# *Backtesting VaR Models: An Expected Shortfall Approach*

Timotheos Angelidis
Department of Economics,
University of Crete,
Gallos Campus,74100 Rethymno, Greece
E-mail address: taggelid@alba.edu.gr.
Corresponding Author.

Stavros Degiannakis
Department of Statistics, Athens University of Economics and Business, 76, Patision str., Athens GR-104 34, Greece.
Tel.: +30-210-8203-120.
E-mail address: sdegia@aueb.gr.

**Abstract**

Academics and practitioners have extensively studied Value-at-Risk (VaR) to propose a unique risk management technique that generates accurate VaR estimations for long and short trading positions and for all types of financial assets. However, they have not succeeded yet as the testing frameworks of the proposals developed, have not been widely accepted. A two-stage backtesting procedure is proposed to select a model that not only forecasts VaR but also predicts the losses beyond VaR. Numerous conditional volatility models that capture the main characteristics of asset returns (asymmetric and leptokurtic unconditional distribution of returns, power transformation and fractional integration of the conditional variance) under four distributional assumptions (normal, GED, Student-t, and skewed Student-t) have been estimated to find the best model for three financial markets, long and short trading positions, and two confidence levels. By following this procedure, the risk manager can significantly reduce the number of competing models that accurately predict both the VaR and the Expected Shortfall (ES) measures.

**Keywords**: Backtesting, Value-at-Risk, Expected Shortfall, Volatility Forecasting, Arch Models.
**JEL**: C22, C52, G15.

## 1. Introduction

The need of major financial institutions to measure their risk started in 1970s after an increase in financial instability. Baumol (1963) first attempted to estimate the risk that financial institutions faced. He proposed a measure based on standard deviation adjusted to a confidence level parameter that reflects the user's attitude to risk. However, this measure is not different from the widely known Value-at-Risk (VaR), which refers to a portfolio's worst outcome that is likely to occur at a given confidence level. According to the Basle Committee, the VaR methodology can be used by financial institutions to calculate capital charges in respect of their financial risk.

Since JP Morgan made available its RiskMetrics system on the Internet in 1994, the popularity of VaR and with it the debate among researchers about the validity of the underlying statistical assumptions increased. This is because VaR is essentially a point estimate of the tails of the empirical distribution. The free accessibility of the RiskMetrics and the plethora of available datasets triggered academics and practitioners to find the best-performing risk management technique. However, even now, the results are conflicting and confusing.

Giot and Laurent (2003a) calculated the VaR number for long and short equity trading positions and proposed the APARCH[i] model with skewed Student-t conditionally distributed innovations (APARCH-skT) as it had the best overall performance in terms of the proportion of failure test. In a similar study, Giot and Laurent (2003b) suggested the same model to the risk managers to estimate the VaR number for six commodities, even if a simpler model (ARCH-skT) generated accurate VaR forecasts. Huang and Lin (2004) argued that for the Taiwan Stock Index Futures, the APARCH model under the normal (Student-t) distribution must be used by risk managers to calculate the VaR number at the lower (higher) confidence level.

Although the APARCH model comprises several volatility specifications, its superiority has not been proved by all researchers. Angelidis and Degiannakis (2005) opined that "a risk manager must employ different volatility techniques in order to forecast accurately the VaR for long and short trading positions", whereas Angelidis et al. (2004) considered that "the Arch structure that produces the most accurate VaR forecasts is different for every portfolio". Furthermore, Guermat and Harris (2002) applied an exponentially weighted likelihood model in three equity portfolios (US, UK, and Japan) and proved its superiority to the GARCH model under the normal and the Student-t distributions in terms of two backtesting measures (unconditional and conditional coverage). Moreover, Degiannakis (2004) studied the forecasting performance of various risk models to estimate the one-day-ahead realized volatility and the daily VaR. He proposed the fractional integrated APARCH model with skewed Student-t conditionally distributed innovations

(FIAPARCH-skT) that efficiently captures the main characteristics of the empirical distribution. Focusing only on the VaR forecasts, So and Yu (2006) argued, on the other hand, that it was more important to model the fat tailed underlying distribution than the fractional integration of the volatility process. The two papers, one by Degiannakis (2004) and the other by So and Yu (2006), among many others, highlight that different volatility techniques are applied for different purposes.

Contrary to the contention of the previous authors, including Mittnik and Paolella (2000), that the most flexible models generate the most accurate VaR forecasts, Brooks and Persand (2003) pointed out that the simplest ones, such as the historical average of the variance or the autoregressive volatility model, achieve an appropriate out-of-sample coverage rate. Similarly, Bams et al. (2005) argued that complex (simple) tail models often lead to overestimation (underestimation) of the VaR.

VaR, however, has been criticized on two grounds. On the one hand, Taleb (1997) and Hoppe (1999) argued that the underlying statistical assumptions are violated because they could not capture many features of the financial markets (e.g. intelligent agents). Under the same framework, many researchers (see for example Beder, 1995 and Angelidis et al., 2004) showed that different risk management techniques produced different VaR forecasts and therefore, these risk estimates might be imprecise. Last, but not least, the standard VaR measure presumes that asset returns are normally distributed, whereas it is widely documented that they really exhibit non-zero skewness and excess kurtosis and, hence, the VaR measure either underestimates or overestimates the true risk.

On the other hand, even if VaR is useful for financial institutions to understand the risk they face, it is now widely believed that VaR is not the best risk measure. Artzner et al. (1997, 1999) showed that it was not necessarily sub-additive, i.e., the VaR of a portfolio may be greater than the sum of individual VaRs and therefore, managing risk by using it may fail to automatically stimulate diversification. Moreover, it does not indicate the size of the potential loss, given that this loss exceeds the VaR. To remedy these shortcomings, Delbaen (2002) and Artzner et al. (1997) introduced the Expected Shortfall (ES) risk measure, which equals the expected value of the loss, given that a VaR violation occurred. Furthermore, Basak and Shapiro (2001) suggested an alternative risk management procedure, namely limited expected losses based risk management (LEL-RM), that focuses on the expected loss also when (and if) losses occur. They substantiated that the proposed procedure generates losses lower than what VaR-based risk management techniques generate.

ES is the most attractive coherent risk[ii] measure and has been studied by many authors (see Acerbi et al. 2001; Acerbi, 2002; and Inui and Kijima, 2005). Yamai and Yoshiba (2005) compared the two measures—VaR and ES—and argued that VaR is not reliable during market turmoil as it can mislead rational investors, whereas ES can be a better choice overall. However, they pointed out that gains on efficient management by using the ES measure are substantial whenever its estimation is accurate. In other cases, they advise the market practitioners to combine the two measures for best results.

Our study sheds light on the issue of volatility forecasting under risk management environment and on the evaluation procedure of various risk models. It compares the performances of the most well known risk management techniques for different markets (stock exchanges, commodities, and exchange rates) and trading positions. Specifically, it estimates the VaR and the ES by using 11 ARCH volatility specifications under four distributional assumptions, namely normal, Student-t, skewed Student-t, and generalized error distribution. We investigated 44 models following a two-stage backtesting procedure to assess the forecasting power of each volatility technique and to select one model for each financial market. In the first stage, to test the statistical accuracy of the models in the VaR context, we examined whether the average number of violations is statistically equal to the expected one and whether these violations are independently distributed. In the second stage, we employed standard forecast evaluation methods by comparing the returns of a portfolio, whenever a violation occurs with the ES forecast.

The results of this paper are important for many reasons. VaR summarizes the risk exposure of the investor in just one number, and therefore portfolio managers can interpret it quite easily[iii]. Yet, it is not the most attractive risk measure. On the other hand, ES is a coherent risk measure and hence its utility in evaluating the risk models can be rewarding. Currently, however, most researchers judge the models only by calculating the average number of violations. Moreover, even if the risk managers hold both long and short trading positions to hedge their portfolios, most of the research has been applied only on long positions. Therefore, it is possible to investigate if a model can capture the characteristics of both tails simultaneously.

This study, to best of our knowledge, is the first that estimates the VaR and ES numbers[iv] for three different markets simultaneously and therefore, we can infer if these markets share common features in risk management framework. Therefore, we combined the most well-known and concurrent parametric models with four distributional assumptions to find out which model has the best overall performance. Even though we did not include all ARCH specifications available in the literature, we estimated the models that captured the most important characteristics of the financial time series and those that were already used or were extensions of specifications that were implemented in similar studies. Finally, we employed a two-stage procedure to investigate the forecasting power of each volatility technique and to guide on VaR model selection process. Following this procedure, we could select a risk model that predicts the VaR number accurately and minimizes, if a VaR violation occurs, the difference between the realized and the expected losses In contrast to this, earlier research focused mainly on the unconditional coverage of the models.

To summarize, this study juxtaposes the performance of the most well-known parametric techniques, and shows that under the proposed backtesting procedure, for each financial market, there is a small set of models that accurately estimate the VaR number for both long and short trading positions and two confidence levels. Moreover, contrary to the findings of the previous research, the more flexible models do not necessarily generate the most accurate risk forecasts, as a simpler specification can be selected regarding two dimensions: (a) distributional assumption and (b) volatility specification. For distributional assumption, standard normal or GED is the most appropriate choice depending on the financial asset, trading position, and confidence level. Besides the distributional choice, asymmetric volatility specifications perform better than symmetric ones, and in most cases, fractional integrated parameterization of volatility process is necessary.

The rest of the paper is organized as follows: Section 2 describes the ARCH models and presents the calculation of VaR and ES, whereas section 3 describes the evaluation framework of VaR and ES forecasts. Section 4 presents preliminary statistics for the dataset, explains the estimation procedure, and presents the results of the empirical investigation. Section 5 presents the conclusions.

## 2. ARCH Volatility Models

To fix notation, let $\{y_t\}_{t=0}^{T} = \{\ln(p_t/p_{t-1})\}_{t=0}^{T}$ refer to the continuously compounded return series, where $p_t$ is the closing price at trading day $t$. The return series follows the stochastic process:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= c_0(1-c_1) + c_1 y_{t-1} \\ \varepsilon_t &= z_t \sigma_t \\ z_t &\overset{i.i.d.}{\sim} f[E(z_t)=0, V(z_t)=1; \theta] \\ \sigma_t^2 &= g(I_{t-1}), \end{aligned} \tag{1}$$

where $E(y_t \mid I_{t-1}) \equiv \mu_t(\theta)$ denotes the conditional mean, given the information set available at time $t-1$, $I_{t-1}$, $\{\varepsilon_t\}_{t=0}^{T}$ is the innovation process with unconditional variance $V(\varepsilon_t) = \sigma^2$ and conditional variance $V(\varepsilon_t \mid I_{t-1}) \equiv \sigma_t^2(\theta)$, $f(.)$ is the density function of $\{z_t\}_{t=0}^{T}$, $g(.)$ is any of the functional forms presented in Table 1 and $\theta$ is the vector of the unknown parameters.

[Insert Table 1 about here]

We take into consideration the following conditional volatility specifications: GARCH$(p,q)$ of Bollerslev (1986), EGARCH$(p,q)$ of Nelson (1991), TARCH$(p,q)$ of Glosten et al. (1993), APARCH$(p,q)$ of Ding et al. (1993), IGARCH$(p,q)$ of Engle and Bollerslev (1986), FIGARCH$(p,q)$ of Baillie et al. (1996), FIGARCHC$(p,q)$ of Chung (1999), FIEGARCH$(p,q)$ of

Bollerslev and Mikkelsen (1996), FIAPARCH$(p,q)$ of Tse (1998), FIAPARCHC$(p,q)$ of Chung (1999), and HYGARCH$(p,q)$ of Davidson (2004). To summarize, the selected volatility models include, besides others, the simplest GARCH model as also the most complex ones, such as FIAPARCHC and HYGARCH. All the selected models reflect the most recent developments in financial forecasting.

