution of returns.

The realized intra-day volatility at day t is computed as:

$$h_t^2=\frac{\hat{\sigma}_{oc}^2+\hat{\sigma}_{co}^2}{\hat{\sigma}_{oc}^2}\sum_{j=1}^{m-1}\left(100\left(\ln\left(P_{(j+1/m),t}\right)-\ln\left(P_{(j/m),t}\right)\right)\right)^2 , \tag{5}$$

where $P_{(m),t}$ are the asset prices at day t with m observations per day, $\hat{\sigma}_{oc}^2=100^2T^{-1}\sum_{t=1}^{T}\left(\ln\left(P_{(1),t}\right)-\ln\left(P_{(1/m),t}\right)\right)^2$ is the open to close sample variance, and $\hat{\sigma}_{co}^2=100^2T^{-1}\sum_{t=1}^{T}\left(\ln\left(P_{(1/m),t}\right)-\ln\left(P_{(1),t-1}\right)\right)^2$ is the close to open sample variance. To scale the intra-day returns, we followed Koopman et al. (2005) who suggested accounting for overnight returns without inserting the noisy effect of daily returns. To avoid market microstructure frictions without lessening the accuracy of the continuous record asymptotics, we used five-minute linearly interpolated prices.

Following Angelidis et al. (2004) and references therein, we did not select the order of k, p, and q according to a model selection criterion, such as the Akaike Information Criterion (AIC) or the Schwarz Bayesian Criterion (SBC), as a good in-sample performance of a model is not a prerequisite for its good out-of-sample precision[1]. Given that the statistical properties of AIC and SBC selection criteria, at least in the ARCH context, are unknown and in a majority of empirical studies, the use of one lag has been proven to work effectively in forecasting volatility for both ARCH and ARFIMAX frameworks, we chose to set $p = q = k = 1$.

2.1. Forecast schemes

The schemes of computing the one-step-ahead volatility forecasts of the three models are briefly presented in the following paragraphs:

- **AR(1)TARCH(1,1) model**

According to a rolling sample of s trading days and at each point in time t, for $t = s, s+1, ..., T+s-1$, we estimated the parameters $\left(c_0^{(t)}, c_1^{(t)}, v^{(t)}, g^{(t)}, \delta^{(t)}, d^{(t)}, a_0^{(t)}, a_1^{(t)}, b_1^{(t)}, \gamma^{(t)}\right)$ of (2) and then we forecasted the daily conditional variance as:

$$\sigma_{t+1|t}^2 = a_0^{(t)} + \left(a_1^{(t)} + \gamma_1^{(t)} d_t\right)\varepsilon_t^2 + b_1^{(t)}\sigma_t^2. \qquad (6)$$

- **AR(1)FIAPARCH(1,1)-skT**

Similarly, we estimated the parameters $\left(c_0^{(t)}, c_1^{(t)}, v^{(t)}, g^{(t)}, \delta^{(t)}, d^{(t)}, a_0^{(t)}, a_1^{(t)}, b_1^{(t)}, \gamma^{(t)}\right)$ of (3) to forecast the variance as:

---

[1]A representative example of the inability of the in-sample model selection methods to suggest models with superior volatility forecasting ability is given by Degiannakis and Xekalaki (2007). They showed that the commonly used in-sample methods of model selection such as AIC, SBC, and Mean Squared Error (MSE), among others, did not lead to the selection of a model that tracks close future volatility.

$$\sigma_{t+1|t}^{2} = \left(a_0^{(t)} + \left(1 - \left(1 - b_1^{(t)} L\right)^{-1} a_1^{(t)} L (1-L)^{-d^{(t)}}\right)\left(|z_t \sigma_t| - \gamma^{(t)} z_t \sigma_t\right)^{\delta^{(t)}} \right)^{2/\delta^{(t)}}. \quad (7)$$

- **AR(1)ARFIMAX(1,1)-skT**

The estimation of the intra-day model was made by following the next three steps:

