A Robust VaR Model under Different Time Periods and Weighting Schemes

Angelidis, Timotheos and Benos, Alexandros and Degiannakis, Stavros

Department of Financial and Management Engineering, Aegean University, Group Risk Management, National Bank of Greece, Department of Statistics, Athens University of Economics and Business

2007

A Robust VaR Model under Different Time Periods and Weighting Schemes

Angelidis, Timotheos[•]

Benos, Alexandros[*]

Degiannakis, Stavros[⊗]

Abstract

This paper analyses several volatility models by examining their ability to forecast the Value-at-Risk (VaR) for two different time periods and two capitalization weighting schemes. Specifically, VaR is calculated for large and small capitalization stocks, based on Dow Jones (DJ) Euro Stoxx indices and is modeled for long and short trading positions by using non parametric, semi parametric and parametric methods. In order to choose one model among the various forecasting methods, a two-stage backtesting procedure is implemented. In the first stage the unconditional coverage test is used to examine the statistical accuracy of the models. In the second stage a loss function is applied to investigate whether the differences between the models, that calculated accurately the VaR, are statistically significant. Under this framework, the combination of a parametric model with the historical simulation produced robust results across the sample periods, market capitalization schemes, trading positions and confidence levels and therefore there is a risk measure that is reliable.

JEL Nos: C22; C52; C53; G15

Keywords: Asymmetric Power ARCH, Backtesting, Extreme Value Theory, Filtered Historical Simulation, Value-at-Risk.

1. Introduction

[•]Department of Financial and Management Engineering, Aegean University, 31, Fostini Street, 82100 Chios, Greece.

[*]Group Risk Management, National Bank of Greece.

[⊗]Department of Statistics, Athens University of Economics and Business.

Regulatory authorities led financial institutions into calculating Value-at-Risk (VaR) for compliance to market risk capital requirements after the increase of financial uncertainty in the 90's. VaR is a statistic of the dispersion of a distribution and refers to a portfolio's worst outcome likely to occur over a predetermined period and a given confidence level. In order for a risk manager to forecast accurately the VaR he/she must develop a model that accommodates the non-symmetrical fat tails of the empirical distribution. Several methods have been proposed to estimate the risk that the financial institutions face, but until now no model has been deemed as adequate for all financial datasets, sample frequencies, trading positions, confidence levels and sub-periods.

The existing methods of forecasting quantiles of the underlying distribution can be classified in three categories: fully parametric methods that model the entire distribution and the volatility dynamics; non-parametric ones, such as the historical simulation that relies on actual prices and semi-parametric ones, such as filtered historical simulation and extreme value theory, which combine the two previous methods.

Gurmat and Harris (2002) proposed an exponentially weighted likelihood model, as they pointed out that, for three equity portfolios (U.S., U.K. and Japan), it calculated the VaR more accurate than that of the GARCH model under either the normal or the Student-t distributions. Bali and Theodossiou (2006) combined the skewed generalized Student-t distribution with 10 GARCH specifications and argued that the TS-GARCH, proposed by Taylor (1986) and Schwert (1989), and the EGARCH, introduced by Nelson (1991), had the best overall performance, as they accurately estimate both VaR and the Expected Shortfall measure. Giot and Laurent (2003a, 2003b) suggested the APARCH model under a skewed Student-t distribution to researchers in order to forecast the VaR both for long and short trading positions, since the exception rates of the model were too close to the expected ones at all confidence levels. In a similar work, Huang and Lin (2004) also argued that the APARCH model must be used, but they noted that the normal (Student-t) distribution was preferred at the lower (higher) confidence level. Furthermore, Degiannakis (2004) proposed to portfolio managers the fractional integrated APARCH model under the skewed Student-t distribution in order to forecast both the one-day-ahead realized volatility and the daily VaR. Similar to the aforementioned work, Mittnik and Paolella (2000) and Mittnik et al. (2000) recommended more general structures for both the volatility process and the distribution in order to improve the VaR forecasts, while So and Yu (2006) argued that it was more important to model the fat tailed underlying distribution than the fractional integration of the volatility process.

Historical Simulation[1] (HS) is a non-parametric technique that is based only on the empirical distribution of returns. It uses historical returns and derives the VaR number for a specific confidence interval as the corresponding quantile of the empirical historical distribution. It therefore accommodates non-normal distributions and accounts for "fat tails" and non-zero skewness. However, based on this method, there is no consistent method of estimating the volatility innovation.

Filtered Historical Simulation (FHS), which was presented by Hull and White (1998) and Barone-Adesi et al. (1999), tries to fill the gap between the non-parametric and the parametric methods by taking the best part of them. More specifically, it forecasts variance through a parametric volatility model and uses the quantile of the standardized returns in order to calculate the VaR. Barone-Adesi and Giannopoulos (2001) argued the FHS produces risk forecasts that accommodate the current state of the market and therefore it is better than the Historical Simulation. Angelidis and Benos (2006) reached to the same conclusion, claiming that at the higher confidence level the FHS performed better than the parametric and the semi-parametric methods.

Under the same framework, the Extreme Value Theory[2] (EVT), which models only the tails of the distribution, has been suggested as an alternative VaR methodology. Chan and Gray (2006) provided evidence in favour of the EVT since when compared to a number of other parametric models and simple historical simulation based approaches, it produced the appropriate unconditional and conditional coverage. Brooks et al. (2005) suggested to “treat the tails as being distinct from the rest of the distribution and to model them separately but to incorporate information from both”.

To sum up, the choice of an adequate model for volatility forecasting is far from resolved. Our study sheds light on the volatility forecasting methods under a risk management framework, since it juxtaposes the performance of the most well known

[1] For more information on Historical Simulation, see Hendricks (1996), and Vlaar (2000) among others.
[2] For more information on EVT techniques on VaR modelling see Jondeau and Rockinger (1999), Ho et al. (2000), MacNeil and Frey (2000), Rozario (2002), Bali (2003), Jondeau and Rockinger (2003), Seymour and Polakov (2003), Byström (2004) and Gençay and Selçuk (2004).

techniques for different market capitalization schemes, trading positions and sample periods and therefore provides an elaborate exposition of the risk management techniques.