Similarly, the chosen density functions of $\{z_t\}_{t=0}^{T}$ are widely applied in finance. In seminal Engle's (1982) paper, the density function was assumed the standard normal, which is described as:

$$f(z_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_t^2}{2}}. \tag{13}$$

However, as the empirical distribution of financial assets is fat-tailed, Bollerslev (1987) introduced the Student-t distribution:

$$f(z_t;v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}\left(1+\frac{z_t^2}{v-2}\right)^{-\frac{v+1}{2}}, \tag{14}$$

where $\Gamma(.)$ is the gamma function. As $v$ tends to infinity, the Student-t tends to the normal distribution. As Student-t is not the only fat tailed distribution available, we also considered the generalized error distribution (GED), which is more flexible than the Student-t as it can include both fat and thick tailed distributions. It was introduced by Subbotin (1923) and applied in ARCH framework by Nelson (1991). Its density function is given in the following equation:

$$f(z_t;v,\lambda) = \frac{v\exp\left(-0.5|z_t/\lambda|^v\right)}{\lambda 2^{\left(1+1/v\right)}\Gamma(v^{-1})}, \tag{15}$$

where $\lambda \equiv \sqrt{2^{-2/v}\Gamma(v^{-1})\Gamma(3v^{-1})}$ and $v > 0$ are the tail-thickness parameters (i.e. for $v = 2$, $z_t$ is standard normally distributed and for $v < 2$, the distribution of $z_t$ has thicker tails than the normal distribution). Finally, given that in VaR framework both the long and short trading positions are important, the skewed Student-t distribution[v] is also applied:

$$f(z_t;v,g) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}\left(\frac{2s}{g+g^{-1}}\right)\left(1+\frac{sz_t+m}{v-2}g^{-d_t}\right)^{-\frac{v+1}{2}}, \tag{16}$$

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, $s = \sqrt{g^2 + g^{-2} - m^2 - 1}$ and $m = \Gamma((v-1)/2)\sqrt{(v-2)}\left(\Gamma(v/2)\sqrt{\pi}\right)^{-1}(g - g^{-1})$ are the standard deviation and the mean, respectively.

Having estimated the vector of the unknown parameters, it is straightforward to calculate VaR using the following equation:

$$VaR_{t+1|t} = \mu_{t+1|t} + F\left(a;\theta^{(t)}\right)\sigma_{t+1|t}\ , \tag{17}$$

where $\mu_{t+1|t}$ and $\sigma_{t+1|t}$ are the conditional forecasts of the mean and the standard deviation at time $t+1$, given the information at time $t$, and $F\left(a;\theta^{(t)}\right)$ is the $a^{th}$ quantile of the assumed distribution, which is computed based on the vector of parameters estimated at time $t$. The functional forms of the one-step-ahead conditional variance predictions, $\sigma^2_{t+1|t}$, are presented in Table 2.

[Insert Table 2 about here]

As we have already mentioned, ES is defined as the conditional expected loss, given a VaR violation. Specifically, for long trading positions, it is calculated as

$$ES_{t+1|t} = E\left(y_{t+1} \mid \left(y_{t+1} \leq VaR_{t+1|t}\right)\right). \tag{29}$$

In particular, ES is a probability-weighted average of tail loss and therefore, to calculate it, we follow Dowd (2002) who suggested that for any distributional assumption "slice the tail into a large number $\kappa$ of slices, each of which has the same probability mass, estimate the VaR associated with each slice and take the ES as the average of these VaRs". To implement this approach, we set $\kappa = 5000$ to increase the accuracy.

## 3. Evaluate VaR and ES Forecasts

Having presented various risk management techniques, we now discuss their formal statistical evaluation. Given that VaR is never observed, not even after violation, we have to first calculate the VaR values and then rank the risk models by examining the statistical properties of the forecasts. Specifically, in the first stage, a model is deemed adequate only if it has not been rejected by both the unconditional and the independence hypotheses. The first hypothesis examines if the average number of violations is statistically equal to the excepted one and the second hypothesis if these violations are independent. However, risk managers who use these tests cannot rank the *adequate* models, because a model with greater *p*-value need not be superior to its competitors and hence, cannot be the best-performing model.

We extended the forecast evaluation approach of Lopez (1999) and Sarma et al. (2003) as the ES measure was introduced in the second stage by creating a loss function that calculated the difference between the actual and the expected losses when a violation occurred. For all the best-performing models of the first stage, we implemented Hansen's (2005) superior predictive ability (SPA) test to evaluate their differences statistically. As Yamai and Yoshiba (2005) pointed out, the two risk measures must be combined to take the most of them and hence, under the proposed backtesting framework, the selected models not only calculate the VaR number accurately but also minimize the difference between the actual loss and the ES.

### 3.1. First Stage Evaluation

The most widely used test, developed by Kupiec (1995), examines whether the observed exception rate is statistically equal to the expected one. Under the null hypothesis that the model is adequate, the appropriate likelihood ratio statistic is:

$$LR_{uc} = 2\ln\left(\left(1-\frac{N}{\tilde{T}}\right)^{\tilde{T}-N}\left(\frac{N}{\tilde{T}}\right)^{N}\right) - 2\ln\left((1-\rho)^{\tilde{T}-N}\rho^{N}\right) \sim X_1^2, \tag{30}$$

where $N$ is the number of days over a period $\tilde{T}$ that a violation occurred and $\rho$ is the desired coverage rate. Therefore, the risk model is rejected if it generates too many or too few violations, but based on it, the risk manager can accept a model that generates dependent exceptions.

Christofersen (1998) proposed a more elaborate criterion, which simultaneously examines if (i) the total number of failures is equal to the excepted one and (ii) the VaR failure process is independently distributed. The appropriate likelihood ratio test of the first hypothesis is given by equation (30) and that of the second one by the following equation:

$$LR_{in} = 2\left(\ln\left((1-\pi_{01})^{n_{00}}\pi_{01}^{\;n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}\right) - \ln\left((1-\pi_0)^{n_{00}+n_{10}}\pi_0^{\;n_{01}+n_{11}}\right)\right) \sim X_1^2, \tag{31}$$

where $n_{ij}$ 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 been made, whereas $i, j = 0$ indicates the opposite, which implies that the process of VaR failures must be spread over the entire sample[vi]. The main advantage of using these two tests is that the risk managers can reject a VaR model that generates too few or too many *clustered* violations and thereby identify the reason for its failure. However, they cannot rank the models based only on the *p*-values of these tests.

### 3.2. Second Stage Evaluation

The statistical adequacy of the VaR forecasts is obtained by previous backtesting tests: the unconditional coverage (equation 30) and the independence test (equation 31). If a model is not rejected, it forecasts VaR accurately. However, in most cases, more than one model is deemed adequate and hence, the risk manager cannot select a unique risk management technique.

To overcome this shortcoming of the backtesting measures, Lopez (1999) proposed a forecast evaluation framework based on loss function. The loss function enables the researcher to rank the models and specify a utility function that accommodates the specific concerns of the risk manager. Specifically, he suggested the following loss function:

$$\Psi_{t+1} = \begin{cases} 1 + \left(VaR_{t+1|t} - y_{t+1}\right)^2 & \text{if violation occurs} \\ 0 & \text{else,} \end{cases} \quad (32)$$

which accounts for the magnitude of the tail losses $\left(\left(VaR_{t+1|t} - y_{t+1}\right)^2\right)$ and adds a score of one whenever a violation is observed. The model that minimizes the total loss $\sum_{t=1}^{T} \Psi_t$, is preferred to other models.

Nevertheless, his approach has two drawbacks. First, if the risk management techniques are not filtered by the aforementioned unconditional or conditional coverage procedures, a model that does not generate any violation is deemed the most adequate as $\Psi_{t+1} = 0$. Second, the return, $y_{t+1}$, should be better compared with the ES measure and not with the VaR, as VaR does not give any indication about the size of the expected loss, given a VaR violation. Therefore, with these limitations, the proposed loss functions can be described by the following equations:

$$\Psi_{1,t+1}^{(i)} = \begin{cases} \left| y_{t+1} - ES_{t+1|t}^{(i)} \right| & \text{if violation occurs} \\ 0 & \text{else,} \end{cases} \quad (33)$$

and

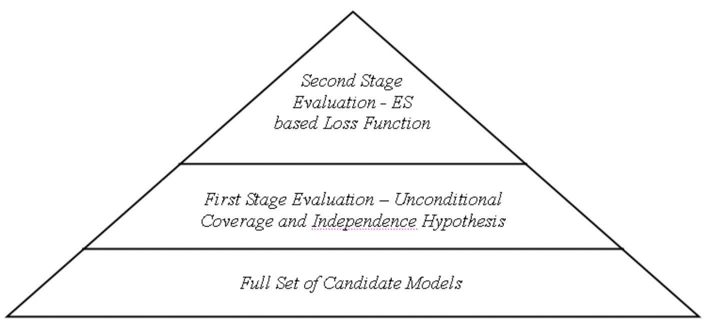
$$\Psi_{2,t+1}^{(i)} = \begin{cases} \left( y_{t+1} - ES_{t+1|t}^{(i)} \right)^2 & \text{if violation occurs} \\ 0 & \text{else.} \end{cases} \quad (34)$$

To judge the models in the second stage, we computed for each model $i$ the mean absolute error (MAE), $\tilde{T}^{-1} \sum_{t=1}^{\tilde{T}} \Psi_{1,t}^{(i)}$, and the mean squared error (MSE), $\tilde{T}^{-1} \sum_{t=1}^{\tilde{T}} \Psi_{2,t}^{(i)}$.

According to the two-stage backtesting procedure, the best performing model will (i) calculate the VaR number accurately, as it will satisfy the prerequisite of correct unconditional and conditional coverage and (ii) forecast the expected loss, given a VaR violation, as it minimizes the total loss value, $\sum_{t=1}^{\tilde{T}} \Psi_{l,t}^{(i)}$.

The statistical significance of the volatility forecasts is investigated by testing Hansen's (2005) Superior Predictive Ability (SPA) hypothesis. For $X_{l,t}^{(i^*,i)} = \Psi_{l,t}^{(i^*)} - \Psi_{l,t}^{(i)}$, the null hypothesis, that the benchmark model $i^*$ is not outperformed by competing models $i$, for $i = 1,\ldots,M$, is investigated against the alternative hypothesis that the benchmark model is inferior to one or more of the competing models. The null hypothesis, $E\left(X_{l,t}^{(i^*,1)} \quad . \quad . \quad . \quad X_{l,t}^{(i^*,M)}\right)' \leq 0$, is tested with the statistic $T_l^{SPA} = \max_{i=1,\ldots,M} \frac{\sqrt{M}\,\overline{X}_{l,i}}{\sqrt{Var\left(\sqrt{M}\,\overline{X}_{l,i}\right)}}$, where $\overline{X}_{l,i} = \widetilde{T}^{-1}\sum_{t=1}^{\widetilde{T}} X_{l,t}^{(i^*,i)}$. The estimation of $Var\left(\sqrt{M}\,\overline{X}_{l,i}\right)$ and $p$-values of the $T_l^{SPA}$ statistic are obtained by using the stationary bootstrap of Politis and Romano (1994).