- Based on $\{h_k^2\}_{k=t-s+1}^{t}$ for s trading days, we estimated the parameters $\left(d'^{(t)}, w_0'^{(t)}, w_1'^{(t)}, a_1'^{(t)}, b_1'^{(t)}, \gamma'^{(t)}, \sigma_u^{2(t)}\right)'$ of (4.d)–(4.e).
- We estimated $h_{k|k-1}^2$, for $t-s+1 \le k \le t$, and forecasted $\tilde{h}_{t+1|t}^2$ following Equation (4.c).
- Based on $h_{k|k-1}^2$, for $t-s+1 \le k \le t$, and $\{y_k\}_{k=t-s+1}^{t}$ for s trading days, we estimated the parameters $\left(c_0'^{(t)}, c_1'^{(t)}, \sigma'^{(t)}, v'^{(t)}, g'^{(t)}\right)$ of (4.a)–(4.b) and then forecasted the one-day-ahead conditional variance as:

$$\sigma_{t+1|t}^{2} = \sigma'^{2(t)} \tilde{h}_{t+1|t}^{2}. \quad (8)$$

3. Intra-day and inter-day datasets

The intra-day dataset was obtained from Olsen and Associates and comprises three European stock indices: The CAC (from January 3, 1995 to September 8, 2003, totaling 2,177 trading days), the DAX30 (from July 3, 1995 to December 29, 2003, totaling 2,136 trading days), and the FTSE100 (from January 2, 1998 to December 30, 2003, totaling 1,485 trading days) indices[2].

Panel A of Table 1 lists the descriptive statistics of daily log-returns. There are indications of non-zero skewness and excess kurtosis relative to that of the normal distribution, and therefore, each volatility model must consider these characteristics. Under the assumption that the log-returns are

[2] Even if it would be interesting to compare the performance of the three models over the same timeframe, this was not possible as the available intra-day dataset did not cover the same periods. On the other hand, by using different sample periods, we were able to investigate whether the risk management techniques are robust across various time periods and specifically select a model that is not affected by the chosen sample period.

i.i.d. normally distributed, the sample skewness, $\hat{s}$, and kurtosis, $\hat{\kappa}$, are distributed normally with variances $V(\hat{s}) = 6/T$ and $V(\hat{\kappa}) = 24/T$, respectively. Only the skewness parameter of CAC40 index belongs to the 95% confidence interval, and thus there are indications that only the distribution of this index is symmetric.

Panel B of Table 1 lists the descriptive statistics of the annualized volatility ($\sqrt{252h_t^2}$). The most volatile indices are the CAC40 and the DAX30, while the safest market is that of the U.K. Our findings are in line with the previous studies (i.e. Giot and Laurent, 2004) as the risk that investors face is not normally distributed since it exhibits positive skewness and excess kurtosis relative to that of the standard normal distribution. Figures of daily log-returns, intra-day standard deviation and logarithmic variance, as well as tables that present the estimated parameters of the three models are available upon request.

Table 1. Descriptive statistics of the daily log-returns, $y_t = 100(\ln(P_t) - \ln(P_{t-1}))$, and the annualized intra-day standard deviation. CAC40 (January 1995–September 2003), DAX30 (July 1995–November 2003) and FTSE100 (January 1998–December 2003).

	Panel A Daily log-Returns			Panel B Annualized Realized Volatility (Standard Deviation)		
	CAC40	DAX30	FTSE100	CAC40	DAX30	FTSE100
Mean	0.027877	0.029954	-0.00933	24.0149	23.6102	16.993
Median	0.039493	0.108051	-0.00468	19.9144	21.0806	15.2741
Maximum	8.885403	7.449508	5.912383	152.336	120.007	104.619
Minimum	-7.69300	-8.87687	-6.35088	6.61509	3.25114	5.30327
Std. Dev.	1.572413	1.702469	1.317675	15.5795	13.3137	7.32575
Skewness	0.093838	-0.17528	-0.13993	3.71153	1.76413	2.81134
Kurtosis	5.640179	5.183852	4.756223	21.9002	9.20653	21.2393
Jarque-Bera	658.8339	435.6011	195.8197	38775	4538.41	22555.4
Probability	0.00	0.00	0.00	0.00	0.00	0.00

4. Empirical results

The purpose of this section is to evaluate the forecasting ability of the models. Specifically, to achieve this goal, first, we used statistical measures to calculate the distance between the predicted and the realized volatility. Second, in a risk management environment, we examined whether VaR forecasts exhibit conditional coverage. Finally, in a simulated option-pricing framework, we evaluated the models by finding which one generates the highest profits for the investors that used it.

For all models and equity indices, we used a rolling sample of 1,000 observations to generate the out-of-sample forecasts. The initial volatility forecasts are generated for January 18, 1999, July 6, 1999, and January 29, 2002 for the CAC40, DAX30, and FTSE100 indices, respectively[3]. The parameters of the models are estimated using the G@RCH and ARFIMA packages of Ox. Given that the estimated parameters describe the trading behavior, the estimations must incorporate the most recent information. Thus, they are re-estimated each trading day. Figure 1 plots, indicatively, the realized intra-day volatility (Equation 5) for the CAC40 index and the corresponding risk forecasts (Equations 6-8) of the three models.

