European University Institute

ROBERT SCHUMAN CENTRE FOR ADVANCED STUDIES

# EUI Working Papers

RSCAS 2012/47
ROBERT SCHUMAN CENTRE FOR ADVANCED STUDIES
Loyola de Palacio Programme on Energy Policy

## RISK MANAGEMENT IN THE ENERGY MARKETS AND VALUE-AT-RISK MODELLING: A HYBRID APPROACH

Kostas Andriosopoulos and Nikos Nomikos

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ROBERT SCHUMAN CENTRE FOR ADVANCED STUDIES

LOYOLA DE PALACIO PROGRAMME ON ENERGY POLICY

# *Risk management in the energy markets and Value-at-Risk modelling: a Hybrid approach*

KOSTAS ANDRIOSOPOULOS AND NIKOS NOMIKOS

EUI Working Paper **RSCAS** 2012/47

ISSN 1028-3625



Printed in Italy, September 2012
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu/RSCAS/Publications/
www.eui.eu
cadmus.eui.eu

## Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies (RSCAS), created in 1992 and directed by Stefano Bartolini since September 2006, aims to develop inter-disciplinary and comparative research and to promote work on the major issues facing the process of integration and European society.

The Centre is home to a large post-doctoral programme and hosts major research programmes and projects, and a range of working groups and *ad hoc* initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration and the expanding membership of the European Union.

Details of the research of the Centre can be found on:
http://www.eui.eu/RSCAS/Research/

Research publications take the form of Working Papers, Policy Papers, Distinguished Lectures and books. Most of these are also available on the RSCAS website:
http://www.eui.eu/RSCAS/Publications/

The EUI and the RSCAS are not responsible for the opinion expressed by the author(s).

### *Loyola de Palacio Energy Policy Chair*

The Loyola de Palacio Energy Policy Chair was created in October 2008 at the RSCAS in honour of Loyola de Palacio, former Vice President of the European Commission and Commissioner for Energy and Transportation in the Prodi Commission. It promotes research in the area of energy policy. It is funded by contributions from donors. Professor Jean-Michel Glachant is the holder of the Chair.

The Chair focuses on the fields of energy economics, law, regulation, as well as geo-politics. It addresses topics such as the *achievement of the EU internal energy market; sustainable energy systems and the environment; energy security of supply; the EU model of energy regulation; the EU energy competition policy; the EU policy towards carbon free energy systems in 2050.*

The series of working papers aims at disseminating the work of academics on the above-mentioned energy policy issues.

*For further information*
Loyola de Palacio Energy Policy Chair
Nicole Ahner (scientific coordinator)
Email contact: Nicole.Ahner@eui.eu
Robert Schuman Centre for Advanced Studies
European University Institute
Via delle Fontanelle, 19
I-50016 San Domenico di Fiesole (FI)
Fax: +39055 4685755
http://www.loyola-de-palacio-chair.eu

## Abstract

This paper proposes a set of VaR models appropriate to capture the dynamics of energy prices and subsequently quantify energy price risk by calculating VaR and ES measures. Amongst the competing VaR methodologies evaluated in this paper, besides the commonly used benchmark models, a MC simulation approach and a Hybrid MC with Historical Simulation approach, both assuming various processes for the underlying spot prices, are also being employed. All VaR models are empirically tested on eight spot energy commodities that trade futures contracts on NYMEX and the Spot Energy Index. A two-stage evaluation and selection process is applied, combining statistical and economic measures, to choose amongst the competing VaR models. Finally, both long and short trading positions are considered as it is extremely important for energy traders and risk managers to be able to capture efficiently the characteristics of both tails of the distributions.

## Keywords

Energy markets, Mean Reversion Jump Diffusion, Value-at-Risk, Hybrid Monte Carlo & Historical Simulation.

## I. Introduction*

The events and especially the aftershocks of the recent financial crisis have been unprecedented, at least in terms of the speed and magnitude of the shock, and the potential long-term impact on the global real economy. As Rogoff and Reinhart (2008) point out, most of the 18 major banking crises and a number of more minor crises that they recorded since World War II, with a major market event appearing at least every 10 years, were caused by excess liquidity in the economy along with a general misjudgement on the benefits of a certain type of innovation. The recent financial crisis of 2007 was no different. Financial innovation in the form of sophisticated securitized instruments contributed to a false sense of security around systemic risk reduction, while at the same time excess liquidity was pouring into the developed countries' financial and housing markets, mostly by investments coming from the emerging markets.

