JEL Classification: C22, G17, Q47
Keywords: value-at-risk, expected shortfall, long memory, GARCH-class models, asymmetries, market risk

Estimation and Performance Assessment of Value-at-Risk and Expected Shortfall Based on Long-Memory GARCH-Class Models*

Chaker ALOUI—King Saud University, Riyadh, Saudi Arabia (cmaloui@ksu.edu.sa), *corresponding author*
Hela BEN HAMIDA—Imam Muhammad Ibn Saud Islamic University, Riyadh, Saudi Arabia (benhamida_hela@yahoo.fr)

Abstract
In this paper, we explore the relevance of asymmetry, long memory and fat tails in modeling and forecasting the conditional volatility and market risk for the Gulf Cooperation Council (GCC) stock markets. Various linear and non-linear long-memory GARCH-class models under three density functions are used to investigate this relevancy. Our results reveal that non-linear GARCH-class models accommodating long memory and asymmetry can better capture the volatility of returns. In particular, we find that some stock returns' behaviors are well described by dual long memory in the mean and the conditional variances. Interestingly, the FIAPARCH volatility model with skewed Student distribution is found to be the best suited for estimating the value at risk and expected shortfall for short and long trading positions. This model outperforms the other competing long-memory GARCH-class models and simple GARCH and EGARCH models. Overall, long-memory, asymmetry, persistence and fat tails are important empirical facts in the GCC markets that should be taken into account when modeling and predicting volatility and assessing total risk. Our findings offer several useful implications for policy regulation, risk assessment and hedging, stock-price forecasting and portfolio asset allocations.

1. Introduction

Originally used by financial institutions for internal risk control and asset management, value-at-risk (VaR) took on greater importance when the Basel Committee recommended its use through the 1996 amendment to the 1988 Basel Accord (Basel Committee on Banking Supervision [BCBS], 1996). In 2006, BCBS refined the regulation inherent to the use of VaR allowing greater flexibility for financial institutions to use their own internal VaR models subject to such models being recognized by the regulator. VaR quantifies the maximum loss for a portfolio of assets under normal market conditions over a given period of time and at a certain confidence level. The expected shortfall (ES) is an alternative tool to VaR that is more sensitive to the shape of the loss distribution in the tail of the distribution. It quantifies the expected value of the loss, provided that a VaR violation has occurred. Within the literature, a variety of increasingly complex models, including both parametric and non-parametric models, are used to estimate the VaR. These models can take into account some major empirical facts of financial asset returns such as

* The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding of this research through the Research Group Project [RGP-VPP-211].

clustering volatility, long memory (hyperbolic decline of the conditional variance), asymmetry, persistence and fat tails. These models have been estimated using various distributions of the asset return (normal, Student, skewed Student, generalized error distribution [GED], exponential generalized beta, and stable Paretian) for different financial and commodity assets, and different confidence levels.[1] Presently, the empirical literature fails to offer a consensus regarding the best model for computing VaR. Specifically, the forecasting ability of each model depends on the sample period, the selected asset and the joint distribution of the asset's return innovations. Moreover, some VaR models provide conflicting results when estimated for short and long trading positions that can be explained by the asymmetry in the asset return distributions (see, among others, Shao *et al.*, 2009). At the same time, the empirical success of long-memory (LM) GARCH-class models to predict financial asset volatility and to estimate VaR and ES has been widely evidenced in the empirical literature.[2] Moreover, related literature offers evidence that long memory GARCH-class models under a non-Gaussian asset return distribution are the most suitable specifications to quantify risk using both VaR and ES. In this vein, Grané and Veiga (2008) reveal that LM GARCH-class models outclass short-memory GARCH specifications though similarly to the major previous papers on VaR, they limit their out-of-sample forecasting assessment to just one trading day.

The need for a consistent instrument for risk assessment and management is the outcome of several significant factors: *firstly*, the great increase of stock market transactions mainly in some emerging and frontier countries which, however, result in higher uncertainty and increased stock market volatility. Indeed, immense capital inflow was recorded during the last decade in emerging stock markets, including those of the Middle East and North Africa (MENA). The net private capital flows to emerging economies rose from USD 582,807 million in 2006 to USD 766,025 million by the end of 2012[3] (Institute of International Finance Report, 2012) mainly because foreign investors expected more profitable opportunities to reduce their portfolio risk through diversification. *Secondly*, some well-known financial disasters have led banks, funds and other financial institutions to narrow their attention to the development and implementation of reliable tools for risk quantification. *Thirdly*, the recent BCBS reports assert that the reputation of VaR was badly damaged during the most recent global financial crisis, which reminded banks, funds and regulators alike that during turbulent periods markets can produce losses far in excess of the maximum amounts forecast by VaR.

In this spirit, the research on stock market risk often uses VaR and ES based on GARCH-class models to evaluate the validity and forecasts of volatility models. A model is said to be best suited for modeling the conditional volatility of stock markets if it provides the most accurate VaR and ES forecasts. Since the dust has not yet settled and academicians are still out on the suitability of volatility models for modeling stock market volatility dynamics, this study will therefore evaluate the accuracy of various linear and non-linear GARCH-class models using different

[1] See, among others, Shao *et al.*, 2009; Aloui and Mabrouk, 2010; Xekalaki and Degiannakis, 2010; Mabrouk and Saadi, 2012; Mabrouk and Aloui, 2010; Degiannakis *et al.*, 2013.
[2] See, among others, Giot and Laurent, 2003, 2004; Tang and Shieh, 2006; Aloui and Mabrouk, 2010; Mabrouk and Saadi, 2012.
[3] Institute of International Finance; capital flow data, available at http://www.iif.com/emr/.

evaluation criteria in addition to VaR and ES. Furthermore, there are only a few works that assess the market risk for GCC countries and they often fall short of adequately characterizing volatility behavior.

More concretely, our study sheds light on the issue of volatility forecasting under the risk management environment and on the evaluation procedure of various risk models. We conduct a comparative analysis of the performance of the most well-known risk management techniques for different GCC stock markets. Explicitly, we ask whether accounting jointly for asymmetry, LM and fat tails in the return distribution as the major empirical facts of stock market volatility provides better estimations of VaR and ES. Empirically, this study estimates VaR and ES by using long-memory GARCH-class frameworks under three density functions (Normal, Student-t, and skewed Student-t). In addition, it compares[4] their forecasting power to standard GARCH and exponential GARCH (EGARCH) models. The underlying idea is to ensure that, taking asymmetry into account, LM and fat tails offers better forecasts of stock market volatility and more accurate estimates of VaR and ES than standard GARCH-class models.

We believe that the results of this study are important for three main reasons. *Firstly*, even though it is not considered the most attractive risk measure, VaR summarizes the risk exposure of the investor in just one number and therefore portfolio managers can interpret it quite easily. *Secondly*, ES is closely associated with VaR; it is a coherent risk measure and therefore its utility in evaluating the risk models can be rewarding. However, most researchers currently assess the models only by calculating the average number of violations. *Thirdly*, even if risk managers hold both long and short trading positions to hedge their portfolios, most of the existing studies have been implemented only in long positions.

This study, to the best of our knowledge, is the first that estimates VaR and ES using daily data for all GCC stock markets, namely Abu Dhabi, Bahrain, Dubai, Kuwait, Saudi Arabia and Qatar and we can therefore infer if these markets share common features in the risk management framework. Thus, we combined the most well-known and concurrent LM GARCH-class parametric to find out which model has the best overall performance. Despite the fact that we did not include all ARCH specifications available in the literature, we narrow our attention to GARCH-class models that captured the most underlying characteristics of the data and those that were already used in similar studies.

The remaining sections of this paper are organized as follows: Section 2 reviews the existing studies. Section 3 exposes the econometric frameworks of the LM GARCH-class models. Section 4 presents the technique for forecasting VaR and ES. Section 5 provides the empirical results and their implications for portfolio management. Section 6 concludes the paper.

2. What Does the Empirical Literature Say?

There is growing literature that addresses the issue of stock market volatility forecasting and risk quantification using VaR and ES. More recent research attempts to combine in one model more than one volatility characteristic (asymmetry effects,

[4] The authors are grateful to an anonymous referee for suggesting this point.

persistence, long-memory, and fat-tails) because modeling stock market volatilities may require the incorporation of several volatility characteristics separately or simultaneously.