4.1. *Volatility forecasts*

We measure the accuracy of the models in forecasting the one-day-ahead conditional variance via three loss functions: (i) the MSE, (ii) the Heteroskedasticity-Adjusted Squared Error (HASE), and (iii) the Logarithmic Error (LE). The loss functions are presented in the following equations:

---

[3] To strike a balance between the necessity of having an initial sample that is large enough for the models to be estimated accurately and the total out-of-sample observations to be as many as possible, we chose to work with a rolling sample of 1,000 observations and to create 1,777, 1,136, and 485 forecasts for the CAC40, DAX30, and FTSE100 indices, respectively, without imposing the initial forecasts to be made for the same date.

$$MSE = T^{-1}\sum_{t=1}^{T}\left(h_{t+1}^{2} - \sigma_{t+1|t}^{2}\right)^{2} \tag{9}$$

$$HASE = T^{-1}\sum_{t=1}^{T}\left(1 - h_{t+1}^{2}/\sigma_{t+1|t}^{2}\right)^{2}, \tag{10}$$

$$LE = T^{-1}\sum_{t=1}^{T}\ln\left(h_{t+1}^{2}/\sigma_{t+1|t}^{2}\right)^{2}, \tag{11}$$

where h_{t+1}^{2} is the realized volatility[4] used as the measure of the true, but unobservable, variance at day $t+1$, $\sigma_{t+1|t}^{2}$ is the one-day-ahead variance forecast and T is the number of the forecasts. In the case of the intra-day model, the one-day-ahead conditional variance is estimated according to (4.c), so $\sigma_{t+1|t}^{2} \equiv \tilde{h}_{t+1|t}^{2}$.

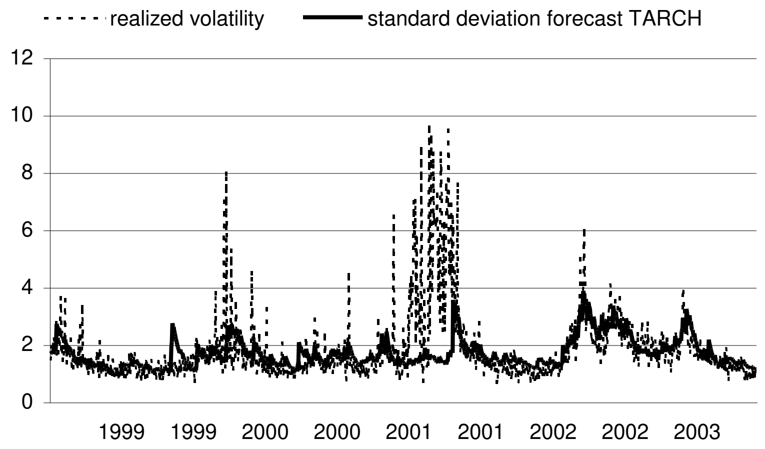
Squared distance between observed and predicted values is the most popular measure in evaluating forecasting accuracy. However, when volatility is the variable under study, symmetric loss functions may produce unreliable results due to the highly non-linear environment. Therefore, we will also evaluate volatility forecasts according to more elaborate loss functions. HASE and LE functions, which take into account the heteroskedastic framework, were introduced by Bollerslev and Ghysels (1996) and Pagan and Schwert (1990), respectively.

Hansen and Lunde (2006) have stated that the substitution of a noisy proxy such as the squared daily returns for the true but unobservable conditional variance can result in an inferior model being chosen as the best one. On the contrary, the realized volatility as a proxy variable does not lead to favor an inferior model. Moreover, they provided evidence that the MSE loss function ensures the equivalence of the ranking of volatility models that is induced by the true volatility and its proxy.

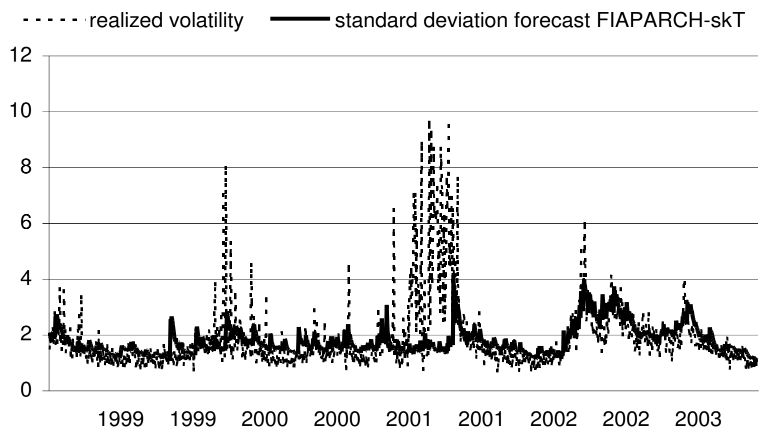
[4] The realized volatility is computed according to Equation (5).