The latest economic recession and its subsequent shock waves significantly affected international trade, the commodity markets and most specifically the energy markets. Oil markets rallied upwards for almost a year after the crisis started, peaking at $145 per barrel, then suddenly collapsed to $31 per barrel within a few months, quickly then recovering some of the lost ground, trading above $60 per barrel until now. These recent energy markets' dynamics can be attributed not only to the prevailing supply and demand conditions, but also to the growth of speculative investments by a more diverse and sophisticated body of market players, including investment banks, hedge funds, pension funds, Exchange Traded Funds (ETFs) and Exchange Traded Notes (ETNs) that follow the commodity markets. This increased sophistication and analytical skills that were brought in to the energy markets, made the use of forecasting models, hedging tools, and risk management techniques, and thus in extension the VaR applications, essential tools for quantifying energy price risk. In this newly created energy environment, precise monitoring and protection against market risk has become a necessity. Power utilities, refineries or any other energy market player can use valuable information derived from the VaR exercise applied in-house, to plan and implement their future risk management strategy.

Following the amendment of the Basel Capital Accord by the Basel Committee on Banking Supervision in 1998, that obliged all member banks to calculate their capital reserve on the basis of VaR, the VaR measurement has become extremely popular both with practitioners as well as academics. As a result, numerous methods have been developed for calculating VaR, proposing techniques that have been significantly refined from the initially adopted Risk Metrics (JP Morgan, 1996), with the goal of providing reliable estimates (Jorion, 2006). The aim of this paper is to investigate whether the widely used in the financial world Value-At-Risk (VAR) and Expected Shortfall (ES) methodologies, along with a new set of proposed models, can be successfully applied in the energy sector. VaR is used to identify the maximum potential loss over a chosen period of time, whereas the ES measures the difference between the actual and the expected loss when a VaR violation occurs.

Although a large body of the empirical literature is focused on forecasting energy prices and their volatilities, according to Aloui and Mabrouk (2010) they are far from finding any consensus about the appropriate VaR model for energy price risk forecasting. This paper attempts to close this gap in the existing literature by proposing a set of models appropriate to capture the dynamics of energy prices and subsequently quantify energy price risk by calculating VaR and ES measures. The methodologies employed include standard VaR approaches like the Risk Metrics, GARCH and many other commonly used models, MC simulations, and a hybrid Monte Carlo with Historical Simulations introduced for the first time in this paper (to the best of the author's knowledge). The model specifications for the MC simulations and the hybrid approach are the MR and MRJD models, modified to allow for

* We would like to thank the participants of the 1st Conference of the Financial Engineering and Banking Society (F.E.B.S.) in Athens, Greece for their helpful comments.

GARCH and EGARCH volatility, and for different speeds of mean reversion after a jump is identified, as described in paper three.

Simulation models are widely used in VaR applications since they help in understanding any potential risks in an investment decision, and in preparing for the possibility of a catastrophic outcome even though it might have a small probability of occurring. There are a number of recently proposed simulation methods for generating reliable VaR estimates due to the flexibility they offer. Huang (2010) proposes a Monte Carlo Simulation VaR model that accommodates recent market conditions in a general manner. By applying the methodology on the S&P 500 returns he finds that the VaR estimation via the proposed optimization process is reliable and consistent, producing better back-testing outcomes for all out-of-sample periods tested. By simulating the value of an asset under a variety of scenarios not only the possibility of falling below the desirable level can be identified, but there can also be measures taken to prevent this event from occurring in the future.

This paper employs a two-stage evaluation and selection process, combining statistical and economic measures, to choose between numerous competing VaR models applied in a number of energy commodities and the Spot Energy Index. The proposed SEI can be closely monitored by the major players of the energy industry and used as the underlying asset to many derivatives products such as futures and forwards, options, swaps, and also as the underlying index of energy ETFs, ETNs, and hedge funds. Amongst the competing VaR methodologies evaluated in this paper, besides the commonly used benchmark models, a MC simulation approach and a Hybrid MC with Historical Simulation approach, both assuming various processes for the underlying spot prices, are also being proposed.