Under the proposed backtesting environment, the risk manager achieves three goals: forecasts VaR accurately and thus satisfies the prerequisites of the Basel Committee for Banking Supervision; selects one model or a family of models among various candidates following a statistical inference procedure; and finally knows in advance the amount that may be needed if a VaR violation occurs, and therefore is better prepared to face the future losses by forecasting the ES measure accurately. The next figure briefly demonstrates the two-stage backtesting procedure. In the first stage, the investor can work with fewer than the available models by applying the two tests (equations 30 and 31). In the next stage, according to the developed loss functions (equations 33 and 34), the ES measure is used to evaluate statistically the best-performing models.




## 4. Empirical Analysis

To evaluate all the available volatility models, we generated out-of-sample VaR and ES forecasts for S&P500 equity index, Gold Bullion $ per Troy Ounce commodity and US dollar/British pound exchange rate, obtained from Datastream for the period April 4$^{th}$ 1988 to April 5$^{th}$ 2005. The daily prices, the log-returns, and the autocorrelations for the absolute log-returns are presented in Figure 1. Volatility clustering is clearly visible and suggests the presence of heteroskedasticity. The absolute log-returns are significantly positive serial autocorrelated over long lags, whereas the sample autocorrelations decrease too fast at the first lags; at higher lags however, the decrease becomes slower, indicating the long-memory property of volatility process and the utility of fractionally integrated volatility specifications.

[Insert Figure 1 about here]

A question that naturally arises is the order of $p$ and $q$ of the conditional volatility specifications. We choose to set $p = q = 1$, given that in the majority of empirical volatility forecasting studies, the order of one lag has proven to work effectively. So and Yu (2006) concluded that "the best fitted model according to AIC (Akaike, 1973) and SBC (Schwarz, 1978) criteria does not necessarily lead to better VaR estimates", whereas Degiannakis and Xekalaki (2006) demonstrated that in the volatility forecasting arena, the best-performing model could not be selected according to any in-sample model selection criterion.

Based on a $\breve{T} = 3000$ rolling sample, we generated $\tilde{T} = 1435$[vii] out-of-sample forecasts (the parameters are re-estimated each trading day) to calculate the 95% and 99% VaR and ES values for long and short trading positions. The parameters of the models were estimated using the G@RCH (Laurent and Peters, 2002) package of Ox (Doornik, 2001).

[Insert Table 3 about here]

[Insert Table 4 about here]

[Insert Table 5 about here]

The MAE and MSE values (equations 33 and 34), the average values of the VaR and ES measures, the exception rates, and the *p*-values of the two backtesting measures are presented in Tables 3 to 5 for all the models that survived the first evaluation (equations 30 and 31)[viii].

Irrespective of the volatility models and the financial assets, ARCH specifications under the Student-t distribution and its corresponding skewed version overestimate VaR numbers at both confidence levels. A similar observation was made in several earlier studies (see Guermat and Harris, 2002 and Billio and Pelizzon, 2000 among others). Even at a 99% confidence level, they did not show any major improvement, as the average realized exception rates were significantly lower than the expected ones. The introduction of the asymmetry parameter ($\xi$) in the underlying distribution did not make any significant difference. In most cases, the VaR numbers were overestimated, mainly because $\log(\xi)$ was close to zero and therefore, the two distributions in the VaR context, were similar[ix].

ES is at least 0.25% greater than VaR, which implies that the risk manager must adjust accordingly the capital that holds to face the unforeseen losses. Moreover, this adjustment should be mainly implemented for equity and commodity assets, as for these assets ES is almost 0.7% greater than VaR.

For each financial asset, there appears to be a different model that forecasts the VaR number accurately. For example, So and Yu (2006) favored using different models for stock indexes and exchange rates; for stock indexes, they favored an asymmetric specification and for exchange rates, a symmetric function was preferred.

Specifically, for the S&P500 index, five models (FIEGARCH-N, EGARCH-N, APARCH-N, TARCH-N, and FIGARCH-GED) generate adequate VaR forecasts, as the *p*-values of the backtesting measures are greater than 10% for both confidence levels and both trading positions. Even if the more complex models generate, in some cases, the most accurate VaR forecasts (i.e. FIEGARCH-GED for 95% confidence level and long trading position), they do not give the best overall performance. This finding is in line with that of Brooks and Persand (2003) but not with the argument of Mittnik and Paolella (2000) that more general ARCH structures are needed. Highlighting this conclusion is the observation that the IGARCH-GED model generates exception rates that are close to the expected ones only for the short trading positions, whereas it is rejected for the long trading positions, because either the model generates clustered violations or the model misestimates the *true* VaR number. As far as the underlying distribution is concerned, there are indications that standard normal is the best overall choice, as four out of five models are normally distributed.

The GED and normal distribution are the best overall choices for Gold. Between the two, GED is considered more appropriate for the commodity market. For example, if the risk manager is interested only in the higher confidence level and for short trading positions, he/she should use the GED distribution. Any other model would generate inaccurate risk forecasts. To summarize, five models (GARCH-GED, IGARCH-GED, FIAGARCH-GED, FIAGARCHC-GED, and FIAPARCHC-GED) generated accurate predictions for both confidence levels and both trading positions. The risk manager can select any of these models, irrespective of the trading position, and satisfy the requirements of the Basel Committee.

For $/£ exchange rate, the choice of the most appropriate distribution is not straightforward, even if the Student-t and skewed Student-t distributions are rejected. For long (short) trading position and at 99% confidence level, the best overall distribution is the GED (normal), whereas for the other two cases, the results are mixed. EGARCH under the normal distribution appears to have the best overall performance, as only this model generates adequate VaR forecasts for long and short trading positions and for both confidence levels. At the lower confidence level and for long (short) trading position, the exception rate of the model equals 4.67% (4.25%), whereas the corresponding rates at the higher confidence level are 1.39% (0.91%). Furthermore, according to the two loss functions, the EGARCH under the normal distribution model is always ranked first except for higher confidence level and long trading position. Therefore, it is plausible to consider this model, which forecasts the VaR number accurately for trading positions and confidence levels, the most appropriate specification.

The difference among the VaR models cannot be evaluated statistically as neither the greatest *p*-value of the backtesting criteria nor the lowest value of the loss functions indicates the superiority of a model. Therefore, to evaluate the reported differences statistically, we implemented the SPA test taking the following as benchmark models: FIEGARCH-N, EGARCH-N, APARCH-N, TARCH-N, and FIGARCH-GED for S&P500, GARCH-GED, IGARCH-GED, FIGARCH-GED, FIGARCHC-GED, and FIAPARCHC-GED for Gold and EGARCH-N for US dollar to British pound. These models predicted the VaR number accurately for all cases (long and short trading positions, and at 95% and 99% confidence levels).

[Insert Table 6 about here]

Table 6 presents the *p*-values of the SPA test for the null hypothesis that the benchmark model $i^*$ outperforms all the competing models. For example, in the case of S&P500 index and for all cases, the benchmark model (FIEGARCH-N) has superior forecasting ability, as the *p*-value of the test is always greater than 10%. All other benchmark models, at least in one case, have equal predictive power and therefore, there are indications that among the various candidate techniques only one survived the proposed evaluation framework. In the case of Gold, the GARCH-GED and the IGARCH-GED models are statistically superior to their competitors, whereas at least for 95% confidence level and short trading position, FIGARCH-GED, FIGARCHC-GED, and FIAPARCHC-GED models do not generate significantly better forecasts. Finally, for the US $ to UK £ exchange rate, the forecasting ability of EGARCH-N model is superior to those of other models. Also, it is to be noted that the evaluation of the models is robust to the choice of the used loss function, because irrespective of the measurement method, we select the same models as the most appropriates[x].

According to the two-stage backtesting procedure, the risk manager has two choices: (a) to select one model for each trading position and each confidence level from those models that have not been rejected by the backtesting measures and (b) to use the model that forecasts the VaR number accurately for both trading positions and both confidence levels. Naturally, the second choice is better, because it reduces the complexity and computational costs. Consequently, the researcher focuses only on one model for each financial asset. Moreover, by employing the two-stage backtesting procedure, the researcher evaluates statistically the differences between the models, and selects, in most cases, only one volatility specification.

In summary, only some models can forecast the VaR number accurately in all cases. Specifically, in the case of S&P500 index, the FIEGARCH-N generates adequate forecasts for both confidence levels and both trading positions, whereas in the case of Gold, two models (GARCH-GED and IGARCH-GED) give the best overall performance. Lastly, for the US $ to UK £ exchange rate, EGARCH-N is considered the best specification.

## 5. Conclusions

We examined the performance of the most recently developed risk management techniques utilizing a proposed combined backtesting procedure. Specifically, for S&P500 equity index, Gold commodity and US $ to UK £ exchange rate, we computed the VaR and ES measures for two confidence levels (95% and 99%) and for two (long and short) trading positions. We investigated whether the models forecast accurately the expected number of violations, generate independent violations, and predict the ES number. As Hansen (2005) rightly suggested, a filtering procedure must be accounted for the full data exploration, before a legitimate statement of the statistical differences among the candidate models. The reduction of the under consideration models was achieved because the evaluation was made in two stages. In the first stage, the framework developed by Kupiec (1995) and Christofersen (1998) was implemented and in the second, the SPA hypothesis testing was applied.

Different volatility models achieve accurate VaR and ES forecasts for each dataset. In summary, the proposed models are the following:

| Market | Model |
|---|---|
| S&P500 | FIEGARCH-N |
| Gold Bullion $ per Troy Ounce | GARCH-GED/ IGARCH-GED |
| US dollar / British pound | EGARCH-N |

Although the most appropriate conditional volatility models are not the same for the three financial assets, they share some common characteristics. The Student-t and skewed Student-t distributions overestimate the *true* VaR. Asymmetry in volatility specification is inevitable, as all the selected models incorporate some form of asymmetry, whereas fractional integration is also important in forecasting the one-day-ahead VaR and ES numbers.

A VaR model selection procedure is proposed. As multiple risk management techniques exhibit unconditional and conditional coverage, the utility function of risk management must be brought into picture to evaluate statistically the differences among the adequate VaR models. Since the investor is also interested on the loss function, given a VaR violation, we introduce the ES measure to the loss function. According to the SPA test, the risk manager can select, for each financial asset, a separate model that forecasts both the risk measures accurately. Therefore, the number of under consideration techniques is reduced to a smaller set of competing models.

**Acknowledgement**

We would like to thank Peter R. Hansen for the availability of his program codes.

**References**


[1]. Acerbi, C. (2002). Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion. *Journal of Banking and Finance*, **26(7)**, 1505-1518.

[2]. Acerbi, C., Nordio, C., Sirtori, C. (2001). Expected Shortfall as a Tool for Financial Risk Management. *Working Paper, http://www.gloriamundi.org/var/wps.html*.

[3]. Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. *Proceedings of the Second International Symposium on Information Theory.* B.N. Petrov and F. Csaki (eds.), Budapest, 267-281.

[4]. Angelidis, T. and Degiannakis, S. (2005). Modeling Risk for Long and Short Trading Positions. *Journal of Risk Finance*, **6(3)**, 226-238.

[5]. Angelidis, T., Benos, A. and Degiannakis, S. (2004). The Use of GARCH Models in VaR Estimation. *Statistical Methodology*, **1(2)**, 105-128.

[6]. Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1997). Thinking Coherently. *Risk*, **10**, 68-71.

[7]. Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance*, **9**, 203-228.

[8]. Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **74**, 3-30.