For example, Härdle and Mungo (2008) take into consideration both asymmetric effects and LM when computing VaR and ES. They estimate two LM GARCH-class models, namely the Fractional Integrated Power ARCH (FIAPARCH) model and the Hyperbolic GARCH (HYGARCH) model, under different stock return innovation distributions. They point out that models simultaneously considering asymmetry and LM perform better in predicting the one-day-ahead and five-days-ahead VaR for both short and long trading positions. Degiannakis (2004) analyzes the forecasting performance of a few risk models and estimates the one-day-ahead realized volatility and daily VaR. The author shows that the FIAPARCH model under skewed Student-t distribution is able to capture the main empirical facts of stock market volatility. Kasman *et al.* (2010) investigate the existence of dual LM for stock markets operating in the region of Central and Eastern Europe and provide strong evidence of LM in stock returns and volatilities. McMillan *et al.* (2008) compute VaR using a board set of linear and non-linear GARCH-class models for emerging stock markets in the Asia-Pacific region. They conclude that it is worthwhile to take into account both asymmetries and long dependence in stock market volatility in order to offer a more accurate VaR. Considering several LM GARCH-class models, Aloui and Mabrouk (2010) show that taking into account dual LM in the mean process and the conditional volatility of commodity returns offers accurate VaR and ES forecasts for both short and long trading positions. Tang and Shieh (2006) examine the LM features of three stock index futures markets. They estimate the FIGARCH and HYGARCH models under normal, Student and skewed Student densities, and discover that the HYGARCH model with skewed Student distribution performs better. Marzo and Zagalia (2007) compare processes based on normal, Student and GED distributions and show that the EGARCH model delivers the best performance followed by the GARCH-GED model. Dimitrakopoulos *et al.* (2010) compare the performance of VaR and the Extreme Value Theory (ETV) for equity portfolios in sixteen emerging and four developed stock markets. These authors discover that the VaR models turn out to be conservative risk forecasts. Mabrouk and Saadi (2012) assess the performance of FIAPARCH, HYGARCH and FIGARCH models in estimating the one-day-ahead VaR of seven stock markets using Student and skewed Student distributions. They find that the FIAPARCH model, under a skewed Student distribution, outperforms all of the other competing models. Albleib and Pohmeier (2012) suggest a methodology of VaR computation based on the optimal combination that accurately predicts losses. They estimate a board set of GARCH-class models including a simple GARCH model, Risk-Metrics and FIGARCH model and show that for the one-day-ahead VaR forecasts, some familiar distributions such us Student-t distribution, skewed Student and EVT convey better VaR estimates.

In a more recent paper, Žikovic and Filler (2013) provide conflicting conclusions. These authors suggest a new methodology for ranking the performance of VaR and ES models based on a nonparametric test using data covering sixteen developed and emerging stock markets for the pre- and post-crisis periods. Žikovic and Filler (2013) find that for a large number of VaR-based models there is no statistically significant difference. Accordingly, the top performers are the conditional extreme value GARCH model and models based on volatility updating. ES results are similar to VaR results with the models being even more closely matched (Žikovic and Filler, 2013, p. 327). Sethapramote *et al.* (2014) examine the accuracy of VaR estimations in the stock exchange of Thailand using a board set of long-memory GARCH-class frameworks. They conclude that VaR estimates using the FIGARCH model with normal distribution are more accurate than those generated using the short-memory GARCH model.

Evidence in favor of LM, asymmetries and fat-tails on developed stock markets has been largely covered in the literature. Despite the fact that emerging markets in the last two decades have attracted the attention of international investors as a means of higher returns such as through diversification of international portfolio risk, only a few studies have investigated the issue of volatility persistence and VaR-ES forecasts. For instance, Al-Maghyereh and Awartani (2012) show that VaR accuracy is improved when they use a FIAPARCH model under skewed Student distribution for the stock markets of the United Arab Emirates (UAE). Al-Maghyereh and Al-Zoubi (2006) are concerned with MENA stock markets. They estimate VaR using tails distributions of return series using EVT, which allows comparison with the variance-covariance method, and GARCH-class models under different distributions. Theses authors show that MENA stock returns exhibit fat tails and thus EVT is the best-suited approach. In a more recent study, Onour (2010) investigates the extreme downside risk for major oil-producing Middle Eastern countries and examines the impact of the global financial crisis. He estimates VaR and ES under GED distribution and shows that the spillover effect of the global crisis varied from country to country, but the most severely affected market among the group of six markets was the Dubai financial market.

3. Long-Memory GARCH-Class Models

3.1 The Autoregressive Fractionally Integrated Moving Average Model

Granger and Joyeux (1980) considered the Autoregressive Fractionally Integrated Moving Average (ARFIMA) (*p, d, q*) models to test the LM property in financial time series. Formally, the model is written as follows:

$$\Phi(L)(1-L)^d(X_t-\mu)=\theta(L)\varepsilon_t\,,\ \ \varepsilon_t=z_t\sigma, \qquad z_t\sim N(0,1) \tag{1}$$

where $(L)=1-\theta_1 L-\theta_2 L^2-\ldots-\theta_P L^P$, and $\theta(L)=1-\theta_1 L-\theta_2 L^2-\ldots-\theta_P L^P$ are the autoregres-sive (AR) and moving-average (MA) polynomials assumed to have all roots outside the unit circle and ε_t is a white noise process. When $-0.5<d<0.5,$ the X_t process is stationary and invertible and if $0<d<0.5,$ then the process is stationary and exhibits a LM. The ARFIMA model nests to ARMA (i.e. short memory) when $d=0$ and ARIMA when $d=1$ (i.e. infinite LM).

3.2 The Fractional Integrated GARCH Model

Baillie *et al.* (1996) extended the standard GARCH model by considering an eventual fractional integration. They suggested the FIGARCH model, which is able to distinguish between short memory and LM in the conditional variance behavior. Formally, the FIGARCH (*p, d, q*) model is defined as follows:

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

or

$$\begin{aligned}\sigma_t^2&=\omega+\beta(L)\sigma_t^2+\left[1-\beta(L)\right]\varepsilon_t^2-\varphi(L)(1-L)^d\,\varepsilon_t^2\\&=\omega\left[1-L\right]^{-1}+\lambda(L)\varepsilon_t^2\end{aligned} \tag{3}$$

where (L) is the lag operator, $\lambda(L)=\sum_{i=1}^{\infty}\lambda_i L^i$ and $0\le d\le 1$. $\lambda(L)$ is an infinite summation which, in practice, has to be truncated. Following Baillie *et al.* (1996), $\lambda(L)$ should be truncated at 1,000 lags.[5] $(1-L)^d$ is the fractional differencing operator. It can be defined as follows:

$$\begin{aligned}(1-L)^d&=\sum_{k=0}^{\infty}\frac{\Gamma(d+1)L^k}{\Gamma(k+1)\Gamma(d-k+1)}=1-dL-\frac{1}{2}d(1-L)L^2-\frac{1}{6}d(1-d)(2-d)L^3-\ldots\\&=1-\sum_{k=1}^{\infty}c_k(d)L^k\end{aligned} \tag{4}$$

where

$c_1(d)=d, c_2(d)=\frac{1}{2}d(1-d)$, etc.

3.3 The Fractional Integrated Asymmetric Power ARCH Model

Tse (1998) extended the FIGARCH (*p, d, q*) model to take into consideration asymmetry and the long-memory feature in the process of the conditional variance by introducing the function $\left(|\varepsilon_t|-\gamma\varepsilon_t\right)$ of the APARCH process. The FIAPARCH (*p, d, q*) can be written as follows:

$$\sigma_t^2=\omega\left[1-\beta(L)\right]^{-1}+\left\{1-\left[1-\beta(L)\right]^{-1}\rho(L)(1-d)^d\right\}\left(|\varepsilon_t|-\gamma\varepsilon_t\right)^{\delta} \tag{5}$$

where δ,γ and λ are the model parameters. The FIGARCH process can take into account some empirical facts on volatility of financial and commodity asset prices. Specifically, (a) if $0<d<1$, then volatility exhibits the long-memory property; (b) when $\gamma>0$, negative shocks have a greater impact on volatility than positive shocks and inversely; (c) $\lambda>0$ is the power term in the volatility structure and it should be specified by the data; (d) the FIGARCH process also nests the FIGARCH process when $\gamma=0$ and $\delta=2$. Therefore, the FIAPARCH process is superior to the FIGARCH because it takes into account asymmetry and LM in the conditional variance behavior.