In contrast to most existing studies on VaR modelling that consider only long positions, this paper examines both long and short trading positions. It is extremely important for energy traders and risk managers to know whether the models they are using can capture efficiently the characteristics of both tails of the distributions, as there are a lot of short players in the market alongside the long players. When taking short positions there is a risk of increasing prices, whereas when taking long positions the risk comes from falling prices. Thus, the focus should be on the left tail of the returns' distribution for the latter case, and on the right tail for the former case. Within the energy markets, the results of this paper have important implications for the accurate management of energy risk and the development of the fast-growing energy derivatives and ETFs markets.

Furthermore, although the proposed VaR model selection process reduces the numerous competing models to a smaller set, in some cases more than one model is identified as the most appropriate. It is in those cases that the modeller should view the selection process as being more valuable and useful than the actual VaR number obtained, and use in combination to the proposed evaluation process other real world considerations for his/ her final choice. As Poon and Granger (2003) argue in their paper, the most important aspect of any forecasting exercise is by itself the comparison process of competing forecasting models.

The structure of this paper is as follows. Sections 2 and 3 describe the VaR methodologies and the back-testing procedure employed, respectively. Section 4 presents the data used. Section 5 offers the empirical results of the study and, finally, section 6 concludes the paper

## II. VAR Methodologies

VaR is defined as the maximum expected loss in the value of an asset or a portfolio of assets over a target horizon, subject to a specified confidence level. Thus, VaR sums up the risk which an asset or a portfolio is exposed to in a single monetary (or expected return) figure. That makes the VaR approach directly applicable to the field of energy prices. Statistically speaking, the calculation of VaR requires the estimation of the quantiles of the distribution of returns and can be applied to both the left (long positions) and the right (short positions) tails. Generally, the VaR of a long position can be expressed by the following formula:

$$Pr(r_{t+1} \leq VaR^{\alpha}_{t+1}|\Omega_t) = \alpha \tag{1}$$

where, $r_{t+1}$is the return of the asset or portfolio of assets over a time horizon (in this case one day) from t to t+1, α is the confidence level, and $\Omega_t$ is the information set at time t. The VaR for a short position is computed using the same definition, with the only difference of substituting α with 1-α. The ES for a long position, defined as the average loss over the VaR violations from the N out-of-sample violations, is also expressed mathematically as:

$$ES_\alpha = \{E[r_{t+1}|(r_{t+1} \leq -VaR_{t+1}(\alpha))] \tag{2}$$

As far as the energy markets are concerned, there has been a recent increase in the relevant empirical literature on testing VaR models and assessing their performance. These papers include a wide range of models from the standard Variance Covariance, to Historical Simulation variations, Monte Carlo simulation, and a plethora of models of the ARCH-type, also including long memory variations, under different distributional assumptions for the returns' innovation (see among others, Chiu et al., 2010; Aloui and Mabrouk, 2010; Huang et al., 2008; Sadeghi and Shavvalpour, 2006; Giot and Laurent, 2003; Cabedo and Moya, 2003). Moreover, there have also been a few studies estimating VaR on the energy markets using an extreme value theory approach (see among others, Nomikos and Pouliasis, 2011; Marimoutou et al., 2009; Krehbiel and Adkins, 2005). Results however, are contradictory in terms of the accuracy of the VaR models proposed, with plenty of discussions focusing on as to whether the simpler models can outperform the more complex/ flexible ones. Brooks and Persand (2003) find that simple models achieve comparably better VaR forecasts to the more complex ones, while Mittnik and Paolella (2000) show that more accurate VaR forecasts can be achieved with the more flexible models. In addition, Bams et al. (2005) find that amongst the models they examine, the simple models often lead to underestimation of the VaR, whereas the opposite holds for the more complex models that seem to lead to overestimation of the VaR.

Furthermore, following the emerging concept in the literature of combining VaR forecasts, Chiu at al. (2010) propose a composite VaR model to increase forecast effectiveness. In the same lines, Hibon and Evgeniou (2005) suggest that by combining forecasts instead of selecting an individual forecasting model, modelling risk is reduced. Choosing the most suitable VaR model for each commodity and for the SEI is of outmost importance for all energy market players, traders, hedgers, regulators, and policy-makers as modelling risk is reduced, and thus avoiding faulty risk management caused by the selected model's inefficiencies.

In principle, there are three general approaches to compute VaR, each one with numerous variations. The first one is to assume the return distributions for the market risks. The second one is to use the variances and co-variances across the market risks, and the third one is to run hypothetical portfolios through historical data or by using Monte Carlo simulations. Within these three general approaches to VaR, there are many different methodologies available, supported mostly by the internal model's approach that gives banks and investment houses the freedom to choose or develop their own methodology.