Figure 1. Realized intra-day standard deviation and its forecast for the CAC40 index.

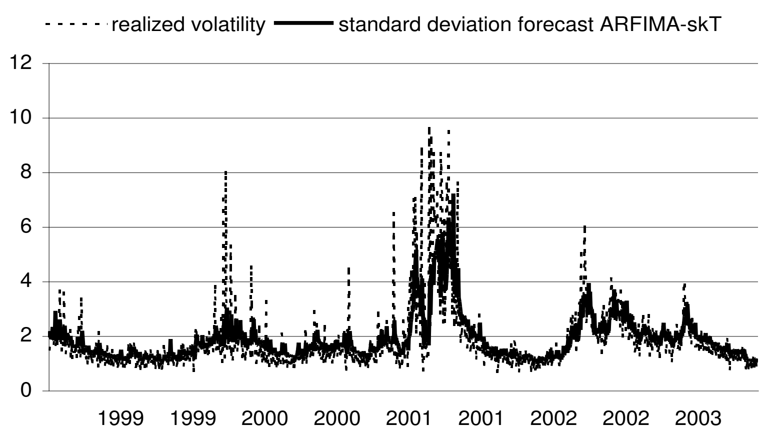
AR(1) TARCH(1,1)



AR(1) FIAPARCH(1,1)-skT



AR(1) ARFIMAX(1,1)-skT



The statistical significance of the volatility forecasts is investigated by (i) the Diedold and Mariano (1995) statistic (DM) and (ii) the Hansen's (2005) Superior Predictive Ability (SPA) hypothesis testing, which are the most frequently used tests in such studies.

Let i be the benchmark model with the lowest loss function value. The DM statistic is the t-statistic derived by the regression of $X_{t,l}^{(i,i^*)} = L_{t,l}^{(i)} - L_{t,l}^{(i^*)}$ on a constant with heteroskedastic and consistent (HAC) standard errors, where $L_{t,l}^{(i)}$ is the value of the loss function l at time t of model i. The null hypothesis, that the benchmark model i has equal predictive ability with model i^*, for $i^* = 1,\ldots,M$, is investigated against the alternative hypothesis that the benchmark model has superior predictive ability. Hansen (2005) introduced the SPA test that is used to compare the forecasting performance of a base model against its M competitors. The null hypothesis that $E\left(X_{t,l}^{(i,1)}, \quad . \quad . \quad . \quad , X_{t,l}^{(i,M)}\right)' \leq 0$ is tested with the statistic $T_l^{SPA} = \max\limits_{i^*=1,\ldots,M} \dfrac{\sqrt{M}\,\overline{X}_{i^*,l}}{\sqrt{Var\left(\sqrt{M}\,\overline{X}_{i^*,l}\right)}}$, where $\overline{X}_{i^*,l} = T^{-1}\sum_{t=1}^{T} X_{t,l}^{(i,i^*)}$. The estimation of $Var\left(\sqrt{M}\,\overline{X}_{i^*,l}\right)$ and the p-value of the T_l^{SPA} are obtained by using the bootstrap method.

According to Table 2, the ARFIMAX is superior to the inter-day models in almost all the cases. The FIAPARCH-skT model, only in the case of FTSE100, has the lowest value of the MSE and HASE loss functions, whereas the hypothesis that it has equal predictive ability with the ARFIMAX model is not rejected. For the HASE loss function, the parsimonious ARCH specification has statistically equivalent forecast ability with the extended ARCH model. For the other two indices, the intra-day model, irrespective of the applied loss functions, generates the most adequate risk forecasts, whereas according to the DM statistic, the hypothesis of equal predictive ability is rejected. Hence, we arrive at the conclusion that the intra-day model clearly produces the most accurate variance forecasts[5]. Results from the SPA test are qualitatively similar to those of the DM test, and are available upon request.

Table 2. The values of the loss functions, the DM statistic for the null hypothesis that a model has equal predictive ability with the benchmark model, and the corresponding p-values.