[5] According to Baillie *et al.* (1996), the implementation of the FIGARCH model necessitates conditioning on pre-sample values and a truncation of the infinite lag polynomial in equation (3). Given the long memory and relatively slow decay of the response to lagged squared innovation, the effect of pre-sample values might be expected to have a bigger impact than with a stationary GARCH process. Based on the above argument, Baille *et al.* (1996) state that truncating at too low a lag may destroy important long-run dependencies. To mitigate these effects, they recommend that the truncating lag be set at 1,000 (Baillie, Bollerslev and Mikkelsen, 1996, p.13)

3.4 The Error's Density Models

The parameters of the volatility models can be estimated by using non-linear optimization procedures to maximize the logarithm of the Gaussian likelihood function. Under the assumption that the random variable is $z_t \sim N(0,1)$, the log likelihood of Gaussian or normal distribution ( ) Norm L can be expressed as:

$$L_{Norm=-\frac{1}{2}} \sum_{t=1}^{T} \left[\ln(2\pi) + \ln\left(\sigma_t^2\right) + z_t^2 \right]$$

where T is the number of observations. However, it is widely recognized that residuals suffer from excess kurtosis. To account for fat tails in the stock return distribution, we consider the Student distribution. If the random variable is $z_t \sim ST(0,1,\nu)$, the log-likelihood function of the Student distribution (L_{Stud}) is to be written as:

$$\begin{aligned} L_{Stud} &= T\left\{ ln\Gamma\left(\frac{\upsilon+1}{2}\right) - ln\Gamma\left(\frac{\upsilon}{2}\right) - \frac{1}{2} ln\left[\pi(\nu-2)\right] \right\} - \\ &-\frac{1}{2}\sum_{t=1}^{T}\left[\ln\left(\sigma_t^2\right) + (1+\nu)\left[ln\left(1+\frac{z_t^2}{\sigma_t^2(\nu-2)}\right)\right]\right] \end{aligned} \tag{6}$$

where $2 < \nu \le \infty$ and $\Gamma(.)$ is the gamma function. In contrast to the normal distribution, the Student distribution is estimated with an additional parameter ν, which stands for the number of degrees of freedom measuring the degree of fat tails in the density. Despite accounting for tail thickness, a Student distribution alone cannot capture the asymmetric feature of density. To account for excess skewness and kurtosis, we consider a skewed Student distribution proposed by Lambert and Laurent (2001). If $z_t \sim Skst(0,1,k,\nu)$, the log likelihood of the skewed Student distribution $\left(L_{Skst}\right)$ is as follows:

$$\begin{aligned} L_{Skst} &= T\left\{ \ln\Gamma\left(\frac{\upsilon+1}{2}\right) - \ln\Gamma\left(\frac{\upsilon}{2}\right) - \frac{1}{2}\ln\left[\pi(\nu-2)\right] + \ln\left(\frac{2}{k+\frac{1}{k}}\right) + \ln(s) \right\} - \\ &-\frac{1}{2}\sum_{t=1}^{T}\left[\ln\left(\sigma_t^2\right) + (1+\nu)\ln\left[1+\left(1+\frac{(sz_t+m)^2}{(\nu-2)}\right)k^{-2I_t}\right]\right] \end{aligned} \tag{7}$$

where $I_t = 1$ if $z_t \ge \frac{m}{s}$ or $I_t = -1$ if $z_t < \frac{m}{s}$, k is an asymmetry parameter. The constants $m = m(k,\nu)$ and $s = \sqrt{s^2(k,\nu)}$ are the mean and standard deviations of the skewed Student distribution:

$$m(k,\nu) = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(k-\frac{1}{k}\right) \tag{8}$$

$$s^2(k,\nu) = \left(k^2 + \frac{1}{k^2} - 1 \right) - m^2 \tag{9}$$

The value of $\ln(k)$ can also represent the degree of asymmetry in the residual distribution. We note that when $\ln(k) = 0$, the skewed Student distribution equals the general Student distribution, $z_t \sim ST(0,1,\nu)$.

3.5 Computing one Step-Ahead VaR and ES under LM GARCH-Class Models

For a 95% confidence level, the one-step-ahead VaR is estimated as follows:

$$VaR_{t+1|t}^{(1-\rho)} = \mu_{t+1|t} + N(\alpha)\sigma_{t+1|t} \tag{10}$$

where $1 - r = 95\%$, $m_{t+1|t}$ and $\sigma_{t+1|t}$ are the conditional forecasts of the mean and the standard deviation at time (t + 1), given the information set available at time (t), respectively. $N(\alpha)$ is the α-th quantile of the normal distribution. In order to back-test the accuracy for the estimated VaRs, we compute the empirical failure rates for both short and long trading positions. The prescribed probability ranges from 0.25% to 5%. By definition, the failure rate is the number of times returns exceed (in absolute value) the forecasted VaR. If the model is "correctly" specified, then the failure rate should be equal to the specified VaR level. Artzner *et al.* (1997) introduce the concept of ES to overcome the main shortcomings of VaR. Shortly, the ES provides the expected value of the loss, given that a VaR violation occurred. We compute the ES as follows: first, we divide the tail of the probability distribution of returns into 5,000 slices, each with identical probability mass, and then we calculate the VaR inherent to each slice and estimate the mean of these VaRs to compute the ES:

$$ES_{t+1|t}^{(1-\varphi)} = E\left(y_{t+1} \middle| \left(VaR_{t+1|t}^{(1-\varphi)} \right) \right) \tag{11}$$

According to Angelidis and Degiannakis (2007), it is possible to measure the squared difference of the losses using the ES given that the VaR does not provide any information regarding the size of the expected loss. They suggest Ψ_{t+1} as a comparison between the actual return and the expected return when a VaR violation occurred. In this way, the best model corresponds to the smallest mean squared error:

$$MSE = \sum_{t=0}^{T-1} \Psi_{t+1} / T \tag{12}$$

$$VaR_{t+\tau|t}^{95\%} = f_{5\%}\left(\left\{ y_{i,t+\tau} \right\}_{i=1}^{5000} \right)$$

The collective accuracy of the VaR figures produced for each of the non-overlapping intervals is then statistically tested using several accuracy forecasting tests.

3.6 Statistical Accuracy of Model-Based VaRs

We employ three alternative tests, namely the Kupiec (1995) test, dynamic quantile test (DQT) suggested by Engle and Manganelli (2004), and the expected loss of VaR forecasts proposed by Giacomini and Kamunjer (2005). Our choice of these alternative tests is motivated by the fact that the common tests of Kupiec and of Engle and Manganelli (2002) exhibit low power when assessing VaR accuracy (Berkowitz *et al.*, 2011). The underlying idea of the Kupiec (1995) test is to estimate the probability of observing a loss greater than the VaR amount. To evaluate the accuracy of the model-based VaR estimates, Kupiec (1995) suggests a likelihood ratio test $\left(LR_{UC}\right)$ for testing whether the failure rate of the model is statistically equal to the expected one (unconditional coverage). Consider that $N=\sum_{t=1}^{T} I_t$ is the number of exceptions in the sample size *T*. Then:

$$I_{t+1} = \begin{cases} 1, & \text{if } r_{t+1} < VaR_{t+1|t(\alpha)} \\ 0, & \text{if } r_{t+1} \geq VaR_{t+1|t(\alpha)} \end{cases} \tag{13}$$

follows a binomial distribution, $N \sim B(T,\alpha)$. If $p = E\left(\frac{N}{T}\right)$ is the expected exception frequency (i.e. the expected ratio of violations), then the hypothesis for testing whether the failure rate of the model is equal to the expected one is expressed as follows: $H_0 : \alpha = \alpha_0$. α_0 is the prescribed VaR level. The appropriate likelihood ratio statistic in the presence of the null hypothesis is given by:

$$LR_{UC} = -2\log\left\{\alpha_0^N \left(1-\alpha_0\right)^{T-N}\right\} + 2\log\left\{\left(\frac{N}{T}\right)^N \left(1-\left(\frac{N}{T}\right)^{T-N}\right)\right\} \tag{14}$$

Under the null hypothesis, LR_{uc} has a $\chi^2(1)$ as an asymptotical distribution. Consequently, the preferred model for VaR prediction should display the property that the unconditional coverage measured by $p = E\left(\frac{N}{T}\right)$ equals the desired coverage level p_0. In addition to the Kupiec LR test, we use the DQT suggested by Engle and Manganelli (2002). The DQT is based on a sequence of VaR violations that is not serially correlated. Formally, considering two new variables $Hit_t\left(\alpha\right) = I\left(y_t < VaR_{t(\alpha)}\right) - \alpha$ and $Hit_t\left(1-\alpha\right) = I\left(y_t > VaR_{t(1-\alpha)}\right) - \alpha$, Engle and Manganelli (2002) suggest jointly testing that:

- A1: $E\left(Hit_t\left(\alpha\right)\right) = 0$ (respectively, $E\left(Hit_t\left(1-\alpha\right)\right)$ for long trading positions (short trading positions);
- A2: $Hit_t\left(\alpha\right)$ $\left(\text{or } Hit_t\left(1-\alpha\right)\right)$ is uncorrelated with the variables included in the information set.