This paper describes various models originating from all three approaches, and compares their performance for accurately calculating VaR for the energy commodity markets. Considering that the proposed MC simulation models jointly take into account two sources of uncertainty, jumps and high volatility with both having some predictable component, the VaR estimates from the proposed specifications are compared to those obtained with more established methods, like the RiskMetrics or Historical Simulation methods. In addition, a Hybrid approach for calculating VaR is developed based on a combination of both the MC Simulations and the Historical Simulation methodologies. Table 1 (panels A to D) summarizes all the VaR models compared in this paper, in total twenty two. All the models listed under panels A and B are variance forecasting models with their sole focus on forecasting tomorrow's volatility. Panels C and D list all the proposed Monte Carlo simulation and Hybrid Monte Carlo – Historical Simulation models. Thus, the major difference between all aforementioned models lies with the methodology used to calculate volatility. The methodology, the main properties and the underlying distribution used in each model, both the more established and the proposed ones are all explained in more detail in the subsequent sections.

Variance-Covariance Model

The Variance-Covariance (V&C) method is a widely used method of computing VaR due to its simplicity and computational efficiency. However, it has a major drawback as it assumes that returns are normally distributed; a rather unrealistic assumption for the energy markets that are characterised by fat-tailed return distributions. Within the family of V&C methods there are several methodologies that can be used to calculate the VaR, based on the way the forecasted variance is calculated. For the purposes of this paper the equally weighted Moving Average (MA) methodology is used, which assumes that future variance can be estimated from a pre-specified window of historical data, weighing equally all the historical observations used. The equally weighted MA model is expressed as:

$$\sigma_t = \sqrt{\frac{1}{(k-1)}\sum_{s=t-k}^{t-1} r_s^2} \tag{3}$$

where, t is the estimation date of the standard deviation of returns over a time window from date t-k to t-1.

RiskMetrics

RiskMetrics (RM) is an Exponentially Weighted Moving Average (EWMA) VaR measure assuming that the standardised returns (returns over the forecasted standard deviation) are normally distributed (JP Morgan, 1996). The RM methodology focuses on the size of the returns but only relative to their standard deviation. A large return, irrespective of the direction, during a period of high volatility could lead to a low standardised return, whereas during a low volatility period it could result to an abnormally high standardised return. This standardization process leads to a more accurate VaR computation as large outliers are considered more frequent than would be expected with a normal distribution. The unconditional standard deviation of the RM model is expressed as:

$$\sigma_t = \sqrt{(1-\lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2}; \; \lambda \in (0,1) \tag{4}$$

where λ is the decay factor, reflecting how the impact of past observations decays while forecasting one-day ahead volatilities. The more recent the observation the largest the impact, with an exponential decay effect as observations move more into the past. The highest (lowest) the value for λ is, the longer (shorter) the memory of past observations is. The value of 0.94 is assigned for λ which is widely used in the literature.

ARCH Models

ARCH (autoregressive conditional heteroscedasticity) models of volatility, initially proposed by Engle (1982), are commonly used by researchers and practitioners to calculate the VaR of their portfolios. Amongst the most popular ARCH formulations used are the GARCH (Bollersev, 1986) and EGARCH (Nelson, 1991) volatility models, because of their ability to capture many of the typical stylised facts of both financial and commodity time series, such as time-varying volatility, persistence, and volatility clustering. According to Engle (2001), models that explicitly allow for the standard deviation to change over time, thus allowing for heteroskedasticity, perform better in forecasting the variance, and thus by extension, in measuring the VaR. Giot and Laurent (2003) and Kuester et al. (2006) conclude that VaR can be captured more accurately using GARCH-type models instead of using non-parametric ones. A key advantage of the GARCH and EGARCH models in terms of calculating VaR is that, according to Christoffersen (2003), the one-day forecast of the variance $\sigma_{t+1|t}^2$, is given directly from the model as $\sigma^2_{t+1}$, which is the conditional volatility following respectively a GARCH or an EGARCH process. A more detailed explanation is given in the following sections.