[9]. Bali, T. and Theodossiou, P. (2006). A Conditional-SGT-VaR Approach with Alternative GARCH Models. *Annals of Operations Research*, forthcoming.

[10]. Bams, D., Lehnert, T. and Wolff, C.C.P. (2005). An Evaluation Framework for Alternative VaR-Models. *Journal of International Money and Finance*, **24**, 944-958.

[11]. Basak, S., and Shapiro, A. (2001). Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices. *Review of Financial Studies*, **14**, 371-405.

[12]. Baumol, W.J. (1963). An Expected Gain Confidence Limit Criterion for Portfolio Selection. *Management Science*, **10**, 174-182.

[13]. Beder, T. (1995). VaR: Seductive but Dangerous. *Financial Analysts Journal*, **51**, 12-24.

[14]. Billio, M. and Pelizzon, L. (2000). Value-at-Risk: A Multivariate Switching Regime Approach. *Journal of Empirical Finance,* **7**, 531-554.

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

[16]. Bollerslev, T. (1987). A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics*, **69**, 542-547.

[17]. Bollerslev, T. and Mikkelsen, H.O. (1996). Modeling and Pricing Long-Memory in Stock Market Volatility. *Journal of Econometrics*, **73**, 151-184.

[18]. Brooks, C. and Persand, G. (2003). Volatility Forecasting for Risk Management. *Journal of Forecasting*, **22**, 1-22.

[19]. Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review*, **39**, 841-862.

[20]. Chung, C.F. (1999). Estimating the Fractionally Integrated GARCH Model. *National Taiwan University*, Working paper.

[21]. Davidson, J. (2004). Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model. *Journal of Business and Economic* Statistics, **22(1)**, 16-29.

[22]. Degiannakis, S. (2004). Volatility Forecasting: Evidence from a Fractional Integrated Asymmetric Power ARCH Skewed-t Model. *Applied Financial Economics,* **14**, 1333-1342.

[23]. Degiannakis, S. and Xekalaki, E. (2006). Assessing the Performance of a Prediction Error Criterion Model Selection Algorithm in the Context of ARCH Models. *Applied Financial Economics,* forthcoming.

[24]. Delbaen, F. (2002). Coherent Risk Measures on General Probability Spaces. *Advances in Finance and Stochastics*. Essays in Honour of Dieter Sondermann, in: K. Sandmann and P. J. Schönbucher (eds.), Springer, 1-38.

[25]. Ding, Z., Granger, C.W.J. and Engle, R.F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, **1**, 83-106.

[26]. Doornik, J.A. (2001). Ox: Object Oriented Matrix Programming, 3.0. *Timberlake Consultants Press*, London.

[27]. Dowd, K. (2002). Measuring Market Risk. *John Wiley & Sons Ltd.*, New York.

[28]. Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica,* **50**, 987-1008.

[29]. Engle, R.F. and Bollerslev, T. (1986). Modelling the Persistence of Conditional Variances. *Econometric Reviews*, **5(1)**, 1-50.

[30]. Fernandez, C. and Steel, M. (1998). On Bayesian Modelling of Fat Tails and Skewness. *Journal of the American Statistical Association*, **93**, 359–371.

[31]. Giot, P. and Laurent, S. (2003a). Value-at-Risk for Long and Short Trading Positions. *Journal of Applied Econometrics*, **18**, 641-664.

[32]. Giot, P. and Laurent, S. (2003b). Market risk in Commodity Markets: A VaR Approach. *Energy Economics*, **25**, 435-457.

[33]. Glosten, L., Jagannathan, R. and Runkle, D. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, **48**, 1779-1801.

[34]. Guermat, C. and Harris, R.D.F. (2002). Forecasting Value-at-Risk Allowing for Time Variation in the Variance and Kurtosis of Portfolio Returns. *International Journal of Forecasting*, **18**, 409-419.

[35]. Hansen, P.R. (2005). A Test for Superior Predictive Ability. *Journal of Business and Economic Statistics*, **23**, 365-380.

[36]. Hansen, P.R. and Lunde, A. (2006). Consistent Ranking of Volatility Models. *Journal of Econometrics*, **131**, 97-121.

[37]. Hoppe, R. (1999). Finance is not Physics. *Risk Professional*, **1(7)**.

[38]. Huang, Y.C and Lin, B-J. (2004). Value-at-Risk Analysis for Taiwan Stock Index Futures: Fat Tails and Conditional Asymmetries in Return Innovations. *Review of Quantitative Finance and Accounting*, **22**, 79-95.

[39]. Inui, K. and Kijima, M. (2005). On the Significance of Expected Shortfall as a Coherent Risk Measure. *Journal of Banking and Finance*, **29**, 853–864.

[40]. Kuester, K., Mittnik, S. and Paolella, M.S. (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics*, **4 (1)**, 53-89.

[41]. Kupiec, P.H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives*, **3**, 73-84.

[42]. Lambert, P. and Laurent, S. (2000). Modelling Skewness Dynamics in Series of Financial Data. *Institut de Statistique*, Louvain-la-Neuve, Discussion Paper.

[43]. Laurent, S. and Peters, J.-P. (2002). G@RCH 2.2: An Ox Package for Estimating and Forecasting Various ARCH Models. *Journal of Economic Surveys*, **16**, 447-485.

[44]. Lopez, J.A. (1999). Methods for Evaluating Value-at-Risk Estimates. *Federal Reserve Bank of New York, Economic Policy Review*, **2**, 3-17.

[45]. Mittnik, S. and Paolella, M.S. (2000). Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates. *Journal of Forecasting*, **19**, 313-333.

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

[47]. Patton, A.J. (2005). Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies, *London School of Economics*, Working Paper.

[48]. Politis, D.N. and Romano, J.P. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association*, **89**, 1303-1313.

[49]. Sarma, M., Thomas, S. and Shah, A. (2003). Selection of VaR models. *Journal of Forecasting*, **22(4)**, 337–358.

[50]. Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-464.

[51]. Schwert, G.W. (1989). Why Does Stock Market Volatility Change Over Time? *Journal of Finance*, **44**, 1115-1153.

[52]. So, M.K.P. and Yu, P.L.H. (2006). Empirical analysis of GARCH models in Value at Risk Estimation. *Journal of International Markets, Institutions and Money*, 16(2), 180-197.

[53]. Subbotin, M.T. (1923). On the Law of Frequency of Error. *Matematicheskii Sbornik*, **31**, 296-301.

[54]. Taleb, N. (1997). Against VaR. *Derivatives Strategy*, April.

[55]. Taylor, S. (1986). Modeling Financial Time Series. *John Wiley & Sons Ltd.*, New York.

[56]. Tse, Y.K. (1998). The Conditional Heteroskedasticity of the Yen-Dollar Exchange Rate. *Journal of Applied Econometrics*, **193**, 49-55.

[57]. Yamai, Y. and Yoshiba, T. (2005). Value-at-risk Versus Expected Shortfall: A Practical Perspective. *Journal of Banking and Finance,* **29(4)**, 997-1015.

## Figures and Tables

Table 1. Panel A. Conditional volatility model specifications.

| Model | | Eq. |
|---|---|---|
| GARCH | $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ | (2) |
| EGARCH | $\ln\sigma_t^2 = \omega(1-\beta) + (1+\alpha L)\left(\gamma_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \gamma_2\left(\left\|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right\| - E\left\|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right\|\right)\right) + \beta\ln\sigma_{t-1}^2$ | (3) |
| TARCH | $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma d_{t-1}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ | (4) |
| APARCH | $\sigma_t^\delta = \omega + \alpha(\|\varepsilon_{t-1}\| - \gamma\varepsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta$ | (5) |
| IGARCH | $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2$ | (6) |
| FIGARCH | $\sigma_t^2 = \omega + \left(1-\beta L-(1-aL)(1-L)^d\right)\varepsilon_t^2 + \beta\sigma_{t-1}^2$ | (7) |
| FIGARCHC | $\sigma_t^2 = \sigma^2(1-\beta) + \left(1-\beta L-(1-aL)(1-L)^d\right)(\varepsilon_t^2 - \sigma^2) + \beta\sigma_{t-1}^2$ | (8) |
| FIEGARCH | $\ln\sigma_t^2 = \omega(1-\beta) + \left((1+\alpha L)(1-L)^{-d}\left(\gamma_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \gamma_2\left(\left\|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right\| - E\left\|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right\|\right)\right)\right) + \beta\ln\sigma_{t-1}^2$ | (9) |
| FIAPARCH | $\sigma_t^\delta = \omega + \left(1-\beta L-(1-aL)(1-L)^d\right)(\|\varepsilon_t\| - \gamma\varepsilon_t)^\delta + \beta\sigma_{t-1}^\delta$ | (10) |
| FIAPARCHC | $\sigma_t^\delta = \sigma^2(1-\beta) + \left(1-\beta L-(1-aL)(1-L)^d\right)\left((\|\varepsilon_t\| - \gamma\varepsilon_t)^\delta - \sigma^2\right) + \beta\sigma_{t-1}^\delta$ | (11) |
| HYGARCH | $\sigma_t^2 = \omega + \left(1-\beta L-(1-aL)\left(1+\zeta\left((1-L)^d - 1\right)\right)\right)\varepsilon_t^2 + \beta\sigma_{t-1}^2$ | (12) |

Table 2. Panel B. One-step-ahead conditional variance predictions.