Engle and Manganelli (2002), suggest the following artificial regression in order to test A1-A2, $Hit_t = X\lambda + \varepsilon_t$; where, X is a $T \times k$ matrix whose first column consists of onesand the next q columns are $Hit_{t-1,...Hit_{t-q}}$. $k-q-1$ and the remaining columns are additional independent variables including the VaR. For Engle and Manganelli (2002), the DQT is given by: $\frac{X'X\hat{\lambda}}{\alpha(1-\alpha)}$ where $\hat{\lambda}$ is the OLS estimates of λ and it follows a $\chi^2(k)$ distribution. However, a significant part of the literature on testing the performance of VaR and ES shows that the tests of Kupiec and Christofersen have low power[6,7] thus if we do not reject the null with these tests, we cannot be impressed by the results, as it is very hard to distinguish between "good" and "bad" forecasts. Therefore, we use for our in-sample and out-of-sample VaR testing an alternative approach originally suggested by Engle and Manganelli (2004). Formally, Engle and Manganelli (2004) use the n-th order autoregression:

$$I_t = \omega + \sum_{k=1}^{n} \beta I_{t-k} + \sum_{k=1}^{n} \beta q_{t-k+1}^{\alpha} + \mu_t \tag{15}$$

where I_{t+1} is 1 if $y_{t+1} < q_t^{\alpha}$ and zero otherwise, while hit sequence I_t is a binary sequence. μ_t is assumed to follow a logistic distribution and it is possible to estimate it as a simple logit model and test whether $Pr[I_t = 1] = q_t^{\alpha}$. Following Avdulaj and Barunik (2013), we rely on simulations suggested by Berkowitz *et al.* (2011) to obtain the *p*-values of this test. In addition, we evaluate the accuracy of the VaR forecasts (i.e. out-of sample forecasting) statistically by implementing the Giacomini and Kamunjer (2005) testing methodology. Briefly, the Giacomini and Kamunjer (2005) define the expected loss made by the forecasting model (m) as follows:

$$L_{\alpha,m} = E\left[\alpha - 1\left(y_{t,t+1} < q_{t,t+1}^{\alpha,m}\right)\right]\left[y_{t,t+1} - q_{t,t+1}^{\alpha,m}\right] \tag{16}$$

The differences in the values of $\left(L_{\alpha,m}\right)$ can then be tested using the Diebold and Mariano (2002) approach, where we test the null hypothesis that the loss function of a benchmark forecaster is the same as the loss function of the selected forecasting model (*m*), under the alternative assumption that the benchmark model is more accurate than the competing one. Since we are concerned with the relevancy of GARCH-class models accounting for asymmetry, LM and fat tails in producing accurate VaR and ES, we select the GARCH model under normal distribution as the benchmark. The competing forecasting model varies across GCC countries.

4. Data, Preliminary Analysis and GARCH-Class Model Estimates

4.1 Descriptive Statistics

The dataset contains seven stock markets, namely those of Saudi Arabia (Saudi Arabia Stock Exchange Index: SASEIDX), Dubai (Dubai Financial Market General Index: DFMGI Index), Abu Dhabi (Abu Dhabi Stock Market Index: ADSMI

[6] We are grateful to an anonymous reviewer for pointing this out and suggesting appropriate tests.

[7] For detailed discussion, see Berkowitz *et al.*, 2011.

Index), Oman (Muscat Stock Market Index: MSM30), Bahrain (Bahrain Stock Exchange All Shares Index: BHSEASI), Kuwait (Kuwait Stock Exchange Index: KWSEIDX) and Qatar (Doha Stock Market Index: DSM Index). Data consist of Bloomberg daily stock indices covering the period (January 3, 2003–January 22, 2013) and totaling 2,620 observations. Stock indices are expressed in local currency. In addition, we use the BGCC index established by Bloomberg as a general index for the GCC stock markets for the entire sample period. We should note that we reserve the last 1,000 observations for the out-of-sample forecasts.

The time-variations of the GCC stock market indices are displayed in *Figure A1* (see *Appendix 1* online). We observe that the GCC indices experienced quite similar trends with some very large swings within the sample period. By taking a look at *Figure A1*, we see that the GCC stock markets have experienced two major break-downs in the course of the ongoing economic and financial crisis from which they have recovered moderately so far. The GCC markets responded sharply to the global financial crisis in the US and Europe since mid-2008 as shown by the drastic drop in all of the GCC indices, reaching their peak in the first quarter of 2009. At that level, the GCC stock market indices had—compared with their levels at the beginning of 2007—fallen by one-fifth in the case of Oman, around one-third in Bahrain, Kuwait and Abu Dhabi, almost 50% in Saudi Arabia, and as much as two-thirds in Dubai.

Table A1 reports (see *Appendix 2* online) some the descriptive statistics for the GCC daily returns as well as their stochastic properties. Panel A shows that the highest average return is recorded for the Qatari market (4.9%) while the lowest average return is for Bahrain (-1.5%). The Dubai and Saudi Arabian markets exhibit a higher level of risk as measured by the standard deviation of daily returns. Further-more, all of the GCC daily returns display clustering volatility and a tendency for large (small) price changes to be followed by other large (small) price changes of either sign and they tend to be time dependent (see *Figure A2* in *Appendix 1* online). Moreover, all of the daily returns display significant negative skewness and we ambiguously reject the null hypothesis of skewness coefficients conforming to the normal distribution value of zero. In addition, all of the return distributions are leptokurtic since they exhibit thicker (fatter) tails than that of a Gaussian distribution. The Jarque-Bera test for the unconditional Gaussian distribution confirms these results and shows a significant departure from normality. The Engle (1982) test for conditional heteroscedasticity and the Ljung-Box test provide strong evidence of ARCH effects in the returns and serial correlations in the squared returns (i.e. volatility), respectively. Panel B relates the results of the Augmented Dickey-Fuller (1979) unit root tests, the Schmidt and Phillips (SP) stationarity test and Zivot-Andrews (1992) unit root test to examine the stationarity property of GCC stock returns. The Zivot-Andrews (1992) unit root test examines the null of the unit root hypothesis against the break-stationary alternative and it is therefore robust to the presence of potential structural breaks in the time variations of GCC returns. The results from these tests reveal that all of the GCC daily returns are stationary and therefore suitable for further analysis.

4.2 LM Tests

In the present study, the dual LM in the conditional mean and variance of stock returns and squared returns is investigated by implanting two tests commonly used in the existing literature: the log-periodogram regression (GPH) of Geweke and Porter-Hudak (1983) and the Gaussian semi-parametric (GSP) of Robinson (1995).[8] It is worth noting that the choice of these various tests is justified by the fact that several scholars cautioned against using Lo's (1991) modified R/S in isolation. Specifically, Monte Carlo experiments conducted by Teverovsky *et al.* (1999) and Willinger *et al.* (1999) on Lo's (1991) statistic reveal that the modified R/S statistic is founded in favor of accepting the null of no LM as the bandwidth increases. *Table A2* (see *Appendix 4* online) reports the results of these tests. We employ the following two bandwidths for the GPH test: $m = T^{0.5}$ and $m = T^{0.6}$ For the GSP test, we use $m = T/4$; and $m = T/8$. From the reported results, we find mixed LM evidence for the returns as both LM tests fail to reject the null hypothesis of no long-range dependence for two markets at the 1% level. More precisely, the Saudi Arabian and Bahraini markets display LM properties for their returns. For the GCC squared returns, the two LM tests conclude favorably on the presence of a LM component at the 1% significance level. As a result, all of the GCC stock market volatilities exhibit LM behavior. Taken as a whole, the results of LM tests indicate the suitability of the GARCH/ARCH models accommodating the LM feature, as far as the modeling and forecasting of stock market return volatility are concerned.

4.3 LM GARCH-Class Estimates

We now discuss the estimation results for the GARCH-class models for the seven GCC markets. For each stock return, our conditional mean equation includes a constant and an autoregressive term, while its conditional variance is modeled by seven competing GARCH-class models taking into consideration the main stylized empirical facts of stock returns. To investigate the relevancy[9] of LM GARCH-class models in modeling conditional volatility and market risk, these models are comparted to other competing standard GARCH and EGARCH models which do not take into account long-memory and fat-tailed returns. As noted earlier, the selected volatility models are estimated by using the Student-t, skewed Student and Quasi-maximum Likelihood (QML) methods. The parameter estimates as well as the results of the diagnostic tests of standardized residuals and squared residuals are displayed in *Tables A3a–A3d* (see *Appendix 5* online).

For all of the GCC stock returns, our results reveal slight evidence of return predictability at the 10% level when the GARCH and EGARCH models are estimated. Regarding the estimates of the conditional variances, we note that the stationarity condition is guaranteed given that the sum of the ARCH and the GARCH coefficients is less than one for all of the selected markets. These coefficients are statistically significant at the 1% level. More importantly, the estimated coefficients for the GARCH and EGARCH models are generally higher than 0.9, indicating that GCC stock market volatilities exhibit a high level of persistence over time. When the LM of stock return volatility is effectively taken into account via the FIAPARCH and FIGARCH specifications, the values of the GARCH parameters in the GARCH and EGARCH frameworks decline noticeably. With regard to the EGARCH estimates,

[8] The log-periodogram regression (GPH) of Geweke and Porter-Hudak (1983) and the Gaussian semi-parametric (GSP) of Robinson (1995) are detailed in *Appendix 3* online.

[9] We are grateful to an anonymous referee for raising this point.

the results indicate that the coefficient EGARCH (δ) is significant and negative for all of the GCC stock markets. This result implies that the negative return shock has a significant and greater impact on the conditional volatility than do positive return shocks of equal size.