The main contribution of the paper is a thorough stress testing of risk management techniques. We implement several volatility models (parametric, semi-parametric and non-parametric) in order to forecast both long and short VaRs, at two confidence intervals (97.5% and 99%) for two equity indices (DJ Euro Stoxx large and small capitalizations). Models are compared over two different time periods in order to investigate whether the risk management techniques are robust across time. Stress testing provides evidence that the risk measures are reliable and help select a model not affected by the chosen sample period, thereby reducing the painful losses due to the use of an inadequate model (model risk).

Furthermore, in the majority of VaR studies no statistical evaluation of the volatility specifications is implemented and hence no robust model for different schemes (i.e. for long and short trading positions or several confidence levels) is being selected. In order to fill this void, we employ a two-stage backtesting procedure, similar to that of Sarma et al. (2003), and propose a specific volatility-forecasting model for risk management. In the first stage, we assess the statistical accuracy of the models and we specifically investigate whether the total number of failures is statistically equal to the expected one. In the second stage, we implement a forecast evaluation method to examine whether the differences between models (which have converged sufficiently), are statistically significant according to Hansen's (2005) Superior Predictive Ability (SPA) test.

For robustness purposes, we use two different equity indices to avoid the results being dependent on a specific financial market and to examine, at the same time, the effect of stock capitalization to the VaR framework. We look specifically for patterns that could bring some information on how the market capitalization affects the model selection procedure. This is an important issue and has not been examined yet; most financial institutions hold portfolios that contain securities of either large or small capitalization and hence it is interesting to know if they can use the same risk model in all cases.

According to the two-stage backtesting evaluation procedure, a risk manager can select a robust model irrespectively of the testing framework. For both sub-periods, FHS generates accurate VaR numbers for both trading positions and confidence levels, as it captures the characteristics of the empirical distribution more efficiently than parametric methods. It also exhibits in most of the cases superior predictive ability based on Hansen's (2005) SPA test.

The rest of the paper is organized as follows. Section 2 provides a description of the various VaR methods, while section 3 describes the evaluation framework. Section 4 presents preliminary statistics for the dataset and presents the results of the empirical investigation while section 5 concludes.

2. Value-at-Risk

2.1 Parametric VaR

We assume that the data generated process of the log-returns, $y_t = \log(p_t / p_{t-1})$, is an ARCH process with constant mean and unconditional variance but time varying conditional variance, σ_t^2, given the information set available at time $t-1$, I_{t-1}:

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

where p_t is the price of the index at the end of day t, $f(.)$ is the density function of z_t, θ is the vector of the unknown parameters of $f(.)$, $g(.)$ is a time-varying, positive and measurable function of I_{t-1} and w is the vector of the unknown parameters for the conditional variance.

The daily parametric VaR, under the assumption that the conditional mean process is essentially zero, is calculated as:

$$
VaR_{t+1|t} = F_a(z_t; \theta)\sigma_{t+1|t}, \tag{2}
$$

where $F_a(z_t; \theta)$ is the α^{th} quantile of the assumed distribution and $\sigma_{t+1|t}$ is the forecast of the conditional standard deviation at time $t+1$, given the estimated parameters θ and w at time t[3]. Since financial time series usually exhibit skewness and kurtosis different

[3]Under the assumption that the portfolio returns are normally distributed, the calculation of VaR is

from that of the standard normal distribution, a risk manager must make assumptions about the underlying distribution and the conditional variance of innovations in order to accurate calculate the parametric VaR. Bollerslev (1986) introduced the Generalized Autoregressive Conditional Heteroskedastic (GARCH) model:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} b_i \sigma_{t-i}^2, \tag{3}$$

where $a_0 > 0$, $a_i \geq 0$ for $i = 1,\ldots,q$, and $b_i \geq 0$ for $i = 1,\ldots,p$[4]. However, the model does not capture the asymmetry of the financial data and therefore the asymmetric GARCH models were introduced. The most popular model is Nelson's (1991) exponential GARCH (EGARCH) model:

$$\log\left(\sigma_t^2\right) = a_0 + \sum_{i=1}^{q}\left(a_i \left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| + \gamma_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right) + \sum_{i=1}^{p}\left(b_i \log\left(\sigma_{t-i}^2\right)\right). \tag{4}$$

The parameters γ_i capture the asymmetric effect. The threshold GARCH (TARCH) model:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \gamma_1 \varepsilon_{t-1}^2 d_{t-1} + \sum_{i=1}^{p} b_i \sigma_{t-i}^2, \tag{5}$$

allows the volatility to response differently to good or bad news as if $\varepsilon_t < 0$, d_t equals to 1 and d_t equals to 0 otherwise. Finally, as there is no apparent reason why one should assume that the conditional variance is linear function of lagged squared returns[5], Ding et al. (1993) introduced the asymmetric power ARCH (APARCH) model:

$$\sigma_t^{\delta} = a_0 + \sum_{i=1}^{q} a_i \left(\left| \varepsilon_{t-i} \right| - \gamma_i \varepsilon_{t-i} \right)^{\delta} + \sum_{i=1}^{p} b_i \sigma_{t-i}^{\delta}, \tag{6}$$

where $a_0 > 0$, $a_i \geq 0$, $-1 < \gamma_i < 1$, $b_i > 0$ and $\delta > 0$. The APARCH comprises most of the presented models. For example, if $\delta = 2$ and $\gamma_i = 0$, specification (6) is equivalent to the GARCH. Given the fact that the GARCH, EGARCH, TARCH and APARCH models

---

simplified as both $\sigma_{t+1|t}$ and $F_a(z_t;\theta)$ have tractable expressions, while this method will be referred as Variance-Covariance. This method is used as benchmark.