	MSE	DM Stat.	p-value	HASE	DM Stat.	p-value	LE	DM Stat.	p-value
				CAC40					
AR(1)TARCH(1,1)	86.126	-2.5889	0.0097	16.393	-2.9093	0.0037	0.6256	-3.2619	0.0011
AR(1)ARFIMAX(1,1)	**50.986**	-	-	**1.2390**	-	-	**0.3117**	-	-
AR(1)FIAPARCH(1,1)-skT	85.757	-2.5869	0.0098	15.605	-2.8721	0.0042	0.6429	-3.5677	0.0004
				DAX30					
AR(1)TARCH(1,1)	10.591	-2.1849	0.0291	0.5017	-3.6353	0.0003	0.2036	-3.0492	0.0023
AR(1)ARFIMAX(1,1)	**9.5729**	-	-	**0.2928**	-	-	**0.1701**	-	-
AR(1) FIAPARCH(1,1)-skT	11.567	-2.4931	0.0128	0.5009	-2.4053	0.0163	0.2203	-4.4358	0.0000
				FTSE100					
AR(1)TARCH(1,1)	4.6015	-1.7817	0.0754	0.3064	-0.1340	0.8934	0.3195	-4.1921	0.0000
AR(1)ARFIMAX(1,1)	4.8093	-1.2637	0.2069	0.5357	-1.3698	0.1714	**0.2340**	-	-
AR(1) FIAPARCH(1,1)-skT	**4.2657**	-	-	**0.3038**	-	-	0.2953	-3.1332	0.0018

Bold face fonts present the best performing model.

4.2. Value-at-risk

VaR at a given probability level a, is the predicted amount of financial loss of a portfolio over a given time horizon. Therefore, daily forecasts of the three volatility models are used to estimate the 95% and 99% VaR numbers as:

$$VaR_{t+1|t} = F\left(a;\theta^{(t)}\right)\sigma_{t+1|t}, \tag{12}$$

[5]Other loss functions were also computed with similar results. The loss functions were also computed in the case of forecasting the one-day-ahead standard deviation and the results were qualitatively similar.

where $F\left(a;\theta^{(t)}\right)$ is the corresponding quantile of the assumed distribution, which is computed based on the vector of parameters estimated at time t, and $\sigma_{t+1|t}$ is the next day's conditional standard deviation forecast. Given that the VaR is never observed, not even after the violation, first, we have to calculate the VaR values and then examine the statistical properties of the forecasts.

Christoffersen (1998) developed a joint test to examine the independence hypothesis and the conditional coverage assumption of the VaR violations[6]. The likelihood ratio statistics of these tests are described in the following equations:

$$LR_{uc} = 2\ln\left(\left(1-\frac{N}{T}\right)^{T-N}\left(\frac{N}{T}\right)^{N}\right) - 2\ln\left((1-\rho)^{T-N}\rho^{N}\right) \sim X_1^2, \tag{13.a}$$

$$LR_{in} = 2\left(\ln\left((1-\pi_{01})^{n_{00}}\pi_{01}^{\ n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}\right) - \ln\left((1-\pi_0)^{n_{00}+n_{10}}\pi_0^{\ n_{01}+n_{11}}\right)\right) \sim X_1^2, \tag{13.b}$$

$$LR_{cc} = -2\ln\left((1-\rho)^{\text{T-N}}\rho^{N}\right) + 2\ln\left((1-\pi_{01})^{n_{00}}\pi_{01}^{\ n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}\right) \sim X_2^2, \tag{13.c}$$

where N is the number of days that a violation has occurred over a period T and ρ is the desired coverage rate. n_{ij} is the number of observations with value i followed by j, for $i, j = 0,1$ and $\pi_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$ are the corresponding probabilities. $i, j = 1$ denotes that a violation has occurred, and $i, j = 0$ indicates the opposite. Finally, $\pi_0 = \pi_{01} = \pi_{11}$, under the null hypothesis of independence. Based on Equation (13.a), the hypothesis that the average number of violations is statistically equal to the expected one is tested, whereas Equations (13.b) and (13.c) investigate the assumptions of independence and conditional coverage, respectively. Under this framework, a risk model is rejected if it generates either too many or too few clustered violations.

Table 3 lists the exception rates and the p-values for the three volatility models. There are strong indications that the TARCH model generates the most accurate VaR forecasts at the higher

[6] A violation occurs if the predicted VaR is not able to cover the realized loss.

confidence level since for each index, all the p-values are greater than 10%[7]. The excellent performance at the higher confidence level of the TARCH model using the normal distribution was rather a surprise, since most researchers reported that these volatility techniques under normal distribution usually underestimate total risk (see Angelidis et al., 2004).

Table 3. Exception rates (N/T) and p-values of the backtesting tests (unconditional coverage (LR_{uc}), independence (LR_{in}) and conditional coverage (LR_{cc}) likelihood ratio statistics).