The results of the LM GARCH-class models (FIGARCH and FIAPARCH) show that only two markets (Saudi Arabia and Bahrain) exhibit LM properties in both the mean return and the conditional variance. For the Saudi Arabian market, the long dependence parameter in the conditional mean ranges between 0.21 and 0.27 when the FIGARCH and FIAPARCH specifications are estimated, respectively. Similarly, the long dependence coefficient ranges between 0.140 and 0.146 for Bahrain. For these countries, the fractional integration parameter in the conditional variance is statistically significant at the 1% level regardless of which nonlinear FIGARCH or FIAPARCH model is implemented and which type of density distribution is used (i.e. normal, Student or skewed Student). For Saudi Arabia (Bahrain), the LM parameter equals 0.34 (0.37) (for the FIAPARCH model) and 0.52 (0.39) (for the FIGARCH model), respectively. More interestingly, the conditional volatility of the Saudi and Bahraini returns reacts asymmetrically to unexpected news in view of the APARCH (γ) being positive and significant. Thus, negative shocks have a more sizable impact on conditional volatility than do positive shocks, which is entirely consistent with the EGARCH implications. The APARCH power term (δ) is positive and significant for the two countries.

Furthermore, for the two selected models, the Student-*df* parameter is positive and thus shows strong evidence of fat tails in stock return volatility. The Log(L), Akaike Information (AIC) and Shibata and Hanan-Quinn (SHIC) criteria we use to rank the various GARCH-class models unanimously select the ARFIMA-FIAPARCH model under skewed Student-t distribution for both the Saudi and the Bahraini returns, indicating the important role of asymmetry effects and LM as the main empirical stylized facts in stock market volatility. Furthermore, the Box-Pierce statistics for the remaining serial correlation for the squared standardized residuals and the residual-based diagnostic test accept the ARFIMA-FIAPARCH as a correct specification.

For the rest of the GCC countries the estimation results of the GARCH-class models are quite similar. They indicate no LM in the mean equation of returns and are well described by an AR(1) process.[10] This result is consistent with what we have learned from the LM tests. The coefficients associated with the ARCH and GARCH terms are highly significant regardless of the density distribution used, suggesting strong support of ARCH and GARCH effects. In addition, the condition for existence of the conditional variance is guaranteed since $\alpha + \beta < 1$. For the EGARCH model, the "leverage" parameter is negatively signed and significant. Moreover, the APARCH (δ) is significant and positive and shows the existence of asymmetry. The fractional difference or LM parameters (d) in the conditional variance are all positive and significant at the 1% level. These findings thus indicate strong evidence of persistence

[10] We note that we checked whether our results are sensitive to the choice of the mean equation; we replicate the same analysis for $AR(k), k: 1,2,3,4,5\ldots10$ for all of the GCC time series. Our unreported results show that the AR-FIGARCH and the ARFIMA-FIAPARCH models under skewed Student-t innovation distribution provide a better fit than all of the competing models.

behavior in the GCC stock market volatilities. However, we note some disparities among markets. Indeed, the LM parameter is lower than 0.5 for Kuwait, Bahrain and Dubai and it is greater than 0.5 for Abu Dhabi and Qatar. More interestingly, the various information criteria used to rank the various GARCH-class models consistently select the AR-FIAPARCH model under skewed Student distribution for all of the GCC countries.

5. Forecasting Performance

5.1 The Evaluation Criteria

As mentioned above, the full sample period provides 2,620 daily observations. We estimate the various long-memory GARCH-class models over the first 1,620 trading days for each GCC market. The parameter estimates generated by the time series from the in-sample period are included in the relevant forecasting formulas and the volatility forecast $\hat{h}_{t+1}$ is computed given the information available at time $t = T\left(=1{,}620\right),\ldots,T+1{,}000\ (=2{,}620)$ (i.e. 1,000 one-step-ahead forecasts are computed). On the other side, before evaluating the forecasting performance of the considered GARCH-class models, it is necessary to have a valid proxy for the true but unobservable stock market volatility. We employ the realized volatility. Accordingly, the squared returns which are assimilated into the natural candidate proxy of the unobservable volatility are considered as a noisy proxy (see Patton, 2010). To assess the forecasting performance of the competing models, we use several criteria such as the mean square error (MSE), the mean absolute prediction error (MAPE), the logarithmic loss function (LL), the Mincer-Zarnowitz (1969) regression and the superior predictive accuracy (SPA) test of Hansen (2005).

5.1.1 The Mincer-Zarnowitz (1969) Regression Test

For the Mincer-Zarnowitz (1969) regression, we recall that this test procedure is based on the regression of the realized volatility h_{t+1} on the constant and forecast volatility $\hat{h}_{t+1}$. Formally, we have:

$$h_{t+1} = \alpha + \beta \hat{h}_{t+1} + \varepsilon_t \tag{17}$$

For an unbiased model, we should verify that the estimated parameters α and β are respectively equal to zero and one.[11] The coefficient of determination is considered as a measure of the predictive power for each model. Consequently, for the model with the largest R^2, the realized volatility is well explained by the forecasted volatility. Despite the fact that the slope coefficient of the Mincer-Zarnowitz regression might be biased, R^2 provides an imminent indicator of the variability in predictions (Andersen and Bollerslev, 1998, p. 890).

5.1.2 The Superior Predictive Accuracy (SPA) Test

To assess the predictive ability of the estimated GARCH-class models, we use Hansen's (2005) SPA test, which is a more efficient test to ensure the significance of the superiority of models. The SPA test examines whether the null hypothesis that

[11] In *Table 1*, we report the Ordinary Least Squares estimates of the parameters in the Mincer-Zarnowitz regression (equation 15). We use the Newey-West correction to calculate the standard errors.

the benchmark model is not outperformed by any of its competitors is rejected or not.[12] Formally, the forecasting performance of the benchmark model,[13] model 0, with respect to k competing models is deduced from the loss function differential $f_{t,k} = I_{0,t} - I_{k,t}$ where $k = 1,\dots,j$ is the total number of competing models. Under the null hypothesis and assuming stationarity for $f_{t,k}$, we expect that on average the forecasting loss function on the benchmark model will be smaller or at least equal to that of model k. The null hypothesis is stated as:

$$T_n^{SPA} = \max_{k=1,..,t} \frac{\sqrt{n} f_k}{\sqrt{\operatorname{var}\left(\sqrt{n} f_k\right)}} \tag{18}$$

where $\overline{f}_k = \frac{1}{n}\sum_{t=1}^{n} f_k$ and $\operatorname{var}\left(\sqrt{n}\,\overline{f}_k\right)$ is the variance of $\sqrt{n}\,\overline{f}_k$. Both $\operatorname{var}\left(\sqrt{n}\,\overline{f}_k\right)$ and the test statistic p-values are consistently estimated via stationary bootstrapping. The last column of *Table 1* reports the results of the SPA test. For each GCC market, models are alternatively used as the benchmark model and the null hypothesis that it is not outperformed by any of it counterparts is tested.

5.2 The Forecasting Performance Results

Table 1 reports the obtained results, while the boldface numbers indicate the best model in terms of volatility forecast accuracy. We can see that under the skewed Student distribution, the ARFIMA-FIAPARCH model provides the best fit for Saudi Arabia and Bahrain. Indeed, the Mincer-Zarnowitz (1969) regression's results show that this dual-LM GARCH-model outperforms the other competing models including standard GARCH and EGARCH models. The coefficient of determination measuring the predictive power for each model indicates that the largest coefficients correspond to the ARFIMA-FIAPARCH model. This result is consistent with the MSE and MAPE criteria. Additionally, the p-values of Hansen's (2005) SPA test indicate that the null hypothesis that this model is not outperformed by any of its counterparts under Student distribution is rejected.

For the other GCC markets, we perceive that FIAPARCH with skewed distribution is the best specification. In fact, the various forecasting measures indicate that this model displays greater forecasting accuracy than the other competing models. Furthermore, the obtained results show that the forecasting accuracy of the FIAPARCH with skewed Student distribution is significantly higher compared to its counterpart under Student and normal distributions. Indeed, the MAE, MAPE and SPA test never select the standard GARCH and EGARCH models. Altogether, the non-linear GARCH-class models, which are able to take into account the main empirical facts of GCC markets returns (asymmetry and/or LM) display greater forecasting accuracy than the linear ones. Furthermore, our results reveal that the FIGARCH model generates

[12] In the SPA test, we evaluated the forecasts using the predefined loss functions. We use the SPA test of Hansen (2005) that incorporates two predefined loss functions. In our case, we refer to the MAPE and MSE.

[13] Here the competing model changes across countries. For example, for Saudi Arabia and Oman, our previous analysis reveals that the ARFIMA-FIAPARCH model with skewed Student distribution outperforms all the other models.