[4]A special case of the GARCH model is the exponentially weighted moving average (EWMA) model, $\sigma_t^2 = 0.06\varepsilon_{t-1}^2 + 0.94\sigma_{t-1}^2$, that was used by RiskMetrics™, when they introduced their analytic VaR methodology.

[5]According to Brooks et al. (2000), "the common use of a squared term is most likely to be a reflection of

are able to capture the thick tailed returns, the volatility clustering and the asymmetry of the data, we limit our analysis to these four families which are the most well known[6].

We turn the discussion to the distributional assumptions of z_t. Engle (1982) introduced the ARCH process under the assumption of normality:

$$f(z_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_t^2}{2}} \tag{7}$$

However, the degree of leptokurtosis induced by the conditional volatility specifications often does not capture all of the leptokurtosis present asset returns, giving evidence that the distribution of z_t is non-normal as well. Bollerslev (1987) proposed the standardized symmetric Student-t distribution with $\nu > 2$ degrees of freedom in order to capture the fat tails of the time series:

$$f(z_t;\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}\left(1+\frac{z_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \tag{8}$$

where $\Gamma(.)$ is the gamma function. However, the density function is symmetric and therefore it is plausible to apply a non-symmetrical distribution to accommodate the non-zero skewness of the financial time series, as Brooks and Persand (2003a), among others, argued that the asymmetry is an important issue in the VaR framework and therefore it must be modeled.

Lambert and Laurent (2001) proposed the standardized skewed Student-t distribution:

$$f(z_t;\xi,\nu) = \begin{cases} \frac{2}{\xi+\xi^{-1}} sf\left(\xi(sz_t+m);\nu\right) & \text{if } z_t < -\frac{m}{s} \\ \frac{2}{\xi+\xi^{-1}} sf\left(\frac{sz_t+m}{\xi};\nu\right) & \text{if } z_t \geq -\frac{m}{s} \end{cases}, \tag{9}$$

where $f(.;\nu)$ is defined in (8), ξ is the asymmetry coefficient, while

---

the normality assumption traditionally invoked regarding financial data".

[6]For more information, see Engle and Patton (2001), Brooks and Persand (2003b), Giot and Laurent (2003a) and Poon and Granger (2003) among others.

$m = \frac{\Gamma\left(\frac{v-1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}(\xi - \xi^{-1})$ and $s^2 = \left(\xi^2 + \xi^{-2} - 1\right) - m^2$ are the mean and the variance of the non-standardized skewed Student-t distribution, respectively. The density is skewed to the right (left) if $\log(\xi) > 0$ $(\log(\xi) < 0)$, while the α^{th}-quantile function, $F_a^*\left(z_t; v, \xi\right)$, of the non-standardized skewed Student-t distribution is calculated as:

$$F_a^*\left(z_t; v, \xi\right) = \begin{cases} \xi^{-1} F_a\left(z_t; v\right)\left(\frac{a}{2}\left(1+\xi^2\right)\right) & \text{if } \alpha < \frac{1}{a+\xi^2} \\ -\xi F_a\left(z_t; v\right)\left(\frac{1-a}{2}\left(1+\xi^2\right)\right) & \text{if } \alpha \geq \frac{1}{a+\xi^2} \end{cases} \tag{10}$$

where $F_a\left(z_t; v\right)$ denotes the α^{th}-quantile function of the standardized symmetric Student-t distribution. Therefore, it is straightforward to estimate the daily VaR, since the quantile function of the standardized skewed Student-t is calculated as:

$$F_a\left(z_t; v, \xi\right) = \frac{F_a^*\left(z_t; v, \xi\right) - m}{s}. \tag{11}$$

2.2 Non-Parametric VaR

VaR based on the HS method is calculated as:

$$VaR_{t+1|t} = F_a\left(\{ y_t \}_{i=1}^{T}\right). \tag{12}$$

Because it relies on actual prices, it accommodates non-normal distributions and therefore accounts for fat tails and skewness. However, this simple approach does not come without a cost, as the choice of the sample length, T, affects the estimates (see for example the work of Van den Goorbergh and Vlaar, 1999).

2.3 Semi-Parametric VaR

The presented methods (parametric and non-parametric) face several drawbacks. For example, a risk manager must make an assumption for the underlying distribution in order to calculate the parametric VaR, while under the framework of the historical simulation technique there is no consistent method of estimating and forecasting the volatility innovation. Hull and White (1998) and Barone-Adesi et al. (1999) combined the two methods in order to lessen the problematic use of the most well known approaches and introduced the VaR estimate based on the FHS method:

$$VaR_{t+1|t} = F_a\left(\{z_{t+1-i|t}\}_{i=1}^{T};\theta\right)\sigma_{t+1|t} \tag{13}$$

where $z_{t+1-i|t} = \varepsilon_{t+1-i|t} / \sigma_{t+1-i|t}$. The combination of a parametric method, such as specifications (3) or (6), with the HS might offer an improvement in the calculation of the VaR, as it accommodates the main characteristics of the empirical distribution (non-zero skewness, fat tails and volatility clustering).

Another method combining the parametric with the non-parametric techniques is EVT, which models only extreme observations and therefore must be seriously considered by risk managers since VaR is only a point estimate of the tails of the distribution[7]. Kuester et al. (2006) extended McNeil and Frey's (2000) EVT framework and argued that a hybrid method, combining a heavy-tailed GARCH filter with an extreme value theory based approach, performed better than the alternative strategies (historical simulation and fully parametric models).

The VaR based on the EVT method is calculated as:

$$VaR_{t+1|t} = \sigma_{t+1|t}\, u \left(\frac{a}{T_u / T}\right)^{-\tau}, \tag{14}$$

where a denotes the VaR confidence level, τ is the Hill estimator of the tail index, T_u is the number of observations beyond the threshold u, which is assumed to be equal to 5% of the total sample size T[8].