	Model					
	AR(1)TARCH(1,1)		AR(1)ARFIMAX(1,1)-skT		AR(1)FIAPARCH(1,1)-skT	
Confidence Level	95%	99%	95%	99%	95%	99%
Exception rates (N/T)						
CAC40	4.93%	1.36%	2.63%	0.08%	2.72%	0.17%
DAX30	6.30%	1.08%	3.87%	0.36%	3.96%	0.47%
FTSE100	6.19%	1.03%	4.74%	0.41%	5.36%	0.21%
P-values of the Unconditional Coverage Likelihood Ratio Statistic (LR_{uc})						
CAC40	90.93%	24.00%	0.00%	0.00%	0.01%	0.04%
DAX30	5.55%	79.11%	7.25%	1.36%	9.95%	3.88%
FTSE100	24.72%	94.57%	79.29%	14.04%	71.84%	3.25%
P-values of the Independence Likelihood Ratio Statistic (LR_{in})						
CAC40	57.34%	50.65%	19.51%	96.71%	18.09%	93.42%
DAX30	3.99%	60.85%	6.26%	86.49%	52.59%	83.15%
FTSE100	87.42%	74.66%	96.26%	89.75%	57.40%	94.87%
P-values of the Conditional Coverage Likelihood Ratio Statistic (LR_{cc})						
CAC40	84.79%	40.21%	0.01%	0.02%	0.02%	0.19%
DAX30	1.94%	84.68%	3.52%	4.68%	21.06%	11.56%
FTSE100	50.56%	94.69%	96.50%	33.45%	80.01%	10.14%

[7]A high cut-off point is preferred to ensure that the successful risk management techniques will not over or under estimate statistically the *true* VaR, as in the former case, the financial institution does not use its capital efficiently, while in the latter case, it cannot cover future losses.

At the lower confidence level, again the TARCH model generates accurate VaR forecasts in two (CAC40 and FTSE100) out of three indices.

4.3. Predicting option prices

As Engle et al. (1997) noted, a natural criterion to compare any pair of competing volatility methods is the incremental profit from replacing the worse forecast with the better one. Thus, the volatility models are evaluated via an option pricing loss function[8]. A trader with a higher (lower) forecast price for the option buys (sells) a straddle on a $1 share of the underlying index from any of the remaining traders with lower (higher) forecast prices. The straddle trading, which is the purchase (or sale) of both a call and a put option with the same maturity day, is used as its rate of return is indifferent to any change in the underlying asset price and is affected only from changes in volatility. Hence, we simulate an options market comprising three fictitious agents who trade straddles based on their volatility forecasts. According to the Black and Scholes (1973) pricing formula, the expected price of a straddle on a $1 share of the underlying asset at time $t+1$ given the information available at time t with one day to maturity and exercise price equal to the exponent of the risk free rate of return, is given by:

$$S_{t+1|t} = 4N\left(0.5\sigma_{t+1|t}\right) - 2, \tag{14}$$

where $N(.)$ denotes the cumulative normal distribution function. The daily profit of each agent from holding the straddle is:

$$\pi_{t+1} = \max\left(\exp\left(y_{t+1}\right) - \exp\left(r_{t+1}\right), \exp\left(r_{t+1}\right) - \exp\left(y_{t+1}\right)\right), \tag{15}$$

where y_t denotes the daily log-returns of the underlying asset and r_t is the daily risk free rate. A trade between two agents, i and j, is executed at the average of the reservation prices of the two agents, yielding to trader i a profit of:

[8]Other option pricing functions could have been applied. However, in our case of one-day maturity options, we decided to create a pricing loss function based on the Black and Scholes formula that is widely accepted and straightforwardly computed.

$$\pi_{t+1}^{(i,j)} = \begin{cases} \pi_{t+1} - \left(S_{t+1|t}^{(i)} + S_{t+1|t}^{(j)}\right), & \text{if } S_{t+1|t}^{(i)} > S_{t+1|t}^{(j)} \\ \left(S_{t+1|t}^{(i)} + S_{t+1|t}^{(j)}\right) - \pi_{t+1}, & \text{if } S_{t+1|t}^{(i)} < S_{t+1|t}^{(j)} \end{cases}. \quad (16)$$

We create a loss function, that calculates the cumulative returns, and examine whether the forecast method with the highest profit has statistically superior ability. For T trading days and $M = 3$ agents, the i^{th} agent's average daily profit is computed as:

$$\pi^{(i)} = T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{2} \pi_t^{(i,j)}. \quad (17)$$

Based on the Diebold-Mariano method, we test the null hypothesis of equivalent predictive ability of agents-models i and i^*, against the alternative hypothesis that model i is superior to model i^*. For $z_t^{(i,i^*)} = \left(\sum_{j=1}^{2} \pi_t^{(i,j)} - \sum_{j=1}^{2} \pi_t^{(i^*,j)}\right)$, the Diebold-Mariano statistic is the t-statistic derived by the regression of $z_t^{(i,i^*)}$ on a constant with HAC standard errors. A positive value of $z_t^{(i,i^*)}$ indicates that model i is superior to model i^*.