Table 1 The In-Sample Forecasting Performance of the GARCH-Class Models

	α	β	R^2	MSE	MAPE	LL	SPA test (p-values)
				Saudi Arabia			
ARFIMA-FIGARCH							
Student	0.44 (0.62)	0.77 (0.22)	0.07	0.272	177.6	6.69	0.01
Skewed St.	0.68 (0.42)	0.82 (0.06)	0.19	0.250	159.3	5.12	0.00
Normal	0.55 (0.06)	0.81 (0.56)	0.13	0.240	166.8	5.78	0.00
ARFIMA-FIAPARCH							
Student	0.01 (0.15)	0.78 (0.06)	0.16	0.236	131.23	4.78	0.00
Skewed St.	0.09 (0.28)	0.79 (0.75)	**0.21**	**0.214**	**129.89**	**4.65**	**0.47**
Normal	0.53 (0.05)	0.80 (0.22)	0.09	0.263	136.9	5.68	0.00
GARCH-Normal	0.07 (0.08)	0.66 (0.06)	0.17	0.289	133.06	5.13	0.01
EGARCH-Normal	0.12 (0.16)	0.68 (0.06)	0.19	0.241	131.05	5.08	0.00
				Abu Dhabi			
AR-FIGARCH							
Student	0.42 (0.11)	1.02 (0.16)	0.53	0.443	165.5	6.78	0.00
Skewed St.	0.11 (0.22)	1.01 (0.25)	0.52	0.426	147.4	6.12	0.00
Normal	0.22 (0.85)	0.98 (0.19)	0.49	0.421	150.2	6.22	0.00
AR-FIAPARCH							
Student	0.67 (0.06)	0.79 (0.26)	0.31	0.445	143.5	6.08	0.00
Skewed St.	0.77 (0.21)	0.87 (0.33)	**0.54**	**0.421**	**132.9**	**4.97**	**0.63**
Normal	0.63 (0.21)	0.96 (0.52)	0.36	0.449	144.0	5.09	0.00
GARCH-Normal	0.26 (0.13)	1.11 (0.49)	0.29	0.423	144.7	5.10	0.00
EGARCH-Normal	0.32 (0.12)	0.96 (0.56)	0.31	0.422	144.3	5.12	0.00
				Dubai			
AR-FIGARCH							
Student	0.37 (0.02)	0.98 (0.03)	0.43	0.477	196.74	9.69	0.00
Skewed St.	0.97 (0.16)	0.96 (0.02)	0.47	0.444	176.85	7.96	0.00
Normal	0.14 (0.25)	1.02 (0.06)	0.41	0.365	182.33	8.65	0.00

AR-FIAPARCH							
Student	0.54 (0.02)	1.01 (0.01)	0.24	0.221	133.2	14.8	0.00
Skewed St.	0.76 (0.01)	1.03 (0.16)	**0.43**	**0.220**	**131.7**	**13.9**	**0.33**
Normal	0.63 (0.06)	1.02 (0.06)	0.36	0.245	133.2	12.3	0.00
GARCH-Normal	0.21 (0.33)	1.09 (0.08)	0.39	0.263	135.6	15.1	0.01
EGARCH-Normal	0.13 (0.16)	0.96 (0.09)	0.35	0.285	131.9	14.9	0.00
			Oman				
AR-FIGARCH							
Student	0.37 (0.01)	0.97 (0.16)	0.03	0.150	114.2	6.79	0.00
Skewed St.	0.97 (0.04)	0.64 (0.33)	0.07	0.430	113.6	6.12	0.00
Normal	0.52 (0.09)	0.63 (0.12)	0.11	0.450	115.7	6.33	0.00
AR-FIAPARCH							
Student	0.31 (0.16)	1.02 (0.66)	0.33	0.379	99.32	7.01	0.00
Skewed St.	0.46 (0.22)	1.67 (0.16)	**0.39**	**0.376**	**96.61**	**6.33**	**0.63**
Normal	0.12 (0.11)	1.33 (0.02)	0.24	0.380	98.01	6.95	0.00
GARCH-*Normal*	0.24 (0.23)	0.98 (0.26)	0.28	0.423	99.02	6.98	0.00
EGARCH-*Normal*	0.33 (0.12)	0.68 (0.09)	0.29	0.380	101.2	6.99	0.01
			Qatar				
AR-FIGARCH							
Student	0.43 (0.06)	1.02 (0.22)	0.03	0.143	22.33	5.78	0.00
Skewed St.	0.11 (0.25)	1.01 (0.33)	0.21	0.417	23.06	5.12	0.00
Normal	0.12 (0.11)	1.03 (0.06)	0.32	0.552	24.01	5.16	0.01
AR-FIAPARCH							
Student	0.37 (0.22)	0.77 (0.93)	0.44	0.219	20.02	7.97	0.00
Skewed St.	0.13 (0.09)	1.02 (0.06)	**0.44**	**0.135**	**19.66**	**7.78**	**0.31**
Normal	0.14 (0.26)	0.86 (0.04)	0.32	0.165	21.33	6.23	0.00
GARCH-*Normal*	0.66 (0.18)	0.90 (0.16)	0.39	0.178	22.09	6.03	0.00
EGARCH-*Normal*	0.51 (0.16)	1.21 (0.06)	0.38	0.195	22.00	5.88	0.00

			Kuwait				
AR-FIGARCH							
Student	0.42 (0.23)	1.02 (0.05)	0.03	0.236	16.57	5.82	0.00
Skewed St.	0.21 (0.06)	1.01 (0.23)	0.11	0.338	14.37	4.86	0.00
Normal	0.24 (0.12)	0.96 (0.13)	0.16	0.296	15.01	5.68	0.00
AR-FIAPARCH							
Student	0.77 (0.01)	0.57 (0.12)	0.44	0.230	17.21	7.06	0.00
Skewed St.	0.21 (0.22)	0.88 (0.25)	**0.53**	**0.201**	**14.30**	**7.00**	**0.23**
Normal	0.20 (0.06)	0.95 (0.13)	0.47	0.210	15.01	5.55	0.01
GARCH-Normal	0.36 (0.05)	1.02 (0.22)	0.23	0.225	14.98	4.99	0.00
EGARCH-Normal	0.49 (0.03)	0.89 (0.14)	0.36	0.216	14.68	5.69	0.00
			Bahrain				
ARFIMA-FIGARCH							
Student	0.43 (0.01)	1.02 (0.19)	0.41	0.243	156.4	6.03	0.00
Skewed St.	0.33 (0.12)	1.01 (0.25)	**0.41**	**0.217**	147.4	6.02	**0.48**
Normal	0.11 (0.16)	0.85 (0.06)	0.39	0.252	145.3	6.00	0.01
ARFIMA-FIAPARCH							
Student	0.77 (0.23)	0.82 (0.12)	0.41	0.279	155.5	7.99	0.00
Skewed St.	0.44 (0.09)	0.87 (0.06)	0.40	0.243	153.2	7.98	0.00
Normal	0.06 (0.08)	1.02 (0.18)	0.39	0.268	150.3	7.44	0.00
GARCH-Normal	0.33 (0.10)	0.86 (0.12)	0.36	0.283	161.8	7.22	0.00
EGARCH-Normal	0.34 (0.09)	0.98 (0.11)	0.39	0.262	160.9	7.63	0.00

Notes: This table reports the mean losses of the different volatility models over the out-of-sample period with respect to the evaluation criteria. α and β refer to the estimated coefficients of the Mincer-Zarnowitz (1969) regression. While R^2 is the determination coefficient. Standard errors calculated using the Newey-West correction are reported in parentheses. MSE is the mean square error. MAPE is the mean absolute prediction error. LL is the logarithmic loss function. SPA is the superior predictive accuracy test of Hansen (1995).

more accurate forecasts than the other models with skewed Student density and is the best-fitting model for only the Bahraini stock market.

Even though our results point to the superiority of non-linear GARCH-class models over the short horizon (one-day-forecasting) relative to linear models (GARCH and EGARCH), the appropriate choice of the "correct" GARCH-class model is a challenging task since the considered non-linear GARCH-class model categorically outclasses the other competing models. In fact, the forecasting accuracy changes across the GCC markets. Such disparities provide better investment opportunities for funds operating in the region and require market operators to take into consideration asymmetry, LM and fat tails when checking the relevance of a particular GARCH-class framework.

6. The Role of LM and Asymmetry in VaR and ES Estimations

As far as market risk matters, portfolio managers may wish to know the best suitable model that can be used to forecast stock market behavior and to estimate the VaR and ES for their portfolios. Given that our previous forecasting analysis provides significant support for non-linear GARCH-class models accommodating long memory, asymmetry and fat tails, we now produce VaR and ES estimates based on the selected model for each market. For comparison purposes, we refer to the standard GARCH model with normal distribution. Our main idea is to check the relevancy of the selected long-memory GARCH-class models in forecasting the VaR and ES. In addition, we extend our previous analysis to research the best-fitting non-linear GARCH-class model for the GCC multi-country portfolio and we estimate the VaR and ES for short and long trading positions.