3. Evaluation Framework

Kupiec (1995) developed the unconditional coverage test and demonstrated that the proportion of failure[9] follows a binomial distribution. Consequently, the appropriate likelihood ratio statistic, under the null hypothesis that the observed exception frequency equals to the expected one $(N/T' = \rho)$, is given by:

$$LR_{uc} = 2\log\left(\left(1-\frac{N}{T'}\right)^{T'-N}\left(\frac{N}{T'}\right)^{N}\right) - 2\log\left((1-\rho)^{T'-N}\rho^{N}\right) \sim \mathrm{X}_1^2, \tag{15}$$

where N is the number of days over a T' period that the portfolio loss was greater than

[7]The volatility models for the FHS and the EVT are based on quasi-maximum likelihood method assuming a normal distribution, as in Diebold et al. (1998) and McNeil and Frey (2000).

[8]For more information see Balkema and de Hann (1974), Pickands (1975), McNeil and Frey (2000) and Christoffersen (2003) among others.

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

the VaR forecast. The main goal of this test is to examine whether the failure rate of a model is statistically equal to the expected one and therefore to ensure that the financial institution will not misallocate its capital[10].

A backtesting measure neither reflects the specific concerns of the risk managers nor can be used to select a unique volatility technique. An alternative method to evaluate VaR models is the implementation of a loss function, an idea which was first introduced by Lopez (1999), since it is possible to specify a real-world risk manager utility function and to perform a model selection procedure, as in most of the cases there are more than one risk models that are deemed as adequate and hence a risk manager can not select a unique volatility forecasting technique. He suggested measuring the *accuracy* of the VaR forecasts on the basis of the distance between the observed returns and the forecasted VaR values if a violation occurs:

$$\Psi_{t+1} = \begin{cases} \left(y_{t+1} - \text{VaR}_{t+1|t}\right)^2 & \text{if a violation occurs} \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

A preferred VaR model is the one that minimizes the total loss value $\left(\Psi = \sum_{t=1}^{T'} \Psi_t\right)$.

Sarma et al. (2003) and Angelidis et al. (2004) evaluated the most accurate VaR models by implementing a Diebold and Mariano (1995) test in order to select one among the various candidates. However, based on this testing procedure a risk manager cannot juxtapose all the models simultaneously. Therefore, in order to conduct a multiple comparison of the forecasting performance of a benchmark model against its competitors, we apply Hansen's (2005) SPA hypothesis testing.

Let i denote the benchmark model and $\Psi_t^{(i)}$ be the value of the loss function at time t of model i. The null hypothesis that, the benchmark model i is superior to its M competitors is tested with the statistic $T^{SPA} = \max_{i^*=1,\ldots,M} \frac{\sqrt{M}\,\overline{X}_{i^*}}{\sqrt{Var\left(\sqrt{M}\,\overline{X}_{i^*}\right)}}$, where $\overline{X}_{i^*} = T'^{-1} \sum_{t=1}^{T'} X_t^{(i,i^*)}$ and $X_t^{(i,i^*)} = \Psi_t^{(i)} - \Psi_t^{(i^*)}$ for $i^* = 1,\ldots,M$. The p-values of the T^{SPA} statistic are computed according to the bootstrap method of Politis and Romano (1994).

[10]Christoffersen's (1998) conditional coverage test was also implemented, but the results were qualitatively similar and therefore are not reported. They are available upon request.

In case the null hypothesis is rejected, the benchmark model is not superior to its competitors.

4. Empirical Investigation

4.1 Data

To evaluate the volatility models, we generate out-of-sample VaR forecasts for two equity portfolios, large and small capitalization of DJ Euro Stoxx, obtained from DataStream for the period of January 2nd, 1987 to July 29th, 2005. Descriptive statistics for the log-returns of the two indices are presented in Table 1.

4.2 Statistical Evaluation of the VaR models

We split the dataset in two sub-groups, in order to investigate whether the adequacy of the risk management techniques is robust across time. The two samples cover the periods from 2 January 1987 to 18 March 1996 and from 19 March 1996 to 29 July 2005, respectively. Each sub-group contains 2399 trading days. For both sub-groups, we use a recursive sample of 1750 observations, leaving 649 trading days for the out-of-sample evaluation. Each trading day the parameter vector $\psi = (c_0, w, \theta)'$, which denotes the whole set of the unknown parameters for the conditional mean, variance and density function, is re-estimated in order to produce the VaR forecasts both for long and short trading positions[11]. The exception rates and the p-values of the unconditional coverage test are presented for the two sub-samples in Tables 2 and 3, respectively.

The Variance-Covariance (VC) method, especially for the first sub-sample, is not an appropriate technique, as it is rejected by the backtesting measure. In the second sub-sample, however, the VaR forecasts are more accurate, a finding that indicates that this risk measure is not robust across time and therefore not consistent.

Volatility models based on the normal distribution (GARCH, EGARCH, TARCH and APARCH) perform better than VC. In the second sub-group, ARCH models under the normal distribution perform better than in the first one as the short trading positions of the second sub-group were accurately calculated in all the cases but the 97.5% VaR of the large capitalization. In general, for both confidence intervals, these risk management techniques accurate calculate long positions but overestimate VaR for the short ones

[11] If the numerical maximization of the log-likelihood function failed to converge more than six times, we excluded the models from our results. This number of failures (six) is almost the 1% of the out-of-sample.

mainly in the first sub-sample.

ARCH models under the Student-t and the corresponding skewed distribution overestimate VaR in the majority of cases, a result also documented by several studies (see Guermat and Harris 2002 and Billio and Pelizzon 2000 among others). An exception to this general remark is VaR estimates for the large capitalization index, when we investigate the long trading position in the second sub-sample.