### *A. GARCH & Filtered GARCH*

Under the GARCH volatility specification the return series is assumed to be conditionally normally distributed, with the VaR measures being calculated by multiplying the conditional standard deviation by the appropriate percentile point on the normal distribution, following Sarma et al. (2003). The conditional volatility following a GARCH(1,1) process is expressed as:

$$\sigma_t = \sqrt{\beta_0 + \beta_1 \varepsilon^2_{t-1} + \beta_2 \sigma^2_{t-1}} \tag{5}$$

where, $\beta_0, \beta_1,$ and $\beta_2$ are positive constants, with $\beta_1 + \beta_2 < 1$ expressing the "non-explosivity" condition, $\varepsilon^2_{t-1}$ representing the previous periods' return innovations, and $\sigma^2_{t-1}$ being the last period's forecast variance (GARCH term). Once $\sigma_t$ is forecasted, the VaR estimates are obtained using the relevant percentile points on the normal distribution for the 99% and 95% VaR, under both long and short positions. Daily volatility forecasts are computed using a rolling estimation window of 1827 daily observations each. The process is then rolled forward until all the data is exhausted[1].

Next, the VaR based on the Filtered GARCH (F-GARCH) process is also calculated. The term filtered refers to the fact that instead of using directly the forecasted variance from the GARCH model, a set of shocks $z_i$ is used, as explained below, which are returns filtered by the forecasted variance . The VaR is estimated from the empirical percentile, which is based on observed information, using the following mathematical expression:

$$VaR_{t+1} = \hat{\sigma}_{GARCH_{t+1}} Percentile\{(z_i)^t_{i=t-T}, a\}; T = 1827\ days \tag{6}$$

where $z_i = r_i / \hat{\sigma}_{GARCH_{t+1}}$ are the standardised residuals and $\hat{\sigma}_{GARCH_{t+1}}$ is the forecasted GARCH volatility using an estimation sample window of width $T = 1827$ days.

### *B. EGARCH & Filtered EGARCH*

To cope with the skewness commonly observed in commodities markets, and to capture the potential presence of an "inverse leverage" effect[2], the more flexible model of persistence, the Exponential GARCH (EGARCH) model is used, which is expressed as:

$$\sigma_t = \sqrt{e^{\beta_0 + \beta_1 * \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \beta_2 * \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta_3 * \ln\left(\sigma^2_{t-1}\right)}} \tag{7}$$

where, β0 denotes the mean of the volatility equation. The coefficients β1 and β2 measure the response of conditional volatility to the magnitude and the sign of the lagged standardised return innovations, respectively; as such, these coefficients measure the asymmetric response of the conditional variance to the lagged return innovations. When $\beta_2 = 0$, there is no asymmetric effect of the past shocks on the current variance, while when $\beta_2 \neq 0$ asymmetric effects are present in response to a shock; for instance, $\beta_2 > 0$ indicates the presence of an "inverse leverage" effect. Finally, $\beta_3$ measures the degree of volatility persistence.

---

1 The starting coefficients for the GARCH models are obtained from the Yule-Walker equations, and the log-likelihood function is maximized using the Marquardt optimization algorithm.

2 Financial markets tend to exhibit a negative correlation between volatility and price, an effect known as "leverage", with negative shocks having a greater impact on volatility compared to positive ones.

As in the case with the GARCH model, the Filtered EGARCH (F-EGARCH) process is also calculated. Again, the term filtered refers to the fact that a set of returns filtered by the forecasted EGARCH variance is used. The VaR is estimated using the following mathematical expression:

$$VaR_{t+1} = \hat{\sigma}_{EGARCH_{t+1}} Percentile\{(z_i)_{i=t-T}^{t}, a\}; T = 1827\ days \tag{8}$$

where $z_i = r_i / \hat{\sigma}_{EGARCH_{t+1}}$ are the standardised residuals and $\hat{\sigma}_{EGARCH}$ is the forecasted EGARCH volatility using as estimation sample window of width $T = 1827$ days.

### *C. Monte Carlo Simulation*

Another popular method for estimating VaR is Monte Carlo simulation which is based on the assumption that prices follow a certain stochastic process (GBM, JD, MR-JD etc.), and thus by simulating these processes one can yield the distribution of the asset's value for the predetermined period. By simulating jointly the behaviour of all relevant market variables to generate possible future values, the MC simulations method allows for the incorporation of future events affecting the market as well as the additions of jumps or extreme events, thus accurately modelling the market's behaviour. In VaR applications, the required quantile for both the left and the right tails can be obtained directly from the random paths. MC simulation is a powerful tool for energy risk management that owes its increased popularity to its flexibility. It can incorporate in the modelling procedure all the important characteristics of the energy markets' behaviour such as seasonality, fat tails, skewness and kurtosis, and is also able to capture both local and non-local price movements. It is mostly due to this flexibility that Duffie and Pan (1997), and So et al. (2008) conclude that the MC approach is probably the best VaR methodology. The only troubling issue with the MC approach is the fact that it is relative complex to implement, and that it can be computationally demanding.