Table 4 lists the daily profit for each agent-model and the corresponding t-statistic of the DM test. The highest return is achieved by the agent who used the TARCH model in the case of the CAC40 index and by the agent who followed the FIAPARCH-skT model in the cases of the DAX30 and FTSE100 indices. The two ARCH models have statistically equal predictive ability, while on the other hand, the ARFIMAX-skT forecast-driven agent achieves statistically lower returns in all the cases. The results for the TARCH model are in line with the work of Christoffersen and Jacobs (2004) who noted that the simple GARCH models must be applied to estimate the option prices[9].

Following Engle et al.'s (1993) approach, we assume various levels of exercise prices to investigate whether our results are sensitive to them. As the results are qualitatively similar among

[9]As Bollerslev and Mikkelsen (1996) argued, the importance of using fractionally integrated variance models stems from the added flexibility in pricing options with maturity of two months or longer. Thus, the added value of ARFIMAX and FIAPARCH specifications may be investigated in a future research with long-term options.

various levels of exercise prices, we indicatively present the results for an exercise price that equals to e^{-3r_t}.

Table 4. Average daily profits for each agent-model, the DM statistic and the corresponding p-values for each agent-model against the best performing agent-model.

Average daily profits for each agent-model						
	Exercise price e^{r_t}			Exercise price e^{-3r_t}		
	CAC40	DAX30	FTSE100	CAC40	DAX30	FTSE100
AR(1)TARCH(1,1)	**0.1018%**	0.0944%	-0.0065%	**0.1020%**	0.0946%	-0.0065%
AR(1)ARFIMAX(1,1)-skT	-0.1410%	-0.2084%	-0.0689%	-0.1410%	-0.2087%	-0.0694%
AR(1)FIAPARCH(1,1)-skT	0.0392%	**0.1141%**	**0.0754%**	0.0390%	**0.1141%**	**0.0758%**
DM Statistic and the corresponding p-values						
AR(1)TARCH(1,1)	-	0.254	1.049	-	0.251	1.053
		(0.399)	(0.147)		(0.400)	(0.146)
AR(1)ARFIMAX(1,1)-skT	2.406	3.411	1.646	2.199	3.407	1.654
	(0.008)	(0.000)	(0.050)	(0.014)	(0.000)	(0.049)
AR(1)FIAPARCH(1,1)-skT	0.748	-	-	0.635	-	-
	(0.227)			(0.262)		

Bold face fonts present the best performing model. P-values of the DM statistic are presented in parentheses.

Engle et al. (1993) added three more agents who trade straddles based on the average, the minimum, and the maximum levels of the daily forecasts. The average of conditional independent forecasts converges rapidly to a perfect forecast, so that any failure of the average forecast indicates a departure from the quality of the individual forecasts. Also, in case of a downward (upward) bias, the maximum (minimum) forecast will beat all individual forecasts that are biased. Profits are then re-computed in the simulated options market, which now comprises six traders. The results show that there are no differences in the performance of the three agents. Moreover, there is no evidence of any bias, as the average forecast takes the first two places and the agents who base their trades on the minimum and the maximum of the daily forecasts achieve the lowest returns. Since there are no differences with the previous findings, we do not present the detailed results, but they are available upon request.

5. Conclusions

The most frequently raised question in the finance literature is which model is to be used to forecast the volatility of asset returns. Given that investors mainly focus on predicting the prices of options, calculating the VaR, and forecasting volatility, the issue of choosing one model for all cases is quite complicated and extremely interesting.

In this paper, we examined whether an intra-day or an inter-day model generates the most accurate forecasts in three European equity markets under the framework of two financial applications, i.e., VaR forecasting and prediction of option prices, plus a volatility forecasting exercise. In the realized volatility forecasting arena, the intra-day model clearly produces the most accurate variance forecasts. When we investigated the performance of the models in a risk management environment, we arrived at a contrary conclusion. In general, for both confidence levels the TARCH model under normal distribution forecasts the VaR accurately. Finally, in the simulated option pricing framework, although the FIAPARCH-skT model generated higher returns, the two ARCH specifications had statistically equivalent predictive powers.