| Model | | Eq. |
|---|---|---|
| GARCH | $\sigma_{t+1\mid t}^{2}=\omega^{(t)}+\alpha^{(t)}\varepsilon_{t\mid t}^{2}+\beta^{(t)}\sigma_{t\mid t}^{2}$ | (18) |
| EGARCH | $\sigma_{t+1\mid t}^{2}=\exp\left(\omega^{(t)}\left(1-\beta^{(t)}\right)+\left(1+\alpha^{(t)}L\right)\left(\gamma_1^{(t)}\frac{\varepsilon_{t\mid t}}{\sigma_{t\mid t}}+\gamma_2^{(t)}\left(\left\lvert\frac{\varepsilon_{t\mid t}}{\sigma_{t\mid t}}\right\rvert-E\left\lvert\frac{\varepsilon_{t\mid t}}{\sigma_{t\mid t}}\right\rvert\right)\right)+\beta^{(t)}\ln\sigma_{t\mid t}^{2}\right)$ | (19) |
| TARCH | $\sigma_{t+1\mid t}^{2}=\omega^{(t)}+\alpha^{(t)}\varepsilon_{t\mid t}^{2}+\gamma^{(t)}d_t\varepsilon_{t\mid t}^{2}+\beta^{(t)}\sigma_{t\mid t}^{2}$ | (20) |
| APARCH | $\sigma_{t+1\mid t}^{2}=\left(\omega^{(t)}+\alpha^{(t)}\left(\lvert\varepsilon_{t\mid t}\rvert-\gamma^{(t)}\varepsilon_{t\mid t}\right)^{\delta^{(t)}}+\beta^{(t)}\sigma_{t\mid t}^{\delta^{(t)}}\right)^{2/\delta^{(t)}}$ | (21) |
| IGARCH | $\sigma_{t+1\mid t}^{2}=\omega^{(t)}+\alpha^{(t)}\varepsilon_{t\mid t}^{2}+\left(1-\alpha^{(t)}\right)\sigma_{t\mid t}^{2}$ | (22) |
| FIGARCH | $\sigma_{t+1\mid t}^{2}=\omega^{(t)}+\left(a^{(t)}-\beta^{(t)}\right)\varepsilon_{t\mid t}^{2}+\sum_{i=1}^{\infty}\left(\frac{d^{(t)}\Gamma\left(i-d^{(t)}\right)}{\Gamma\left(1-d^{(t)}\right)\Gamma(i+1)}L^{i}\left(\varepsilon_{t+1\mid t+i}^{2}-a^{(t)}\varepsilon_{t\mid t+i}^{2}\right)\right)+\beta^{(t)}\sigma_{t\mid t}^{2}$ | (23) |
| FIGARCHC | $\sigma_{t+1\mid t}^{2}=\left(a^{(t)}-\beta^{(t)}\right)\varepsilon_{t\mid t}^{2}+\sum_{i=1}^{\infty}\left(\frac{d^{(t)}\Gamma\left(i-d^{(t)}\right)}{\Gamma\left(1-d^{(t)}\right)\Gamma(i+1)}L^{i}\left(\varepsilon_{t+1\mid t+i}^{2}-a^{(t)}\varepsilon_{t\mid t+i}^{2}\right)\right)+\beta^{(t)}\sigma_{t\mid t}^{2}$ | (24) |
| FIEGARCH | $\sigma_{t+1\mid t}^{2}=\exp\left(\begin{array}{c}\omega^{(t)}\left(1-\beta^{(t)}\right)+\left[\gamma_1^{(t)}\frac{\varepsilon_{t\mid t}}{\sigma_{t\mid t}}+\gamma_2^{(t)}\left(\left\lvert\frac{\varepsilon_{t\mid t}}{\sigma_{t\mid t}}\right\rvert-E\left\lvert\frac{\varepsilon_{t\mid t}}{\sigma_{t\mid t}}\right\rvert\right)\right]\\+a^{(t)}\left[\gamma_1^{(t)}\frac{\varepsilon_{t-1\mid t}}{\sigma_{t-1\mid t}}+\gamma_2^{(t)}\left(\left\lvert\frac{\varepsilon_{t-1\mid t}}{\sigma_{t-1\mid t}}\right\rvert-E\left\lvert\frac{\varepsilon_{t-1\mid t}}{\sigma_{t-1\mid t}}\right\rvert\right)\right]+\beta^{(t)}\ln\sigma_{t\mid t}^{2}\\+\sum_{i=1}^{\infty}\left(\frac{\Gamma\left(i+d^{(t)}\right)}{\Gamma\left(d^{(t)}\right)\Gamma(i+1)}\right)\left(L^{i}+a^{(t)}L^{i+1}\right)\left[\gamma_1^{(t)}\frac{\varepsilon_{t\mid t+i}}{\sigma_{t\mid t+i}}+\gamma_2^{(t)}\left(\left\lvert\frac{\varepsilon_{t\mid t+i}}{\sigma_{t\mid t+i}}\right\rvert-E\left\lvert\frac{\varepsilon_{t\mid t+i}}{\sigma_{t\mid t+i}}\right\rvert\right)\right]\end{array}\right)$ | (25) |
| FIAPARCH | $\sigma_{t+1\mid t}^{2}=\left(\begin{array}{c}\omega^{(t)}+\left(a^{(t)}-\beta^{(t)}\right)\left(\lvert\varepsilon_{t\mid t}\rvert-\gamma^{(t)}\varepsilon_{t\mid t}\right)^{\delta^{(t)}}+\beta^{(t)}\sigma_{t\mid t}^{\delta}\\+\sum_{i=1}^{\infty}\left(\frac{d^{(t)}\Gamma\left(i-d^{(t)}\right)}{\Gamma\left(1-d^{(t)}\right)\Gamma(i+1)}L^{i}\left(\lvert\varepsilon_{t+1\mid t+i}\rvert-\gamma^{(t)}\varepsilon_{t+1\mid t+i}\right)^{\delta^{(t)}}-a^{(t)}\left(\lvert\varepsilon_{t\mid t+i}\rvert-\gamma^{(t)}\varepsilon_{t\mid t+i}\right)^{\delta^{(t)}}\right)\end{array}\right)^{2/\delta^{(t)}}$ | (26) |
| FIAPARCHC | $\sigma_{t+1\mid t}^{2}=\left(\begin{array}{c}\left(a^{(t)}-\beta^{(t)}\right)\left(\lvert\varepsilon_{t\mid t}\rvert-\gamma^{(t)}\varepsilon_{t\mid t}\right)^{\delta^{(t)}}+\beta^{(t)}\sigma_{t\mid t}^{\delta}\\+\sum_{i=1}^{\infty}\left(\frac{d^{(t)}\Gamma\left(i-d^{(t)}\right)}{\Gamma\left(1-d^{(t)}\right)\Gamma(i+1)}L^{i}\left(\lvert\varepsilon_{t+1\mid t+i}\rvert-\gamma^{(t)}\varepsilon_{t+1\mid t+i}\right)^{\delta^{(t)}}-a^{(t)}\left(\lvert\varepsilon_{t\mid t+i}\rvert-\gamma^{(t)}\varepsilon_{t\mid t+i}\right)^{\delta^{(t)}}\right)\end{array}\right)^{2/\delta^{(t)}}$ | (27) |
| HYGARCH | $\sigma_{t+1\mid t}^{2}=\left(\begin{array}{c}\omega^{(t)}+\left(a^{(t)}-\beta^{(t)}\right)\varepsilon_{t\mid t}^{2}+\beta^{(t)}\sigma_{t\mid t}^{2}\\+\sum_{i=1}^{\infty}\left(\frac{d^{(t)}\Gamma\left(i-d^{(t)}\right)}{\Gamma\left(1-d^{(t)}\right)\Gamma(i+1)}L^{i}\zeta^{(t)}\left(\varepsilon_{t+1\mid t+i}^{2}-a^{(t)}\varepsilon_{t\mid t+i}^{2}\right)\right)\end{array}\right)$ | (28) |

Panel A presents the Conditional volatility model specifications, where GARCH = Generalized ARCH, EGARCH = Exponential GARCH, TARCH = Threshold ARCH, APARCH = Asymmetric Power ARCH, IGARCH = Integrated ARCH, FIGARCH = Fractionally Integrated GARCH, FIGARCHC = Chung's FIGARCH, FIEGARCH = Fractionally Integrated EGARCH, FIAPARCH Fractionally Integrated APARCH, FIAPARCHC = Chung's FIAPARCH, HYGARCH = Hyperbolic GARCH. In TARCH model, $d_t = 1$ if $\varepsilon_t < 0$ and $d_t = 0$ otherwise.

Panel B presents One-step-ahead conditional variance predictions, where:

1. $L$ denotes the lag operator, i.e. $L^i \varepsilon_t = \varepsilon_{t-i}$ .

2. EGARCH model: $E\left|\varepsilon_{t|t}\sigma_{t|t}^{-1}\right| = \sqrt{2/\pi}$ for normal, $E\left|\varepsilon_{t|t}\sigma_{t|t}^{-1}\right| = \Gamma\left(\frac{v+1}{2}\right)\frac{2\sqrt{v-2}}{\sqrt{\pi}(v-1)\Gamma(v/2)}$ for Student-t, $E\left|\varepsilon_{t|t}\sigma_{t|t}^{-1}\right| = \lambda 2^{v^{-1}}\Gamma\left(2v^{-1}\right)\left(\Gamma\left(v^{-1}\right)\right)^{-1}$ for GED, and $E\left|\varepsilon_{t|t}\sigma_{t|t}^{-1}\right| = \frac{4g^2}{g+g^{-1}}\frac{\Gamma\left(\frac{v+1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}(v-1)\Gamma(v/2)}$ for skewed Student-t distribution.

3. Fractionally Integrated Models: $\sum_{i=1}^{\infty} d\left(\frac{\Gamma(i-d)}{\Gamma(1-d)\Gamma(i+1)}L^i\right) = (1!)^{-1}dL + (2!)^{-1}d(1-d)L^2 + (3!)^{-1}d(1-d)(2-d)L^3 + \ldots,\ d>0$ and $\sum_{i=1}^{\infty}\left(\frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)}L^i\right) = \frac{1}{1!}dL + \frac{1}{2!}d(1+d)L^2 + +\frac{1}{3!}d(1+d)(2+d)L^3 + \ldots,\ d>0$ .

Table 3. The S&P500 case. Column 1 presents the models that have not been rejected by the backtesting criteria (unconditional coverage and the independence test). Columns 2 and 3 present the values of the MAE and the MSE loss functions multiplied by $10^3$ (in parentheses the ranking of the models is presented). The average values of the VaR and ES forecasts are presented in 4th and 5th columns, respectively. The percentage of violations is presented in 6th column, whereas the 7th and 8th columns present the Kupiec's and Christofersen's p-values, respectively.

| **Model** | **MAE (Rank)** | **MSE (Rank)** | **Av.Var** | **Av.ES** | **Rate** | **Kupiec** | **Chr/sen** |
|---|---|---|---|---|---|---|---|
| **Panel A. Long Position - 95% VaR** | | | | | | | |
| FIEGARCH-GED | 19.209(1) | 18.642(1) | -1.964 | -2.664 | 4.18% | 14.35% | 14.41% |
| EGARCH-N | 19.868(2) | 24.350(12) | -1.848 | -2.324 | 5.16% | 78.62% | 54.10% |
| FIEGARCH-N | 20.028(3) | 21.554(3) | -1.879 | -2.365 | 5.16% | 78.62% | 27.33% |
| TARCH-N | 20.195(4) | 24.638(13) | -1.830 | -2.302 | 5.30% | 61.00% | 32.97% |
| APARCH-N | 20.230(5) | 23.944(10) | -1.870 | -2.352 | 5.23% | 69.59% | 58.03% |
| HYGARCH-N | 20.269(6) | 23.767(9) | -1.894 | -2.389 | 4.95% | 92.75% | 43.06% |
| FIAPARCH-N | 20.681(7) | 25.742(14) | -1.890 | -2.377 | 5.09% | 88.00% | 50.28% |
| FIAPARCHC-N | 21.112(8) | 27.365(15) | -1.942 | -2.441 | 4.46% | 33.93% | 50.20% |
| IGARCH-N | 21.473(9) | 24.213(11) | -1.883 | -2.374 | 5.23% | 69.59% | 30.07% |
| FIAPARCHC-GED | 21.537(10) | 22.817(5) | -1.967 | -2.668 | 4.25% | 18.19% | 79.73% |
| HYGARCH-GED | 21.598(11) | 22.799(4) | -1.907 | -2.616 | 4.88% | 83.15% | 17.99% |
| EGARCH-GED | 21.833(12) | 21.407(2) | -1.952 | -2.659 | 4.53% | 40.64% | 54.01% |
| FIGARCHC-N | 22.221(13) | 27.486(17) | -1.837 | -2.317 | 5.37% | 52.95% | 36.03% |
| TARCH-GED | 22.279(14) | 22.944(7) | -1.856 | -2.534 | 5.09% | 88.00% | 50.28% |
| APARCH-GED | 22.376(15) | 22.903(6) | -1.901 | -2.588 | 4.88% | 83.15% | 74.62% |
| FIAPARCH-GED | 22.388(16) | 23.726(8) | -1.912 | -2.591 | 4.81% | 73.75% | 70.33% |
| FIGARCH-N | 23.691(17) | 28.718(18) | -1.799 | -2.269 | 5.64% | 27.19% | 49.82% |
| FIGARCH-GED | 25.598(18) | 27.420(16) | -1.820 | -2.494 | 5.71% | 22.43% | 13.61% |
| **Panel B. Long Position - 99% VaR** | | | | | | | |
| APARCH-GED | 3.938(1) | 4.635(2) | -3.015 | -3.651 | 0.63% | 12.75% | 73.60% |
| EGARCH-GED | 4.383(2) | 3.914(1) | -3.097 | -3.751 | 0.70% | 22.22% | 70.78% |
| GARCH-GED | 4.711(3) | 5.412(3) | -3.003 | -3.658 | 0.63% | 12.75% | 73.60% |
| FIAPARCH-GED | 4.855(4) | 6.221(4) | -3.014 | -3.637 | 0.63% | 12.75% | 73.60% |
| FIEGARCH-N | 5.450(5) | 7.381(5) | -2.672 | -3.066 | 0.91% | 71.59% | 62.58% |
| FIGARCH-GED | 6.322(6) | 8.158(6) | -2.913 | -3.540 | 0.77% | 35.40% | 68.01% |
| HYGARCH-N | 6.456(7) | 10.304(8) | -2.701 | -3.103 | 1.12% | 66.73% | 54.79% |
| APARCH-N | 6.813(8) | 10.057(7) | -2.656 | -3.046 | 0.98% | 92.57% | 59.93% |
| FIAPARCHC-N | 6.836(9) | 12.936(11) | -2.756 | -3.161 | 0.98% | 92.57% | 12.42% |
| EGARCH-N | 6.965(10) | 10.323(9) | -2.625 | -3.011 | 1.05% | 86.41% | 57.33% |
| TARCH-N | 7.487(11) | 10.782(10) | -2.600 | -2.983 | 1.18% | 49.45% | 52.30% |