6.1 The In-Sample VaR and ES Estimations

To do so, we estimate the in-sample VaR and ES and the out-of-sample VaR and ES forecasts for both short and long trading positions for a one-day investment horizon and confidence levels ranging between 0.25% and 5%. As noted above, we conduct several tests including the Kupiec (1995) test and the DQT proposed by Engle and Manganelli (2004) to evaluate accuracy. The VaR and ES estimates are displayed in *Table A4*[14] (see *Appendix 6* online) With regard to the utilized accuracy tests, we note that, under normal distribution, the standard GARCH model performs poorly for all of the GCC stock markets. For example, the failure rates significantly exceed the prescribed quantiles as they decrease from 5% to 1% and are quite similar for short and long trading positions. Furthermore, for the Saudi Arabian and Bahraini markets, our results indicate that under skewed Student distribution, the ARFIMA-FIAPARCH model performs better than the normal distribution for both short and long trading positions. Particularly for the Saudi stock market, the *p*-values of the Kupiec and DQT tests indicate that the null hypothesis is not rejected for all of the selected quantiles except for the 0.95 confidence level.

Quite comparable results are obtained for the Bahraini stock market, as ARFIMA-FIAPARCH model generates good estimates for the two trading positions. Indeed, the selected accuracy tests fail to reject the null hypothesis for all of the prescribed confidence levels, except for the DQT of Engle and Manganelli (2004) with 0.25% for short positions and 5% for long positions. Therefore, an important finding emerging from these results is that taking into account long memory jointly in stock returns and volatility, asymmetry and fat tails in the return innovations enhance the VaR and the ES estimations for short and long trading positions. For the other markets, we see

[14] We note that, due to space limitations, the VaR and ES generated by the standard GARCH under normal distribution results are not presented here, but are available upon request submitted to the corresponding author.

that the use of the FIAPARCH or FIGARCH models under skewed Student distribution leads to a clear improvement of the in-sample VaR and ES estimations relative to the standard GARCH model whether the undertaken trading action is buy or sell. Indeed, when the LM and asymmetric effects are accounted for in the FIAPARCH model, the Kupiec (1995) and Manganelli (2004) tests fail to reject the null hypothesis for all of the significance levels and for all of the GCC markets return series. These findings lead us to conclude that the VaR estimates provided by the long-memory GARCH-class models are more accurate than those of the standard GARCH model over the in-sample period. Thus, taking LM into account, asymmetry and fat tails in volatility models improve the quality of the VaR estimations. Our results are thus in line with Härdle (2008), Degiannakis *et al.* (2013) and Sethapramote *et al.* (2014), suggesting that taking into account long-range memory and asymmetry could provide better performance in risk management than that of standard GARCH.

6.2 The Out-of-Sample VaR and ES Forecasts

We analyze the forecasting performance of the VaR and ES with respect to the selected GARCH-class models under skewed Student distribution. Our forecasts are based on a window updating the model parameters every 50 observations in the out-of-sample period. Specifically, we compute 1,000 out-of- sample VaR and ES for each stock return. As in the in-sample VaR estimates, the out-of-sample VaR and ES are compared to the observed returns and then both results are recorded for latter assessment using the considered accuracy tests. As mentioned above, we use the Kupiec (1995) test, the DQT of Engle and Manganelli (2004) and the Giacomini and Kamunjer (2005)[15] accuracy testing methodology. We should recall that for the Giacomini and Kamunjer (2005) test procedure, we test the null hypothesis that the loss function of a benchmark forecaster is the same as the loss function of the selected forecasting model (m), under the alternative assumption that the benchmark model is more accurate than the competing one. Since we are testing the relevancy of the selected long-memory GARCH-class models in producing accurate VaR and ES forecasts, we chose the standard GARCH-model as a benchmark. *Table A5* (see *Appendix 7* online) reports the one-day-ahead VaR and ES forecasts for short and long trading positions.

Broadly speaking, the reported results show that the out-of-sample VaR forecasts are quite similar to their counterparts for the in-sample period. This result implies that the long-memory GARCH-class models are able to accommodate major empirical facts on the GCC return volatilities. Specifically, the selected LM GARCH-class models provide better forecasting accuracy than the GCC return volatilities for the out-of-sample period. With reference to the standard GARCH as a benchmark model, we note that the ARFIMA-FIAPARCH and the FIAPARCH models with skewed distribution provide better forecasts for all of the critical levels as the null hypothesis is not rejected in all cases according to the Kupiec and DQT tests. Additionally, we perceive that for long trading positions, the out-of-sample VaR forecasts are marginally better than their equivalents for short positions. The p-values corresponding to the Kupiec (1995) test and the DQT of Engle and Manganelli (2004), statistics are higher for all of the selected markets. Furthermore, when looking at

[15] We are grateful to an anonymous reviewer for suggesting this test.

the estimated loss functions of Giacomini and Kamunjer (2005), we can see that the norm of quantiles implied by different models can be distinguished from each other statistically for all of the prescribed confidence levels. This result indicates that the selected long-memory GARCH-class models under Student and skewed Student distributions provide better VaR forecasts than the standard GARCH model under normal distribution. Considering what has been discussed above, we can claim long dependence either in the return or conditional variance, while asymmetry and fat tails play an important role in forecasting GCC volatilities and thus for forecasting VaR and ES. Our results are in line with the findings of Giot and Laurent (2004), Kang and Yoon (2007), Mabrouk and Aloui (2010), Mabrouk and Saadi (2012), and Degiannakis *et al.* (2013). In *Figure A3* (see *Appendix 8* online), we report the conditional variance and the one-day-ahead VaR plot for the out-of-sample period.

6.3 The GCC Multi-Country Portfolio Risk Assessment

In this sub-section, we extend our previous analysis by researching the best-fitting non-linear GARCH-class model for the GCC multi-country portfolio and computing the VaR and ES for short and long trading positions. To do so, we use the Bloomberg GCC 200 Index (BGCC-index), which is a capitalization-weighted index of the top 200 stocks in the GCC markets based on market capitalization and liquidity. The estimation results show that the ARFIMA-FIAPARCH model under skewed Student distribution is the best model for describing the GCC multi-country portfolio.[16] The fractional integration parameter in the mean return is equal to 0.573 and is significant at the 1% level while it is equal to 0.68 for the conditional volatility. The APARCH coefficients γ and δ are positive and are equal to 0.311 and 1.9743, respectively. Furthermore, the tail parameter is equal to 2.73 and significant at the 1% level while the asymmetry coefficient is negatively signed (-0.0396) and significant at the 1% level. The diagnostic tests show that the ARFIMA(1,d,0)-FIAPARCH(1,d,1) under skewed Student distribution outperforms the same model under normal and Student-t distributions. *Table A6* (see *Appendix 9* online) reports the in-sample VaR and ES estimates for both short and long trading positions while *Figures A4a* and *A4b* (see *Appendix 10* online) report the daily return, conditional variance and one-day-ahead VaR of the GCC multi-country portfolio for the in-sample and out-of-sample periods, respectively.

Despite the fact that they are not as good as the in-sample forecasts, the estimation results show that the ARFIMA(1,d,0)-FIAPARCH(1,d,1) model with skewed distribution performs well for the out-of-sample forecasts. This result confirms the main idea that GARCH-class models produce accurate in-sample estimates but less satisfactory out-of-sample forecasts. Regarding long trading positions, the out-of-sample results are quite similar to those of the in-sample period. However, we note that the p-values for the in-sample VaR estimates are greater than those of the out-of-sample. On the other hand, for short trading positions, the selected ARFIMA-FIAPARCH model with skewed Student-t distribution performs very well for the GCC multi-country portfolio at all confidence levels.

[16] These results are not displayed here, but are available upon request to the corresponding author.

7. Conclusion

This article examines the relevance and usefulness of LM and asymmetry in modeling and forecasting the conditional volatility and market risk for seven stock markets operating in the GCC region. By implementing a GARCH-class methodology, our empirical study makes it possible not only to find the best-fitting GARCH-class model that takes into account the major empirical facts in stock market volatility, but also to examine the ability of competing GARCH-class models to predict GCC market risk using VaR and ES. Our empirical framework is distinguishable from existing studies in at least four main points: firstly, our sample is extended to cover all of the GCC stock markets, and we consider a large sample period in daily frequency. Secondly, we consider a border set of linear and non-linear GARCH-class models including FIGARCH, FIAPARCH, GARCH and EGARCH under three density functions (Normal, Student and skewed Student). Thirdly, we investigate simultaneously for LM in the mean and conditional variance of stock returns. Finally, we estimate jointly the VaR and ES for both short and long trading positions as well as for a GCC multi-country portfolio.