With the MC simulations method the VaR of an asset or a portfolio is quantified as the maximum loss in the random variables distribution, associated with the appropriate percentile. In order to calculate the VaR, first the dynamics of the underlying processes i.e. prices, volatilities etc. need to be specified. Second, N sample paths need to be generated by sampling changes in the value of the asset or individual assets that comprise a portfolio (risk factors), over the desired holding period. Third, all information enclosed in the probability distribution needs to be incorporated. Fourth, using the N sample paths the value of each underlying risk factor needs to be determined, given the assumed process for each one. Finally, the individual values need to be used to determine the value of the asset/ portfolio at the end of the holding period.

The following seven specifications are used for modelling the spot prices of the energy markets examined:

Geometric Brownian Motion (GBM)

Mean Reversion with Ordinary Least Squares (constant) volatility (MR-OLS)

Mean Reversion with GARCH(1,1) volatility (MR-GARCH(1,1))

Mean Reversion with EGARCH(1,1) volatility (MR-EGARCH(1,1))

Mean Reversion with Jump Diffusion and OLS volatility (MRJD-OLS)

Mean Reversion with Jump Diffusion and GARCH(1,1) volatility (MRJD-GARCH(1,1))

Mean Reversion with Jump Diffusion and EGARCH(1,1) volatility (MRJD-EGARCH(1,1))

As with other VaR methodologies, any modifications to the MC simulations approach focus mostly on using various techniques to reduce computational burden. For example, Jamshidan and Zhu (1997) use principal component analysis to narrow down the number of factors used into the simulation process, a procedure they name scenario simulations. Glasserman et al. (2000), guide the MC simulations sampling process using approximations from the V&C approach, resulting in time and resources savings without the loss of precision. The MC simulation along with the hybrid MC-HS methodologies proposed in this paper for estimating the VaR of energy commodities and the SEI are a significant improvement of existing ones due to their flexibility. They allow for any stochastic process to be used for describing the distribution of returns, and at the same time allow for the incorporation in the model of all major features that define the behaviour of energy prices. Such features include seasonality, time varying volatility, volatility clustering, mean reversion, jumps, and most importantly a different speed of mean reversion after a jump occurs.

For estimating all inputs for the MC simulations 1,827 daily observations from the in-sample period are used. Using each time the relevant underlying process 100,000 simulations are run, forecasting the spot prices 623 days ahead. Then, using the average simulated path the daily VaR for each one of the 623 forecasted returns is estimated. The mathematical expression for calculating the VaR using the MC Simulation models is the following:

$$VaR_t = Percentile\{r_t^s, a\} \tag{9}$$

where $r_t^s$ is the total number of simulated returns at time t.

The Mean Reversion Jump Diffusion Model

The more complex Mean Reversion Jump Diffusion model specification is defined as follows. It is assumed that log-prices can be expressed as the sum of a predictable and a stochastic component as follows:

$$\ln S_t = f(t) + Y_t \tag{10}$$

with the spot price represented as:

$$S_t = F(t)e^{Y_t} \tag{11}$$

where $F(t) \equiv e^{f(t)}$ is the predictable component of the spot price St that takes into account the deterministic regularities in the evolution of prices, namely seasonality and trend. Also, $Y_t$ is a stochastic process whose dynamics are given by the following equation:

$$dY_t = a_i(\mu - Y_t)dt + \sigma_t dZ_t + kdq_t \tag{12}$$

where $a_i$ is the mean reversion rate, μ is the long-term average value of $\ln S_t$ in the absence of jumps, $\sigma_t$ is the volatility of the series, $dZ_t$ is a Wiener process, k is the proportional jump size and $dq_t$ is a Poisson process. It is assumed that the Wiener and the Poisson processes are independent and thus not correlated, which further implies that the jump process is independent of the mean-reverting process.