Table 3. Continued

| Model | MAE (Rank) | MSE (Rank) | Av.Var | Av.ES | Rate | Kupiec | Chr/sen |
|---|---|---|---|---|---|---|---|
| **Panel C. Sort Position - 95% VaR** | | | | | | | |
| APARCH-N | 15.106 (1) | 9.702 (1) | 1.921 | 2.402 | 4.53% | 40.64% | 17.32% |
| EGARCH-N | 16.357 (2) | 12.342 (5) | 1.902 | 2.378 | 4.53% | 40.64% | 17.32% |
| FIEGARCH-N | 16.541 (3) | 12.115 (4) | 1.951 | 2.438 | 4.25% | 18.19% | 79.73% |
| TARCH-N | 17.550 (4) | 11.554 (2) | 1.890 | 2.362 | 4.81% | 73.75% | 12.08% |
| IGARCH-GED | 17.715 (5) | 12.351 (6) | 1.972 | 2.683 | 4.11% | 11.15% | 28.27% |
| GARCH-N | 17.790 (6) | 15.400 (10) | 1.948 | 2.429 | 4.32% | 22.71% | 22.29% |
| GARCH-GED | 18.061 (7) | 12.732 (7) | 1.949 | 2.650 | 4.18% | 14.35% | 26.16% |
| FIGARCHC-N | 18.267 (8) | 14.966 (9) | 1.944 | 2.424 | 4.32% | 22.71% | 65.03% |
| APARCH-GED | 18.314 (9) | 11.614 (3) | 1.956 | 2.643 | 4.39% | 27.95% | 20.52% |
| EGARCH-GED | 19.484 (10) | 13.706 (8) | 1.993 | 2.699 | 4.11% | 11.15% | 71.09% |
| FIGARCH-N | 20.041 (11) | 16.353 (12) | 1.906 | 2.376 | 4.81% | 73.75% | 41.42% |
| FIGARCHC-GED | 21.789 (12) | 16.574 (13) | 1.936 | 2.621 | 4.39% | 27.95% | 61.29% |
| TARCH-GED | 23.007 (13) | 16.134 (11) | 1.917 | 2.595 | 5.02% | 97.59% | 33.23% |
| FIGARCH-GED | 23.649 (14) | 17.524 (14) | 1.904 | 2.579 | 4.95% | 92.75% | 78.97% |
| **Panel D. Sort Position - 99% VaR** | | | | | | | |
| APARCH-N | 1.968 (1) | 0.963 (1) | 2.707 | 3.097 | 0.77% | 35.40% | 68.01% |
| IGARCH-GED | 2.669 (2) | 1.600 (5) | 3.124 | 3.790 | 0.63% | 12.75% | 73.60% |
| FIGARCH-GED | 2.726 (3) | 1.564 (4) | 2.997 | 3.624 | 0.63% | 12.75% | 73.60% |
| TARCH-N | 2.747 (4) | 1.200 (2) | 2.661 | 3.044 | 0.98% | 92.57% | 59.93% |
| FIGARCHC-GED | 2.829 (5) | 1.702 (6) | 3.046 | 3.682 | 0.63% | 12.75% | 73.60% |
| FIEGARCH-N | 2.874 (6) | 1.436 (3) | 2.745 | 3.139 | 0.98% | 92.57% | 59.93% |
| GARCH-GED | 3.212 (7) | 1.745 (7) | 3.084 | 3.739 | 0.77% | 35.40% | 68.01% |
| EGARCH-N | 3.380 (8) | 2.441 (10) | 2.679 | 3.065 | 0.98% | 92.57% | 59.93% |
| IGARCH-N | 3.473 (9) | 2.002 (8) | 2.786 | 3.184 | 1.05% | 86.41% | 57.33% |
| HYGARCH-N | 3.600 (10) | 2.231 (9) | 2.811 | 3.212 | 0.98% | 92.57% | 59.93% |
| FIAPARCH-N | 3.702 (11) | 2.959 (13) | 2.740 | 3.135 | 0.98% | 92.57% | 59.93% |
| FIGARCHC-N | 3.828 (12) | 2.622 (12) | 2.727 | 3.116 | 1.05% | 86.41% | 57.33% |
| GARCH-N | 4.337 (13) | 2.597 (11) | 2.733 | 3.124 | 1.18% | 49.45% | 52.30% |
| FIGARCH-N | 4.683 (14) | 3.263 (14) | 2.673 | 3.055 | 1.25% | 35.16% | 49.87% |

Table 4. The Gold Bullion $ per Troy Ounce case. Column 1 presents the models that have not been rejected by the backtesting criteria (unconditional coverage and the independence test). Columns 2 and 3 present the values of the MAE and the MSE loss functions multiplied by $10^3$ (in parentheses the ranking of the models is presented). The average values of the VaR and ES forecasts are presented in $4^{th}$ and $5^{th}$ columns, respectively. The percentage of violations is presented in $6^{th}$ column, whereas the $7^{th}$ and $8^{th}$ columns present the Kupiec's and Christofersen's p-values, respectively.

| Model | MAE (Rank) | MSE (Rank) | Av.Var | Av.ES | Rate | Kupiec | Chr/sen |
|---|---|---|---|---|---|---|---|
| **Panel A. Long Position - 95% VaR** | | | | | | | |
| FIGARCHC-N | 18.342 (1) | 17.047 (8) | -1.426 | -1.785 | 4.11% | 11.15% | 71.09% |
| FIGARCH-N | 18.581 (2) | 17.317 (9) | -1.425 | -1.783 | 4.11% | 11.15% | 76.84% |
| EGARCH-N | 20.169 (3) | 21.477 (10) | -1.436 | -1.801 | 4.18% | 14.35% | 75.38% |
| FIAPARCHC-GED | 20.296 (4) | 15.772 (6) | -1.415 | -2.054 | 4.18% | 14.35% | 75.38% |
| APARCH-GED | 20.311 (5) | 15.128 (2) | -1.450 | -2.113 | 4.11% | 11.15% | 71.09% |
| TARCH-GED | 20.355 (6) | 15.094 (1) | -1.464 | -2.133 | 4.11% | 11.15% | 71.09% |
| FIGARCHC-GED | 20.430 (7) | 15.652 (3) | -1.394 | -2.027 | 4.39% | 27.95% | 88.53% |
| FIGARCH-GED | 20.657 (8) | 15.710 (4) | -1.397 | -2.030 | 4.39% | 27.95% | 88.53% |
| GARCH-GED | 21.165 (9) | 15.753 (5) | -1.438 | -2.104 | 4.32% | 22.71% | 65.03% |
| IGARCH-GED | 21.272 (10) | 16.043 (7) | -1.442 | -2.109 | 4.32% | 22.71% | 65.03% |
| **Panel B. Long Position - 99% VaR** | | | | | | | |
| EGARCH-GED | 2.012 (1) | 0.916 (1) | -2.692 | -3.431 | 0.70% | 22.22% | 70.78% |
| TARCH-GED | 2.463 (2) | 1.469 (2) | -2.540 | -3.224 | 0.70% | 22.22% | 70.78% |
| FIEGARCH-GED | 3.011 (3) | 1.626 (3) | -2.659 | -3.388 | 0.70% | 22.22% | 70.78% |
| FIAPARCH-GED | 3.040 (4) | 1.862 (4) | -2.479 | -3.136 | 0.84% | 52.11% | 65.27% |
| FIAPARCHC-GED | 3.136 (5) | 1.926 (5) | -2.442 | -3.091 | 0.84% | 52.11% | 65.27% |
| HYGARCH-GED | 3.292 (6) | 1.927 (6) | -2.509 | -3.187 | 0.91% | 71.59% | 62.58% |
| GARCH-GED | 3.299 (7) | 2.452 (8) | -2.508 | -3.191 | 0.77% | 35.40% | 68.01% |
| IGARCH-GED | 3.412 (8) | 2.563 (10) | -2.514 | -3.200 | 0.77% | 35.40% | 68.01% |
| FIGARCH-GED | 3.565 (9) | 2.214 (7) | -2.415 | -3.061 | 0.91% | 71.59% | 62.58% |
| APARCH-GED | 4.019 (10) | 2.811 (11) | -2.516 | -3.194 | 0.84% | 52.11% | 65.27% |
| FIGARCHC-GED | 4.077 (11) | 2.505 (9) | -2.412 | -3.057 | 0.98% | 92.57% | 59.93% |
| TARCH-N | 6.038 (12) | 5.005 (12) | -2.120 | -2.428 | 1.25% | 35.16% | 49.87% |
| APARCH-N | 6.317 (13) | 5.158 (13) | -2.100 | -2.406 | 1.32% | 23.99% | 47.50% |
| IGARCH-N | 6.440 (14) | 5.685 (15) | -2.093 | -2.396 | 1.32% | 23.99% | 47.50% |
| GARCH-N | 6.742 (15) | 5.788 (16) | -2.086 | -2.387 | 1.39% | 15.71% | 45.19% |
| FIAPARCH-N | 6.859 (16) | 6.215 (17) | -2.075 | -2.376 | 1.25% | 35.16% | 49.87% |
| HYGARCH-N | 6.860 (17) | 5.635 (14) | -2.085 | -2.386 | 1.39% | 15.71% | 45.19% |