The main assumption of the HS method is that samples are identically and independently distributed. It is hence expected that the HS method will underestimate or overestimate the *true* VaR, if the distribution of the future returns changes. This is indeed the case in our empirical research, as the HS method overestimates the total risk, since for all cases exception rates are lower than expected ones.

Generally speaking, the FHS procedure combined with the ARCH volatility specifications offers a major improvement over either the parametric or the non-parametric methods. For long trading positions, both FHS models estimate the *true* VaR accurately. In the case of the FHS with a GARCH updating volatility technique, the exception rates are too close to the theoretical ones in all sub-samples, indices, confidence levels and trading positions. On the other hand, EVT models perform better at the 99% than at the 97.5% confidence level, but overall they do not outperform the FHS models. This finding is in line with the work of Bekiros and Georgoutsos (2005) who argued that the superiority of the EVT based methods emerges at high confidence levels.

To sum up, ARCH models with normally distributed innovations and FHS with an ARCH updating technique describe more efficient the tails of the empirical distribution than their competing techniques. In conclusion, there are models that generate adequate VaR forecasts for specific index, trading position and sub-group but only one technique, the FHS combined with GARCH volatility (FHS-G), produces accurate VaR forecasts for all cases.

4.3 Model Selection

The different VaR models cannot be compared directly, as neither an exception rate close to the expected one nor a high p-value of a model indicates its superiority among its competitors, even if most research has focused on these measures. According however to the approach presented in Section 3, a statistical evaluation of risk management techniques can be achieved. For this purpose, we compute the loss function in (16) and carry out the SPA hypothesis testing, for all models with a p-value greater than 10% in Kupiec's test[12]. Using a high cut-off point for the p-value, we ensure that the *successful* models will accurately estimate the expected rate.

In Table 4, we present p-values of the SPA test for each sub-sample, capitalization, confidence level and trading position. We test the null hypothesis that the FHS-G model (benchmark) is superior to its competitors, as it is the only model that produces adequate VaR forecasts for all cases. For example, for both sub-samples and for the long position at the higher confidence level the null hypothesis is not rejected and therefore the FHS-G outperforms its competitors. On the other hand, for the first sample and for the short position at the lower confidence level, the benchmark model (FHS-G) does not outperform the models that were not rejected by the unconditional coverage test.

Generally speaking, this model selection procedure informs us that the FHS-G model not only provides adequate VaR forecasts, as it has not been rejected by the unconditional coverage test, but also outperforms its competitors in most cases, since the null hypothesis of the SPA test is rejected at 5% level of significance only in four out of sixteen cases. Consequently, the risk manager can use this technique irrespectively of the sample period, stock portfolio, confidence level and trading position.

5. Conclusions

Given the fact that stress testing is now getting much more attention for market risk management purposes, it can help the risk manager to avoid some painful losses, like those that emerged these recent years. In this paper, we employ several volatility models to forecast daily VaR and to specifically compare the models for over two different time periods in order to investigate techniques' robustness across time.

As backtesting tests might not identify a unique best model for each portfolio, we define a utility function to evaluate models that have already met the prerequisite of the correct unconditional coverage. Under this two stage framework, the model which minimized the total loss, was preferred over the remaining ones. We also implemented a test for forecast error differences and we provide statistical inference for the forecasting ability of the models. In most cases, there were significant differences between the

[12]A similar ranking was also made by Brooks and Persand (2003b).

models that satisfied the conditions of the two backtesting tests. Given the fact that we started with 18 models, we however manage to reduce them to a much smaller set.

Finally, we find out that FHS-G is robust across sub-samples, stock portfolio, confidence level and trading position as the average exception rates are too close to the expected ones. According to SPA test, that same model does not outperform its competing models *only* in four out of sixteen cases, although a risk manager can safely use it in all cases.

Acknowledgement

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

References

1. Angelidis, T. and Benos, A. 2006. Value-at-Risk for Greek Stocks. Multinational Finance Journal, forthcoming.
2. Angelidis, T., Benos, A. and Degiannakis, S. 2004. The Use of GARCH Models in VaR Estimation. Statistical Methodology, 1(2), 105-128.
3. Bali, T.G. and Theodossiou, P. 2006. A Conditional-SGT-VaR Approach with Alternative GARCH Models. Annals of Operations Research, forthcoming.
4. Bali, T.G. 2003. An extreme value approach to estimating volatility and value at risk. Journal of Business, 76, 83-108.
5. Balkema, A. and de Hann, L. 1974. Residual lifetime at great age. Annals of Probability, 2, 792-804.
6. Barone-Adesi, G., Giannopoulos, K. and Vosper, L. 1999. VaR without correlations for nonlinear Portfolios. Journal of Futures Markets, 19, 583-602.
7. Barone-Adesi, G. and Giannopoulos, K. 2001. Non-parametric VaR techniques. Myths and realities. Economic Notes by Banca Monte dei Paschi di Siena SpA, 30, 167-181.
8. Bekiros, D.,S. and Georgoutsos, A. D. 2005. Journal of International Financial Markets, Institutions and Mone, 15(3), 209-228.
9. Billio, M. and Pelizzon, L. 2000. Value-at-Risk: A multivariate switching regime approach. Journal of Empirical Finance, 7, 531-554.
10. Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307-327.