Using equations (10) and (11), the modelling procedure by Dixit and Pindyck (1994) is followed and after applying Ito's Lemma, the proposed model can be discretised in the following logarithmic form:

$$\ln S_t = f_t + \left(\ln S_{t-1} * e^{-a_i \Delta t}\right) + \left(\ln \bar{S} - \frac{\sigma_t^2}{2a_i}\right) * \left(1 - e^{-a_i \Delta t}\right) + \sigma_t * \sqrt{\frac{1 - e^{-2a_i \Delta t}}{2a_i}} * \varepsilon_1 + J(\mu_J, \sigma_J) * I_{(u_t < \Phi \Delta t)} \tag{13}$$

where,

$$a_i = \begin{cases} a_1 = a_{JD}, \text{ when a jump occurs; for a duration equal to jump returns' half-life} \\ a_2 = a, \text{ otherwise} \end{cases} \quad i = 1,2 \tag{13.1}$$

$$\sigma_t = \begin{cases} \sigma_t = \sigma \; [\text{Constant}] \\ \sigma_t = \sqrt{\beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 * \sigma_{t-1}^2} \; [GARCH(1,1)] \\ \sigma_t = \sqrt{e^{\beta_0 + \beta_1 * \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \beta_2 * \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta_3 * \ln(\sigma_{t-1}^2)}} \; [EGARCH(1,1)] \end{cases} \tag{13.2}$$

$$f_t = \gamma_0 \sin\left(\frac{2\pi(t+\tau)}{252}\right) + \gamma_1 t \tag{13.3}$$

$$I_{(u_t < \Phi\Delta t)} = \begin{cases} 1 \text{ when } u_t < \Phi\Delta t, \text{ i.e when a jump occurs} \\ 0 \text{ when } u_t > \Phi\Delta t, \text{ i.e when there is no jump} \end{cases}$$

$$\begin{aligned} &\text{J} \sim N(\mu_J, \sigma_J) \text{ with Mean}: \mu_J = (\bar{\kappa}_J + \sigma_J \varepsilon_2) \text{ and Standard Deviation}: \sigma_J \\ &\varepsilon_1, \varepsilon_2 \sim N(0,1), \rho(\varepsilon_1, \varepsilon_2) = 0 \\ &u \sim U[0,1] \end{aligned} \tag{13.4}$$

where $\ln \bar{S}$ is the long-term mean (μ), $\Phi$ is the average number of jumps per day (daily jump frequency), $\bar{\kappa}_J$ is the mean jump size, σJ is the jump volatility, $\varepsilon_1$ and $\varepsilon_2$ are two independent standard normal random variables, and u is a uniform [0, 1] random variable. The term $\mathrm{I}_{(u_t < \Phi\Delta t)}$ is an indicator function which takes the value of 1 if the condition is true, and 0 otherwise. This condition leads to the generation of random direction jumps at the correct average frequency. When the randomly generated number is below or equal to the historical average jump frequency, the model simulates a jump with a random direction; no jump is generated when the number is above that frequency. When a jump occurs its size is the mean size of the historical jump returns plus a normally distributed random amount with standard deviation σJ. Notice as well that the proposed modelling approach allows for the possibility of both positive and negative jumps to occur[3].

In addition, the model takes into account the fact that most energy prices exhibit a seasonal behaviour that follows an annual cycle. Various methods have been used in the literature for the deterministic seasonal component, from a simple sinusoidal (Pilipovic, 1998) or a constant piece-wise function (Pindyck, 1999; Knittel and Roberts, 2005), to a hybrid of both functions (Lucia and Schwartz, 2002; Bierbrauer et al., 2007). This periodic behaviour is accounted for by fitting a sinusoidal function with a linear trend to the actual prices, as described by $f_t$. The estimation is done using Maximum Likelihood (ML), with the sine term capturing the main annual cycle, and the time trend capturing the long-run growth in prices[4]. Moreover, the possibility for the returns to have a different mean reversion rate after a jump occurs is incorporated into the model. This approach is in

---

3 Merton (1976) in his original jump diffusion model assumes that the jump size distribution is lognormal, and so jumps can occur in only one direction (positive jumps).