Table 4. Continued

| Model | MAE (Rank) | MSE (Rank) | Av.Var | Av.ES | Rate | Kupiec | Chr/sen |
|---|---|---|---|---|---|---|---|
| **Panel C. Sort Position - 95% VaR** | | | | | | | |
| FIEGARCH-GED | 21.571(1) | 31.570(1) | 1.523 | 2.230 | 4.18% | 14.35% | 73.22% |
| APARCH-N | 22.838(2) | 39.708(15) | 1.480 | 1.857 | 4.53% | 40.64% | 52.08% |
| TARCH-N | 23.160(3) | 39.281(14) | 1.492 | 1.872 | 4.67% | 56.09% | 59.93% |
| IGARCH-N | 23.802(4) | 42.209(19) | 1.454 | 1.828 | 4.67% | 56.09% | 59.93% |
| HYGARCH-N | 23.940(5) | 38.789(13) | 1.448 | 1.820 | 4.95% | 92.75% | 76.78% |
| FIAPARCH-N | 23.964(6) | 37.328(5) | 1.455 | 1.826 | 4.88% | 83.15% | 72.46% |
| FIAPARCHC-N | 23.992(7) | 38.619(12) | 1.430 | 1.794 | 4.88% | 83.15% | 72.46% |
| EGARCH-N | 24.157(8) | 45.752(21) | 1.439 | 1.803 | 4.53% | 40.64% | 52.08% |
| GARCH-N | 24.287(9) | 42.504(20) | 1.449 | 1.821 | 4.81% | 73.75% | 68.20% |
| APARCH-GED | 24.547(10) | 37.366(6) | 1.450 | 2.113 | 4.74% | 64.69% | 64.02% |
| GARCH-GED | 24.576(11) | 38.025(11) | 1.439 | 2.104 | 4.88% | 83.15% | 72.46% |
| IGARCH-GED | 24.672(12) | 37.849(10) | 1.442 | 2.109 | 4.88% | 83.15% | 72.46% |
| TARCH-GED | 24.781(13) | 36.966(3) | 1.464 | 2.133 | 4.74% | 64.69% | 91.62% |
| FIEGARCH-N | 25.078(14) | 41.954(18) | 1.382 | 1.738 | 4.81% | 73.75% | 34.85% |
| HYGARCH-GED | 25.455(15) | 36.201(2) | 1.444 | 2.106 | 5.02% | 97.59% | 81.16% |
| FIAPARCH-GED | 26.332(16) | 37.076(4) | 1.437 | 2.084 | 5.09% | 88.00% | 85.58% |
| FIGARCHC-N | 26.620(17) | 41.097(17) | 1.397 | 1.755 | 5.71% | 22.43% | 75.06% |
| FIGARCH-N | 26.706(18) | 41.083(16) | 1.396 | 1.754 | 5.71% | 22.43% | 75.06% |
| FIGARCH-GED | 27.028(19) | 37.510(8) | 1.396 | 2.030 | 5.44% | 45.51% | 92.20% |
| FIGARCHC-GED | 27.081(20) | 37.477(7) | 1.393 | 2.026 | 5.44% | 45.51% | 92.20% |
| FIAPARCHC-GED | 27.445(21) | 37.599(9) | 1.415 | 2.054 | 5.30% | 61.00% | 98.94% |
| **Panel D. Sort Position - 99% VaR** | | | | | | | |
| FIAPARCH-GED | 7.017(1) | 13.462(1) | 2.478 | 3.136 | 1.05% | 86.41% | 14.57% |
| FIAPARCHC-GED | 8.003(2) | 14.782(2) | 2.442 | 3.091 | 1.12% | 66.73% | 16.92% |
| FIGARCH-GED | 8.131(3) | 15.040(3) | 2.415 | 3.060 | 1.05% | 86.41% | 14.57% |
| GARCH-GED | 8.162(4) | 17.210(6) | 2.508 | 3.191 | 0.98% | 92.57% | 12.42% |
| IGARCH-GED | 8.176(5) | 16.925(5) | 2.514 | 3.199 | 0.98% | 92.57% | 12.42% |
| EGARCH-GED | 8.635(6) | 20.424(7) | 2.692 | 3.431 | 0.91% | 71.59% | 10.44% |
| FIGARCHC-GED | 8.909(7) | 15.308(4) | 2.411 | 3.056 | 1.18% | 49.45% | 19.44% |

Table 5. The US $ to UK £ case. Column 1 presents the models that have not been rejected by the backtesting criteria (unconditional coverage and the independence test). Columns 2 and 3 present the values of the MAE and the MSE loss functions multiplied by $10^3$ (in parentheses the ranking of the models is presented). The average values of the VaR and ES forecasts are presented in 4th and 5th columns, respectively. The percentage of violations is presented in 6th column, whereas the 7th and 8th columns present the Kupiec's and Christofersen's p-values, respectively.

| Model | MAE (Rank) | MSE (Rank) | Av.Var | Av.ES | Rate | Kupiec | Chr/sen |
|---|---|---|---|---|---|---|---|
| **Panel A. Long Position - 95% VaR** | | | | | | | |
| EGARCH-N | 8.402 (1) | 2.606 (1) | -0.886 | -1.112 | 4.67% | 56.09% | 61.97% |
| FIEGARCH-N | 8.757 (2) | 2.762 (2) | -0.895 | -1.123 | 4.74% | 64.69% | 44.41% |
| IGARCH-N | 9.172 (3) | 2.824 (3) | -0.864 | -1.084 | 5.09% | 88.00% | 68.64% |
| FIAPARCHC-GED | 9.976 (4) | 2.876 (4) | -0.833 | -1.167 | 5.44% | 45.51% | 50.29% |
| FIGARCHC-N | 9.987 (5) | 3.173 (7) | -0.834 | -1.047 | 5.44% | 45.51% | 20.36% |
| FIAPARCH-N | 10.150 (6) | 3.282 (12) | -0.831 | -1.043 | 5.64% | 27.19% | 40.80% |
| HYGARCH-GED | 10.181 (7) | 3.059 (5) | -0.842 | -1.180 | 5.37% | 52.95% | 53.72% |
| FIGARCH-N | 10.265 (8) | 3.288 (13) | -0.835 | -1.048 | 5.57% | 32.61% | 17.04% |
| HYGARCH-N | 10.400 (9) | 3.362 (15) | -0.830 | -1.041 | 5.57% | 32.61% | 17.04% |
| FIAPARCHC-N | 10.568 (10) | 3.410 (18) | -0.823 | -1.033 | 5.78% | 18.32% | 68.94% |
| GARCH-N | 10.570 (11) | 3.391 (16) | -0.832 | -1.045 | 5.64% | 27.19% | 77.13% |
| FIGARCHC-GED | 10.581 (12) | 3.260 (11) | -0.835 | -1.170 | 5.51% | 38.72% | 47.00% |
| FIGARCH-GED | 10.584 (13) | 3.230 (10) | -0.844 | -1.183 | 5.44% | 45.51% | 50.29% |
| TARCH-N | 10.592 (14) | 3.393 (17) | -0.834 | -1.047 | 5.71% | 22.43% | 72.99% |
| EGARCH-GED | 10.644 (15) | 3.161 (6) | -0.903 | -1.272 | 4.67% | 56.09% | 61.97% |
| FIAPARCH-GED | 10.694 (16) | 3.218 (9) | -0.842 | -1.180 | 5.51% | 38.72% | 47.00% |
| IGARCH-GED | 10.777 (17) | 3.212 (8) | -0.863 | -1.213 | 5.16% | 78.62% | 64.74% |
| GARCH-GED | 11.255 (18) | 3.360 (14) | -0.843 | -1.181 | 5.51% | 38.72% | 85.65% |
| TARCH-GED | 11.646 (19) | 3.482 (19) | -0.845 | -1.183 | 5.64% | 27.19% | 77.13% |
| APARCH-GED | 12.052 (20) | 4.656 (20) | -0.830 | -1.161 | 5.71% | 22.43% | 72.99% |
| **Panel B. Long Position - 99% VaR** | | | | | | | |
| FIGARCHC-GED | 1.340 (1) | 0.360 (1) | -1.376 | -1.701 | 0.77% | 35.40% | 68.01% |
| IGARCH-GED | 1.397 (2) | 0.381 (4) | -1.428 | -1.769 | 0.70% | 22.22% | 70.78% |
| GARCH-GED | 1.453 (3) | 0.382 (5) | -1.389 | -1.716 | 0.84% | 52.11% | 65.27% |
| FIAPARCH-GED | 1.488 (4) | 0.377 (3) | -1.388 | -1.715 | 0.84% | 52.11% | 65.27% |
| EGARCH-GED | 1.545 (5) | 0.402 (6) | -1.499 | -1.858 | 0.70% | 22.22% | 70.78% |
| HYGARCH-GED | 1.546 (6) | 0.407 (7) | -1.388 | -1.714 | 0.84% | 52.11% | 65.27% |
| FIGARCH-GED | 1.569 (7) | 0.411 (8) | -1.391 | -1.719 | 0.84% | 52.11% | 65.27% |
| EGARCH-N | 1.572 (8) | 0.364 (2) | -1.254 | -1.437 | 1.39% | 15.71% | 45.19% |
| TARCH-GED | 1.616 (9) | 0.433 (9) | -1.391 | -1.718 | 0.91% | 71.59% | 62.58% |
| FIEGARCH-GED | 1.658 (10) | 0.543 (11) | -1.586 | -1.975 | 0.63% | 12.75% | 73.60% |
| FIAPARCHC-GED | 1.669 (11) | 0.464 (10) | -1.373 | -1.696 | 0.77% | 35.40% | 68.01% |
| APARCH-GED | 2.020 (12) | 1.171 (12) | -1.364 | -1.684 | 0.91% | 71.59% | 62.58% |

Table 5. Continued

| Model | MAE (Rank) | MSE (Rank) | Av.Var | Av.ES | Rate | Kupiec | Chr/sen |
|---|---|---|---|---|---|---|---|
| **Panel C. Sort Position - 95% VaR** | | | | | | | |
| EGARCH-N | 7.754 (1) | 2.839 (1) | 0.894 | 1.120 | 4.25% | 18.19% | 79.73% |
| IGARCH-N | 8.369 (2) | 3.057 (5) | 0.872 | 1.092 | 4.53% | 40.64% | 54.01% |
| FIGARCHC-N | 8.499 (3) | 3.046 (4) | 0.842 | 1.055 | 4.81% | 73.75% | 70.33% |
| TARCH-N | 8.639 (4) | 3.013 (2) | 0.844 | 1.057 | 4.88% | 83.15% | 74.62% |
| FIGARCH-N | 8.926 (5) | 3.107 (8) | 0.843 | 1.056 | 4.95% | 92.75% | 78.97% |
| HYGARCH-N | 8.950 (6) | 3.107 (9) | 0.838 | 1.050 | 5.02% | 97.59% | 83.37% |
| GARCH-N | 8.966 (7) | 3.085 (7) | 0.842 | 1.055 | 5.02% | 97.59% | 83.37% |
| FIAPARCH-N | 9.016 (8) | 3.220 (14) | 0.839 | 1.051 | 4.95% | 92.75% | 78.97% |
| FIAPARCHC-N | 9.083 (9) | 3.245 (15) | 0.830 | 1.040 | 5.23% | 69.59% | 96.72% |
| FIAPARCH-GED | 9.273 (10) | 3.013 (3) | 0.847 | 1.185 | 4.46% | 33.93% | 22.58% |
| APARCH-N | 9.545 (11) | 3.344 (16) | 0.823 | 1.031 | 5.37% | 52.95% | 94.40% |
| FIGARCH-GED | 9.659 (12) | 3.084 (6) | 0.848 | 1.187 | 4.60% | 48.04% | 57.94% |
| HYGARCH-GED | 9.693 (13) | 3.116 (10) | 0.846 | 1.184 | 4.67% | 56.09% | 61.97% |
| FIGARCHC-GED | 9.818 (14) | 3.153 (12) | 0.839 | 1.174 | 4.74% | 64.69% | 66.11% |
| GARCH-GED | 10.029 (15) | 3.134 (11) | 0.848 | 1.186 | 4.60% | 48.04% | 57.94% |
| TARCH-GED | 10.033 (16) | 3.183 (13) | 0.850 | 1.188 | 4.53% | 40.64% | 54.01% |
| FIAPARCHC-GED | 10.345 (17) | 3.427 (19) | 0.837 | 1.172 | 4.88% | 83.15% | 39.66% |
| IGARCH-GED | 10.376 (18) | 3.392 (17) | 0.867 | 1.217 | 4.46% | 33.93% | 50.20% |
| APARCH-GED | 10.668 (19) | 3.407 (18) | 0.835 | 1.166 | 4.88% | 83.15% | 39.66% |
| **Panel D. Sort Position - 99% VaR** | | | | | | | |
| EGARCH-N | 1.922 (1) | 0.623 (1) | 1.262 | 1.446 | 0.91% | 71.59% | 10.44% |
| FIEGARCH-N | 1.999 (2) | 0.650 (2) | 1.273 | 1.458 | 0.98% | 92.57% | 12.42% |
| TARCH-N | 2.175 (3) | 0.695 (3) | 1.192 | 1.364 | 1.05% | 86.41% | 14.57% |
| IGARCH-N | 2.246 (4) | 0.730 (5) | 1.231 | 1.410 | 1.05% | 86.41% | 14.57% |
| GARCH-N | 2.256 (5) | 0.701 (4) | 1.189 | 1.362 | 1.12% | 66.73% | 16.92% |
| HYGARCH-N | 2.501 (6) | 0.786 (6) | 1.184 | 1.355 | 1.25% | 35.16% | 22.13% |
| FIAPARCH-N | 2.511 (7) | 0.852 (9) | 1.185 | 1.357 | 1.12% | 66.73% | 16.92% |
| FIGARCH-N | 2.585 (8) | 0.816 (7) | 1.190 | 1.363 | 1.25% | 35.16% | 22.13% |
| APARCH-N | 2.614 (9) | 0.849 (8) | 1.162 | 1.330 | 1.25% | 35.16% | 22.13% |
| FIAPARCHC-N | 2.647 (10) | 0.908 (11) | 1.172 | 1.343 | 1.25% | 35.16% | 22.13% |
| FIGARCHC-N | 2.753 (11) | 0.877 (10) | 1.189 | 1.362 | 1.32% | 23.99% | 25.00% |