Regarding the estimation results, we find that non-linear GARCH-class models achieve superior performance in comparison with the other competing models for both the in-sample and out-of-sample periods. Interestingly, our forecasting analysis reveals that long-memory GARCH-class models (FIGARCH and FIAPARCH) outperform the simple GARCH and EGARCH models for all of the selected stock return innovations. Thus, not only LM but also asymmetry effects and the existence of fat tails are prominent for modeling and forecasting the volatility of the GCC markets under study. In addition, we show the correct choice of the FIAPARCH model to compute the VaR and ES for a GCC multi-country portfolio. The relevant models show evidence of strong persistence and asymmetry in the seven GCC stock markets. This has several important implications for both participants and policymakers. On the one hand, investors and fund managers concerned with GCC markets should favor models that accommodate asymmetries, LM and fat tails when they construct their predictions. Policymakers therefore should be aware that negative shocks such as wars, geopolitical risk and economic recession have a greater effect on stock market volatility than positive shocks such as an increase in energy commodity prices or greater improvement of the global GCC economies. On the other hand, in a non-Fickian or non-neutral market, LM in return behavior systematically implies the existence of profit-making arbitrage opportunities, making the considered markets inefficient and unfair. Policymakers thus have an interest in undertaking appropriate actions to improve the efficiency of stock markets in order to ensure their attractiveness as a long-run financing source for promoting economic growth (Aloui and Nguyen, 2014).

Our research presents some shortcomings and can be extended in several ways. *Firstl*y, it would be interesting to estimate the GARCH-class models under other stock return innovation distributions including the generalized error distribution (GED), exponential generalized beta, and stable Paretian, and to enlarge the sample to include other MENA countries. *Secondly*, from a portfolio management perspective, it is useful to examine the relevancy of the selected GARCH-class models in forecasting volatility and assessing risk within a multi-step-ahead approach. In fact, the empirical literature recognizes that long-memory GARCH-class models improve forecasts mainly in multi-step-ahead horizons. *Thirdly*, it would be very important to show how to choose between the competing GARCH-class models based on calculations of the out-of-sample percentage of violations and daily capital charges that funds and other financial institutions must report to financial market regulatory authorities, in compliance with the Basel II Accord rules.

REFERENCES

Al-Maghaireh A, Awartani B (2012): Modeling and forecasting Value-at-risk in the UAE equity markets: the role of long memory, fat tails and asymmetries in return innovations. *Review of Middle East Economics and Finance*, 8:1–22.

Al-Maghyereh A, Al-Zoubi H (2006): Value-at-risk under extreme values: the relative performance in MENA emerging stock markets. *International Journal of Managerial Finance*, 2:154–172.

Aloui C, Bellalah M, Abaoub E (2005): Long-range dependence in daily volatility on Tunisian stock market. *International Journal of Business*, 10:192–215.

Aloui C, Nguyen DK (2014): On the detection of extreme movements and persistent behavior in Mediterranean stock markets: a wavelet-based approach. *Applied Economics*, 46:2611–2622.

Andersen T, Bollerslev T (1998): Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39:885–905.

Angelidis T, Degiannakis S (2007): Backtesting VaR models: a two-stage procedure. *Journal of Risk Model Validation*, 1:1–22.

Artzner P, Delbaen F, Eber JM, Heath D (1997): Thinking Coherently. *Risk*, 10:68–71.

Avdulaj K, Barunik J (2013). Can we still benefit from international diversification? The case of the Czech and German stock markets. *Finance a úvěr-Czech Journal of Economics and Finance*, 63:425–442.

Baillie RT, Bollerslev T, Mikkelson H (1996): Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 14:3–30.

Basel Committee on Banking Supervision (2006): *International convergence of capital measurement and capital standards—A revised framework.* Basel.

Beran J (1994): *Statistics for long memory processes*. Chapman & Hall, New York.

Berkowitz J, Christoffersen P, Pelletier D (2011): Evaluating value-at-risk models with desk-level data. *Management Science*, 52:2213–2227.

Conrad C, Karanasos M, Zeng N (2011): Multivariate fractionally integrated APARCH modeling of stock market volatility: a multi-country study. *Journal of Empirical Finance*, 18:147–159.

Degiannakis S (2004): Volatility forecasting: Evidence from a fractional integrated asymmetric power ARCH skewed-t model. *Applied Financial Economics*, 14:1333–1342.

Degiannakis S, Floros C, Dent P (2013): Forecasting value-at-risk and expected shortfall using fractionally integrated models of conditional volatility: international evidence. *International Review of Financial Analysis*, 27:21–33.

Dimitrakopoulos DN, Kavussanos MG, Spyrou SI (2010): Value at risk models for volatile emerging markets equity portfolios. *Quarterly Review of Economics and Finance*, 50:515–526.

Engle RF (1982): Autoregressive conditional heteroskedasticity with estimates of the variance of U.K inflation. *Econometrica*, 50:987–1008.

Engle R, Manganelli S (1999): *CAViaR: conditional autoregressive value at risk by regression quantiles*. San Diego, Department of Economics—mimeo,
Available at http://community.stern.nyu.edu/rengle/research/.

Engle R, Manganelli S (2004): CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business Economics and Statistics*, 22:367–381.

Geweke J, Porter-Hudak S (1983): The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221–238.

Giacomini R, Komujer I (2005): Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economics Statistics*, 23:416–431.

Giot P, Laurent S (2003): Value-at-risk for long and short positions. *Journal of Applied Econometrics*, 18:641–664.

Giot P, Laurent S (2004): Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, 11:379–398.

Grané A, Veiga H (2007): The effect of realized volatility on stock return estimates. *Universidad Carlos III de Madrid, Statistics and Econometrics Working Papers, no 076316, available at: http://e-archivo.uc3m.es/bitstream/10016/947/1/ws076316.pdf.*

Granger CWJ, Joyeux R (1980): An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1:15–29.

Hansen PR (2005): A test of superior predictive ability. *Journal of Business and Economic Statistics*, 23:365–380.

Hansen PR, Lunde A, Nason JM (2011): The model confidence set. *Econometrica*, 79:453–497.

Kang S, Yoon S-M (2008): Value-at-Risk analysis of the long memory volatility process: the case of individual stock returns. *Korean Journal of Finance*, 21:4–18.

Kasman A (2009): Estimating Value-at-Risk for the Turkish Stock Index Futures in the Presence of Long Memory Volatility. *Republic of Turkey, Central Bank Review*, pp. 1–14.

Kim CS, Phillips PCB (2006): Log periodogram regression in the nonstationary case. Yale University, *Cowless foundation Discussion paper*, no. 1587.

Kupiec P (1995): Techniques for verifying the accuracy of risk management models, *Journal of Derivatives*, 2:173–184.

Lambert P, Laurent S (2001): Modelling financial time series using GARCH-type models and a skewed Student density. Université de Liège—Mimeo.

Lo AW (1991): Long-term memory in stock market prices. *Econometrica*, 59:1279–1313.

Mabrouk S, Aloui C (2010): One-day-ahead value-at-risk estimations with dual long-memory models: evidence from the Tunisian stock market. *International Journal of Financial Management Services*, 2:77–94.

Marzo M, Zagalia P (2007): Volatility forecasting for crude oil Futures. *Stockholm University, Department of Economics, Working Paper*, no. 599.

McMillan DG, Kambouridis D (2009): Are RiskMetrics good enough? Evidence from 31 stock markets. *International Review of Financial Analysis*, 18:117–124.

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

Newey W, West K (1987): A simple, positive semi-definite, heteroskedasticity and autocorrelation Consistent Covariance Matrix. *Econometrica*, 55:703–708.

Onour I (2010): The global financial crisis and equity markets in the Middle East oil exporting countries. *MPRA working paper, no 23332,*
*available on the link:www.*mpra.ub.uni-muenchen.de/23332/1/MPRA_paper_23332.pdf

Robinson PM (1995): Log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, 23:1048–1072.

Rossignolo AF, Fethi MD, Shaban M (2012): Value-at-risk models and Basel capital charges: Evidence from emerging and frontier stock markets. *Journal of Financial Stability*, 4:303–319.

Sethaparamote Y, Prukumpai S, Kanyamee T (2014): Evaluation of value at risk estimation using long memory volatility: evidence from stock exchange of Thailand. *SSRN paper*, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2396531.

Shao XD, Lian YJ, Yin LQ (2009): Forecasting value-at-risk using high frequency data: The realized range model. *Global Finance Journal*, 20:128–136.

Tang TL, Shieh SJ (2006): Long memory in stock index futures markets: A value-at-risk approach. *Physica A*, 366:437–448.

Tse YK (1998): The conditional heteroskedasticity of the Yen-Dollar exchange rate. *Journal of Applied Econometrics*, 193:49–55.

Xekalaki E, Degiannakis S (2010): *ARCH models for financial applications*. John Wiley & Sons.

Yamai Y, Yoshiba T (2005): Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking and Finance*, 29:997–1015.

Žikovic S, Filer RK (2013): Ranking of VaR and ES models performance in developed and emerging markets. *Finance a úvěr-Czech Journal of Economics and Finance*, 63:327–359.