11. Bollerslev, T. 1987. A conditional heteroscedastic time series model for speculative prices and rates of return, Review of Economics and Statistics, 69, 542-547.
12. Brooks, C. Clare, A. D., Dalle Molle J. W. and Persand, G. 2005. A comparison of extreme value theory approaches for determining value at risk, Journal of Empirical Finance, 12, 339-52.
13. Brooks, R.D., Faff, R.W. and McKenzie, M.D. 2000. A multi-country study of power ARCH models and national stock market returns. Journal of International Money and Finance, 19, 377-397.
14. Brooks, C. and Persand, G. 2003a. The effect of asymmetries on stock index return Value-at-Risk estimates. Journal of Risk Finance, 29-42.
15. Brooks, C. and Persand, G. 2003b. Volatility forecasting for risk management. Journal of Forecasting, 22, 1-22.
16. Byström, H.N.E. 2004. Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. International Review of Financial Analysis 13 (2), 133-152.
17. Chan, K.F. and Gray, P. 2006. Using extreme value theory to measure value-at-risk for daily electricity spot prices. International Journal of Forecasting, 22 (2), 283-300.
18. Christoffersen, P. 1998. Evaluating interval forecasts. International Economic Review, 39, 841-862.
19. Christoffersen, P. 2003. Elements of Financial Risk Management. Academic Press.
20. Degiannakis, S. 2004. Volatility Forecasting: Evidence from a Fractional Integrated Asymmetric Power ARCH Skewed-t Model. Applied Financial Economics, 14, 1333-1342.
21. Diebold, F.X. and Mariano, R. 1995. Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13(3), 253-263.
22. Diebold, F.X., Schuermann, T. and Stroughair, J.D. 1998. Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management., The Wharton School, University of Pennsylvania, Working paper, 98-10.

23. 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.
24. Engle, R.F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica, 50, 987-1008.
25. Engle, R.F. and Patton, A.J. 2001. What good is a volatility model? Quantitative Finance, 1(2), 237-245.
26. Gençay, R. and Selçuk, F. 2004. Extreme value theory and Value-at-Risk: Relative performance in emerging markets. International Journal of Forecasting, 20 (2), 287-303.
27. Giot, P. and Laurent, S. 2003a. Value-at-Risk for Long and Short Trading Positions. Journal of Applied Econometrics, 18, 641-664.
28. Giot, P. and Laurent, S. 2003b. Market risk in commodity markets: a VaR approach. Energy Economics, 25, 435-457.
29. 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.
30. Hansen, P.R. 2005. A Test for Superior Predictive Ability. Journal of Business and Economic Statistics, 23, 365-380.
31. Hendricks, D. 1996. Evaluation of Value-at-Risk models using historical data. Economic Police Review, 2, 39-70.
32. Ho, L.-C., Burridge, P., Cadle, J. and Theobald, M. 2000. Value-at-Risk: Applying the extreme value approach to Asian markets in the recent financial turmoil. Pacific-Basin, Finance Journal, 88, 249-275.
33. 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.
34. Hull, J. and White, A. 1998. Incorporating volatility updating into the historical simulation method for VaR. Journal of Risk, 1, 5-19.
35. Jondeau, E. and Rockinger, M. 1999. The tail behavior of stock returns: Emerging versus mature markets. HEC-School of Management, Department of Finance, Working Paper.

36. Jondeau, E. and Rockinger, M. 2003. Testing for differences in the tails of stock-market returns. Journal of Empirical Finance, 10, 559-581.
37. 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.
38. Kupiec, P.H. 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives, 3, 73-84.
39. Lambert, P. and Laurent, S. 2001. Modeling financial time series using garch-type models and a skewed student Density. Universite de Liege, Mimeo.
40. Lopez, J.A. 1999. Methods for evaluating Value-at-Risk estimates. Federal Reserve Bank of New York, Economic Policy Review, 2, 3-17.
41. McNeil, A.J. and Frey, R. 2000. Estimation of tail-related risk measures for heteroskedasticity financial time series: An extreme value approach. Journal of Empirical Finance, 7, 271-300.
42. Mittnik, S. snd Paolella, M.S. 2000. Conditional density and value-at-risk prediction of Asian currency exchange rates. Journal of Forecasting 19, 313-333.
43. Mittnik, S. and Paolella, M.S., Rachev, S.T. 2000. Diagnosing and treating the fat tails in financial return data. Journal of Empirical Finance 7, 389-416.
44. Nelson, D. 1991. Conditional heteroskedasticity in asset returns: A new approach, Econometrica, 59, 347-370.
45. Pickands, J. 1975. Statistical inference using extreme order statistics. Annals of Statistic, 3, 119-131.
46. Politis, D.N. and Romano, J.P. 1994. The Stationary Bootstrap. Journal of the American Statistical Association, 89, 1303-1313.
47. Poon, S.-H. and Granger, W.J. 2003. Forecasting volatility in financial markets: A review. Journal of Economic Literature, XLI, 478-539.
48. Rozario, D.R. 2002. Estimating Value at Risk for the electricity market using a technique from extreme value theory. School of Banking and Finance, University of NSW, Working Paper.
49. Sarma, M., Thomas, S. and Shah, A. 2003. Selection of VaR models, Journal of Forecasting, 22(4), 337-358.

50. Schwert, G.W. 1989. Why Does Stock Market Volatility Change Over Time? Journal of Finance 44, 1115-1153.
51. Seymour, A.J., and Polakow, D.A. 2003. A coupling of extreme-value theory and volatility updating with Value-at-Risk Estimation in emerging markets: A south african test. Multinational Finance Journal, 7, 3-23.
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. Taylor, S. 1986. Modeling Financial Time Series. New York: Wiley.
54. Van den Goorbergh, R.W.J. and Vlaar, P. 1999. Value-at-Risk analysis of stock returns. Historical simulation, variance techniques or tail index estimation? Netherlands Central Bank, DNB Staff Reports 40.
55. Vlaar, P. 2000. Value at Risk models for Dutch bond portfolios. Journal of Banking and Finance, 24, 131-154.

Tables and Figures

.

Table 1. Descriptive statistics of the daily log-returns, for the large and the small capitalization (2nd January of 1987 to 29th July of 2005).