The key argument of this paper is that the choice of a volatility model is a function of the selection criteria implemented and that it is impossible to select one model that would do well according to all criteria. However, it provides guidance on the volatility modeling process, since each financial application (volatility forecasting, risk management and option pricing) reveals the most crucial elements that must be considered. Specifically, for both the VaR and the European option pricing tests, only the price one day later matters, as the option pricing and VaR criteria are based on daily frequency returns. Therefore, the question that this paper tries to answer, under the framework of the three equity indices, can be simplified to: Can the one-day-ahead volatility be better estimated with a model using intra-day data than with a model using daily data? Based on the presented evidence, the answer is clear: Using intra-day data does not help when the criteria are based on daily frequency.

To summarize, the results indicate that there is not a unique model for all cases that can be deemed an adequate one, and therefore investors must be extremely careful when they use one model in all cases. Nevertheless, despite this general conclusion, a researcher must use an inter-day model for inter-day based financial applications and intra-day datasets for intra-day volatility forecasting.

The effects of overnight returns and intra-day noise in the high frequency datasets are still an open area of study. An interesting issue for future research is whether different empirical measures of realized volatility affect the evaluation of volatility specifications' predictability. Finally, an interesting point that can be studied further is the evaluation of the methods that can be used in a multi-period framework as well as for different datasets, e.g., exchange rates or commodities.

Acknowledgement

We would like to thank Peter R. Hansen, Pierre Giot and Sébastien Laurent for the availability of their program codes, Tim Bollerslev and an anonymous referee for their instructive suggestions that helped to improve an early version of the paper.

References

Andersen, T., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 529-626.

Angelidis, T., Benos, A., Degiannakis, S., 2004. The use of Garch models in VaR estimation. Statistical Methodology 1(2), 105-128.

Black, F., Scholes M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81(3), 637-654.

Bollerslev, T., Ghysels, E., 1996. Periodic autoregressive conditional heteroskedasticity. Journal of Business and Economic Statistics 14, 139-157.

Bollerslev, T., Mikkelsen, H.O., 1996. Modeling and pricing long-memory in stock market volatility. Journal of Econometrics 73, 151-184.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841-862.

Christoffersen, P., Jacobs, K., 2004. Which arch model for option valuation? Management Science 50, 1204-1221.

Degiannakis, S., 2004. Volatility forecasting: evidence from a fractional integrated asymmetric power arch skewed-t model. Applied Financial Economics 14, 1333-1342.

Degiannakis, S., Xekalaki, E., 2007. Assessing the performance of a prediction error criterion model selection algorithm in the context of arch models. Applied Financial Economics, 17, 149-171.

Diebold, F.X., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13(3), 253-263.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. Econometrica 50, 987-1008.

Engle, R.F., Hong, C.H., Kane, A., Noh, J., 1993. Arbitrage valuation of variance forecasts with simulated options. Advances in Futures and Options Research 6, 393-415.

Engle, R.F., Kane, A., Noh, J., 1997. Index-option pricing with stochastic volatility and the value of the accurate variance forecasts. Review of Derivatives Research 1, 120-144.

Giot, P., 2005. Market risk models for intraday data. European Journal of Finance 11, 309-324.

Giot, P., Laurent, S., 2003. Value-at-risk for long and short trading positions. Journal of Applied Econometrics 18, 641-664.

Giot, P., Laurent, S., 2004. Modelling daily value-at-risk using realized volatility and arch type models. Journal of Empirical Finance 11, 379-398.

Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801.

González-Rivera, G., Lee, T-H, Mishra, S., 2004. Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. International Journal of Forecasting 20, 629-645.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365-380.

Hansen, P.R., Lunde, A., 2006. Consistent ranking of volatility models. Journal of Econometrics 131, 97-121.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the s&p100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance 12(3), 445-475.

McMillan, D., Speight, A.S., Gwilym, O., 2000. Forecasting u.k. stock market volatility. Journal of Applied Econometrics 10, 435-48.

Noh, J., Engle, R.F., Kane, A., 1994. Forecasting volatility and option prices of the S&P500 index. Journal of Derivatives 2, 17-30.

Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267-290.

Taylor, S.J., 1986. Modeling Financial Time Serie*s*. Wiley and Sons: New York.

Tse, Y.K., 1998. The conditional heteroskedasticity of the yen-dollar exchange rate. Journal of Applied Econometrics 193, 49-55.

Xekalaki, E., Degiannakis, S., 2005. Evaluating volatility forecasts in option pricing in the context of a simulated options market. Computational Statistics and Data Analysis 49(2), 611-629.