Table 6. The p-values of the SPA test for the null hypothesis that the benchmark model is the best model.

| Loss Function | Long Position 95% VaR | Long Position 99% VaR | Short Position 95% VaR | Short Position 99% VaR |
|---|---|---|---|---|
| | | S&P500 (Benchmark Model: FIEGARCH-N) | | |
| MAE | 0.81390 | 0.12790 | 0.61030 | 0.26370 |
| MSE | 0.34780 | 0.17080 | 0.28250 | 0.38550 |
| | | (Benchmark Model: EGARCH-N) | | |
| MAE | 0.89750 | 0.08000[b] | 0.59720 | 0.20970 |
| MSE | 0.32300 | 0.11730 | 0.33850 | 0.43500 |
| | | (Benchmark Model: APARCH-N) | | |
| MAE | 0.87810 | 0.04360[a] | 0.97800 | 0.89900 |
| MSE | 0.34690 | 0.11450 | 0.99820 | 0.99050 |
| | | (Benchmark Model: TARCH-N) | | |
| MAE | 0.88740 | 0.02690[a] | 0.16050 | 0.48620 |
| MSE | 0.35010 | 0.12440 | 0.07160[b] | 0.69910 |
| | | (Benchmark Model: FIGARCH-GED) | | |
| MAE | 0.01340[a] | 0.09190[b] | 0.00600[a] | 0.79410 |
| MSE | 0.06970[b] | 0.19970 | 0.00330[a] | 0.77080 |

Table 6. Continued

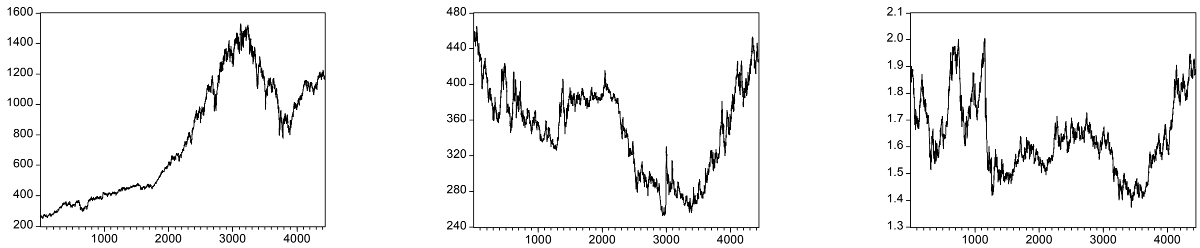
| Loss Function | Long Position 95% VaR | Long Position 99% VaR | Short Position 95% VaR | Short Position 99% VaR |
|---|---|---|---|---|
| Gold Bullion $ per Troy Ounce (Benchmark Model: GARCH-GED) | | | | |
| MAE | 0.36750 | 0.34000 | 0.34680 | 0.56090 |
| MSE | 0.84510 | 0.32590 | 0.43570 | 0.34920 |
| (Benchmark Model: IGARCH-GED) | | | | |
| MAE | 0.31350 | 0.35870 | 0.32110 | 0.53350 |
| MSE | 0.23210 | 0.32260 | 0.43330 | 0.43680 |
| (Benchmark Model: FIGARCH-GED) | | | | |
| MAE | 0.26480 | 0.21560 | 0.06820[b] | 0.38710 |
| MSE | 0.81200 | 0.11440 | 0.20400 | 0.47350 |
| (Benchmark Model: FIGARCHC-GED) | | | | |
| MAE | 0.32250 | 0.11250 | 0.06840[b] | 0.10370 |
| MSE | 0.84020 | 0.07850[b] | 0.19470 | 0.28370 |
| (Benchmark Model: FIAPARCHC-GED) | | | | |
| MAE | 0.37010 | 0.41260 | 0.02870[a] | 0.12790 |
| MSE | 0.83110 | 0.37380 | 0.05430[b] | 0.05330[b] |
| US $ to UK £ (Benchmark Model: EGARCH-N) | | | | |
| MAE | 0.95180 | 0.72780 | 0.96560 | 0.78310 |
| MSE | 0.97730 | 0.89700 | 0.97300 | 0.91270 |

[a] Indicates that the null hypothesis is rejected at 5% level of significance.
[b] Indicates that the null hypothesis is rejected at 10% level of significance.

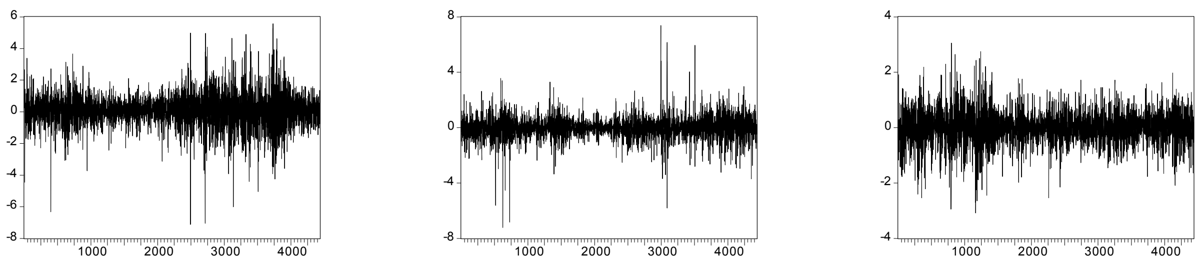
Figure 1. Daily closing prices, log-returns and the lag 1 through 1000 autocorrelations for the absolute log-returns from April 4th, 1988 through April 5th, 2005.

| S&P500 | Gold Bullion $ per Troy Ounce | US $ to UK £ |
|---|---|---|

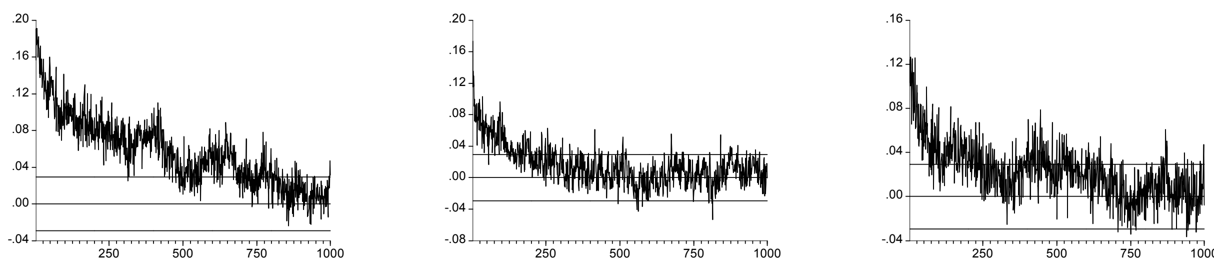
Daily closing prices




Daily log-returns




Sample autocorrelations for the absolute daily log-returns, $Cor(|y_t|,|y_{t-\tau}|)$, for $\tau = 1,...,1000$.*




* The vertical lines present the 95% confidence interval of no serial dependence, $\pm 1.96/\sqrt{T}$ .

[i]Mittnik and Paoella (2000) also used the APARCH model to accommodate the time varying skewness of the exchange rate market.

[ii]A coherent risk measure is defined as one that satisfies the following four properties: (a) sub-additivity, (b) homogeneity, (c) monotonicity, and (d) risk-free condition. These are described in the following equations: (a) $\rho(x)+\rho(y)\leq\rho(x+y)$, (b) $\rho(tx)=t\rho(x)$, (c) $\rho(x)\geq\rho(y)$ if $x \leq y$, and (d) $\rho(x+n)=\rho(x) - n$. For more details on coherent risk measures, see Artzner et al. (1997).

[iii]As the *4:15 report* JP Morgan did.

[iv]Bali and Theodosiou (2006) suggested either the TS-GARCH, proposed by Taylor (1986) and Schwert (1989), or the EGARCH model, introduced by Nelson (1991), as they estimate the VaR and ES measures accurately and give the best overall performance.

[v]The skewed Student-t distribution was introduced by Fernandez and Steel (1998) and was applied by Lambert and Laurent (2000) in ARCH framework. Moreover, Kuester et al. (2006) argued that compared to the normal distribution, substantial improvement in predicting VaR was achieved when asymmetrical fat tailed distribution was used.

[vi]The log-likehood ratio statistic of the combined hypothesis is computed as: $LR_{cc} = -2\ln\left((1-\rho)^{T-N}\rho^{N}\right)+2\ln\left((1-\pi_{01})^{n00}\pi_{01}^{n01}(1-\pi_{11})^{n10}\pi_{11}^{n11}\right)\sim X_2^2$, under the null hypothesis of an independence failure process with failure probability $\rho$.

[vii] $T=\tilde{T}+\breve{T}$.

[viii]We set the cut off point to 10% to ensure that the successful models will neither over nor underestimate the *true* VaR and the sequence of violations will be independent. Detailed results for all the models are available upon request.

[ix]The rolling parameters are available upon request.

[x]Hansen and Lunde (2006) and Patton (2005) noted that not all the loss functions rank the volatility forecasting models consistently. Specifically, Hansen and Lunde noted that some loss functions, including the MAE criterion, can be distorted by the substitution of a proxy for the latent population measure of volatility. Hence, if we take under consideration only the MSE loss function, we would add to the appropriate models the EGARCH-N and APARCH-N volatility specifications for the S&P500, as well as the FIGARCH-GED model for the Gold case.