	Large	Small
Mean	0.03%	0.01%
Median	0.06%	0.05%
Maximum	6.62%	4.52%
Minimum	-8.07%	-7.67%
Std. Dev.	1.20%	0.83%
Skewness	-0.34	-1.12
Kurtosis	8.27	12.22
Jarque-Bera	5648	18007
Probability	0.00	0.00

Table 2. Exception rates and p-values of Kupiec's unconditional coverage test, at the 97.5% and 99% VaR confidence levels, for large and small capitalization and long and short trading positions, over the period 2 January 1987 to 18 March 1996.

Large Capitalization								
	Long Trading Position				Short Trading Position			
	97.5% Conf.L.		99% Conf.L.		97.5% Conf.L.		99% Conf.L.	
Model	Ex.Rate	Kupiec	Ex.Rate	Kupiec	Ex.Rate	Kupiec	Ex.Rate	Kupiec
VC	1.1%	0.9%	0.5%	12.4%	0.5%	0.00%	0.00%	0.0%
G-N	2.3%	75.5%	1.4%	35.0%	0.9%	0.3%	0.3%	3.8%
E-N	1.4%	4.8%	0.8%	54.0%	0.8%	0.1%	0.2%	0.7%
T-N	2.2%	56.7%	1.1%	84.3%	0.9%	0.3%	0.2%	0.7%
A-N	2.0%	40.1%	1.1%	84.3%	0.9%	0.3%	0.2%	0.7%
G-T	0.9%	0.3%	0.3%	3.8%	0.2%	0.0%	0.0%	0.0%
E-T	0.8%	0.1%	0.3%	3.8%	0.2%	0.0%	0.0%	0.0%
T-T	0.9%	0.3%	0.3%	3.8%	0.2%	0.0%	0.0%	0.0%
A-T	0.8%	0.1%	0.3%	3.8%	0.2%	0.0%	0.0%	0.0%
G-ST	0.8%	0.1%	0.3%	3.8%	0.3%	0.0%	0.2%	0.7%
E-ST	0.8%	0.1%	0.3%	3.8%	0.3%	0.0%	0.0%	0.0%
T-ST	0.8%	0.1%	0.3%	3.8%	0.2%	0.0%	0.2%	0.7%
A-ST	0.8%	0.1%	0.3%	3.8%	0.2%	0.0%	0.0%	0.0%
HS	0.9%	0.3%	0.0%	0.0%	0.8%	0.1%	0.0%	0.0%
FHS-G	1.7%	16.3%	0.8%	54.0%	2.6%	84.7%	0.8%	54.0%
FHS-A	1.8%	26.6%	0.8%	54.0%	1.8%	26.6%	0.2%	0.7%
EVT-G	1.2%	2.2%	0.8%	54.0%	1.7%	16.3%	0.6%	29.0%
EVT-A	1.1%	0.9%	0.6%	29.0%	0.9%	0.3%	0.3%	3.8%

Table 2. Continued. Small Capitalization								
	Long Trading Position				Short Trading Position			
	97.5% Conf.L.		99% Conf.L.		97.5% Conf.L.		99% Conf.L.	
Model	Ex.Rate	Kupiec	Ex.Rate	Kupiec	Ex.Rate	Kupiec	Ex.Rate	Kupiec
VC	0.8%	0.1%	0.3%	3.8%	0.0%	0.0%	0.0%	0.0%
G-N	2.3%	75.5%	1.1%	84.3%	0.5%	0.0%	0.0%	0.0%
E-N	1.8%	26.6%	0.6%	29.0%	1.1%	0.9%	0.3%	3.8%
T-N	2.0%	40.1%	1.1%	84.3%	0.8%	0.1%	0.2%	0.7%
A-N	2.0%	40.1%	1.2%	56.5%	0.6%	0.0%	0.0%	0.0%
G-T	1.1%	0.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
E-T	0.8%	0.1%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
T-T	1.1%	0.9%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
A-T	1.1%	0.9%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
G-ST	0.9%	0.3%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%
E-ST	0.6%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
T-ST	0.9%	0.3%	0.0%	0.0%	0.3%	0.0%	0.0%	0.0%
A-ST	0.5%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
HS	0.6%	0.0%	0.2%	0.7%	0.2%	0.0%	0.0%	0.0%
FHS-G	2.3%	75.5%	0.5%	12.4%	1.8%	26.6%	0.5%	12.4%
FHS-A	2.0%	40.1%	0.6%	29.0%	1.7%	16.3%	0.2%	0.7%
EVT-G	1.1%	0.9%	0.3%	3.8%	0.8%	0.1%	0.3%	3.8%
EVT-A	1.4%	4.8%	0.5%	12.4%	0.8%	0.1%	0.2%	0.7%

The models are successively Variance-Covariance (VC), GARCH under normal distribution (G-N), EGARCH-normal (E-N), TARCH-normal (T-N), APARCH-normal (A-N), GARCH-Student-t (G-T), EGARCH-Student-t (E-T), TARCH-Student-t (T-T), APARCH-Student-t (A-T), GARCH-skewed Student-t (G-ST), EGARCH-skewed Student-t (E-ST), TARCH-skewed Student-t (T-ST), APARCH-skewed Student-t (A-ST), Historical Simulation (HS), Filtered Historical Simulation-GARCH (FHS-G), Filtered Historical Simulation-APARCH (FHS-A), Extreme Value Theory-GARCH (EVT-G), Extreme Value Theory-APARCH (EVT-A).

Table 3. Exception rates and p-values of Kupiec's unconditional coverage test, at the 97.5% and 99% VaR confidence levels, for large and small capitalization and long and short trading positions, over the period 19 March 1996 to 29 July 2005.

Large Capitalization								
	Long Trading Position				Short Trading Position			
	97.5% Conf.L.		99% Conf.L.		97.5% Conf.L.		99% Conf.L.	
Model	Ex.Rate	Kupiec	Ex.Rate	Kupiec	Ex.Rate	Kupiec	Ex.Rate	Kupiec
VC	1.1%	0.9%	0.8%	54.0%	1.4%	4.8%	0.8%	54.0%
G-N	2.3%	75.5%	2.0%	2.4%	1.4%	4.8%	0.8%	54.0%
E-N	2.3%	75.5%	1.4%	35.0%	1.1%	0.9%	0.6%	29.0%
T-N	2.5%	95.5%	1.4%	35.0%	1.2%	2.2%	0.6%	29.0%
A-N	2.5%	95.5%	1.4%	35.0%	1.2%	2.2%	0.6%	29.0%
G-T	2.2%	56.7%	0.8%	54.0%	0.9%	0.3%	0.5%	12.4%
E-T	2.2%	56.7%	1.2%	56.5%	0.9%	0.3%	0.2%	0.7%
T-T	1.8%	26.6%	0.9%	84.5%	0.9%	0.3%	0.2%	0.7%
A-T	2.2%	56.7%	0.8%	54.0%	0.9%	0.3%	0.2%	0.7%
G-ST	2.0%	40.1%	0.8%	54.0%	1.4%	4.8%	0.8%	54.0%
E-ST	1.8%	26.6%	0.6%	29.0%	1.4%	4.8%	0.3%	3.8%
T-ST	2.0%	40.1%	0.6%	29.0%	1.2%	2.2%	0.5%	12.4%
A-ST	2.0%	40.1%	0.6%	29.0%	1.2%	2.2%	0.6%	29.0%
HS	1.1%	0.9%	0.2%	0.7%	1.2%	2.2%	0.5%	12.4%
FHS-G	2.3%	75.5%	0.9%	84.5%	2.0%	40.1%	0.8%	54.0%
FHS-A	2.3%	75.5%	0.8%	54.0%	1.5%	9.2%	0.9%	84.5%
EVT-G	2.0%	40.1%	0.9%	84.5%	0.9%	0.3%	0.8%	54.0%
EVT-A	2.0%	40.1%	0.8%	54.0%	1.1%	0.9%	0.8%	54.0%

Table 3. Continued. Small Capitalization								
	Long Trading Position				Short Trading Position			
	97.5% Conf.L.		99% Conf.L.		97.5% Conf.L.		99% Conf.L.	
Model	Ex.Rate	Kupiec	Ex.Rate	Kupiec	Ex.Rate	Kupiec	Ex.Rate	Kupiec
VC	1.9%	26.6%	0.6%	29.0%	1.2%	2.2%	0.5%	12.4%
G-N	2.9%	49.7%	1.4%	35.0%	2.2%	56.7%	0.5%	12.4%
E-N	3.1%	35.9%	1.2%	56.5%	2.6%	84.7%	1.1%	84.3%
T-N	2.8%	66.1%	1.2%	56.5%	2.2%	56.7%	0.9%	84.5%
A-N	2.9%	49.7%	1.2%	56.5%	2.6%	84.7%	0.8%	54.0%
G-T	1.4%	4.8%	0.9%	84.5%	0.5%	0.0%	0.2%	0.7%
E-T	1.8%	26.6%	0.9%	84.5%	1.2%	2.2%	0.5%	12.4%
T-T	1.4%	4.8%	1.1%	84.3%	0.9%	0.3%	0.0%	0.0%
A-T	-	-	-	-	-	-	-	-
G-ST	1.2%	2.2%	0.5%	12.4%	1.5%	9.2%	0.2%	0.7%
E-ST	1.2%	2.2%	0.6%	29.0%	1.5%	9.2%	0.5%	12.4%
T-ST	1.2%	2.2%	0.8%	54.0%	1.5%	9.2%	0.5%	12.4%
A-ST	-	-	-	-	-	-	-	-
HS	0.6%	0.0%	0.2%	0.7%	1.4%	4.8%	0.5%	12.4%
FHS-G	1.7%	16.3%	1.1%	84.3%	3.5%	10.9%	0.9%	84.5%
FHS-A	2.2%	56.7%	1.1%	84.3%	3.7%	6.8%	0.9%	84.5%
EVT-G	1.2%	2.2%	1.1%	84.3%	2.3%	75.5%	1.2%	56.5%
EVT-A	1.2%	2.2%	1.1%	84.3%	2.5%	95.5%	1.2%	56.5%

The models are successively Variance-Covariance (VC), GARCH under normal distribution (G-N), EGARCH-normal (E-N), TARCH-normal (T-N), APARCH-normal (A-N), GARCH-Student-t (G-T), EGARCH-Student-t (E-T), TARCH-Student-t (T-T), APARCH-Student-t (A-T), GARCH-skewed Student-t (G-ST), EGARCH-skewed Student-t (E-ST), TARCH-skewed Student-t (T-ST), APARCH-skewed Student-t (A-ST), Historical Simulation (HS), Filtered Historical Simulation-GARCH (FHS-G), Filtered Historical Simulation-APARCH (FHS-A), Extreme Value Theory-GARCH (EVT-G), Extreme Value Theory-APARCH (EVT-A).

Table 4. P-values of the SPA test for the null hypothesis that the Filtered Historical Simulation GARCH model is superior to its competitors.

First Sub-sample				
	Long Trading Position		Short Trading Position	
Capitalization	97.5% Conf.L.	99% Conf.L.	97.5% Conf.L.	99% Conf.L.
Large	0.33340	0.13090	0.01490[a]	0.11930
Small	0.15590	0.77440	0.03170[a]	-
Second Sub-sample				
	Long Trading Position		Short Trading Position	
Capitalization	97.5% Conf.L.	99% Conf.L.	97.5% Conf.L.	99% Conf.L.
Large	0.01210[a]	0.14760	-	0.09710
Small	0.08710	0.11080	0.00770[b]	0.09190

[a] Indicates that the null hypothesis is rejected at 5% level of significance.
[b] Indicates that the null hypothesis is rejected at 1% level of significance.
[-] Indicates that there are no competing models.