4 The approach used in Pilipovic (1998) is followed to calculate the seasonal component in the data, because this method is more flexible than using dummy variables. According to Lucia and Schwartz (2002) the use of dummy variables does not provide a smooth function for the seasonal component observed in the data, which can cause discontinuities when pricing forward and futures contracts.

line with Nomikos and Soldatos (2008) who use two different coefficients of mean reversion, one for the normal small shocks and another, larger, for the spikes to capture the fast decay rate of jumps observed in the energy markets. Geman and Roncoroni (2006) also analyse the existence of different speeds of mean reversion in the context of mean-reverting jump-diffusion models, by introducing a class of discontinuous processes exhibiting a “jump-reversion” component to represent the sharp upward moves that are shortly followed by drops of the same magnitude. The proposed approach is flexible enough to accommodate the fact that the abnormal events that cause the jumps have different effect in each market and hence, prices tend to remain at the level to which they jump for a longer or shorter period of time, depending on the energy market under investigation. Therefore, prices following a jump are adjusted by using in equation (13) a different mean reversion rate, noted as $a_{JD}$, for a period of time equal to the half-life of jump returns for each energy market; when another jump occurs within the duration of the half-life period used, then $a_{JD}$ is used again for the same number of days, counting from the day following the last jump [see equation (13.1)]. If no other jump occurs within that period, then a2 is used until a new jump occurs. The proposed model, by incorporating this half-life measure, allows for the model to better adapt to the duration of both short- and long-term shocks of a wide magnitude range, exhibited in energy prices. The latter allows for a higher flexibility compared to the model proposed by Nomikos and Soldatos (2008) which fits best mainly the highly volatile electricity markets, as the speed of mean reversion estimated after a spike shock is significantly higher than the normal mean reversion rate. In addition, the model proposed in this paper incorporates in its specification GARCH and EGARCH volatility, to account for volatility clustering and any asymmetries that are usually present in energy prices.

Regarding the mean-reverting part of equation 13, an exact discretization is used for the simulations since the presence of jumps complicates the use of a large $\Delta t$. This is because the drift of the mean-reverting process is a function of the current value of a random variable and in order to simulate the jumps correctly the time step $\Delta t$ must be small relative to the jump frequency. Because the rare large jumps are of biggest interest, if the time interval $\Delta t$ is sufficiently small, the probability of two jumps occurring is negligible $\left((\phi\Delta t)^2 \ll \phi\Delta t\right)$. That makes it valid to assume that there can be only one jump for each time interval; in this case one every day since $\Delta t$ is equal to one day. Especially when $\Delta t$ is increased to one week or one month, as it is usually the case with real option applications that involve pricing medium- and long-term options, it is more important to use an exact discretization for the simulation process, because the overall error from the first-order Euler and the Milstein approximations will be much higher [5]. The random number generation of the Monte Carlo (MC) simulations already introduces an error in the results, therefore using these approximations that need a very small $\Delta t$ and thus also introduce a discretization error, would lead to higher computational cost into the simulations.

As for the two time-varying volatility model specifications of equation (13.2), in the case of the GARCH process, $\varepsilon_{t-1}^2$ represents the previous periods' return innovations and $\sigma_{t-1}^2$ is the last period's forecast variance (GARCH term). As for the EGARCH process, β0 denotes the mean of the volatility equation. The coefficients β1 and β2 measure the response of conditional volatility to the magnitude and the sign of the lagged standardised return innovations, respectively; as such, these coefficients measure the asymmetric response of the conditional variance to the lagged return innovations. When $\beta_2 = 0$, there is no asymmetric effect of the past shocks on the current variance, while when $\beta_2 \neq 0$ asymmetric effects are present in response to a shock; for instance, $\beta_2 > 0$ indicates the

[5] Clewlow and Strickland (2000) use the first-order Euler's approximation in order to get the discrete time version of the Arithmetic Ornstein-Uhlenbeck: $x_t = x_{t-1} + a*(\bar{x} - x_{t-1})*\Delta t + \sigma*\sqrt{\Delta t}*\varepsilon_t$ where the discretization is only correct in the limit of the time step tends to zero.

presence of an "inverse leverage" effect. Finally, $\beta_3$ measures the degree of volatility persistence. Knittel and Roberts (2005) suggest that a positive shock in electricity prices represents an unexpected demand shock which has a greater impact on prices relative to a negative shock of the same size, as a result of convex marginal costs and the competitive nature of the market. Moreover, Kanamura (2009) suggests that this inverse leverage effect, i.e. positive correlation between prices and volatility, is a phenomenon often observed in energy markets, whereas evidence from the stock markets suggests that the opposite relationship exists between volatility and prices, namely the "leverage" effect[6]. Hence, intuitively, the asymmetry parameter is expected to be positive and significant for most energy markets, implying that positive shocks have greater effect on the variance of the log-returns compared to negative shocks, consistent with the presence of an "inverse leverage" effect.

### *D. Historical Simulation & Filtered Historical Simulation*

The historical simulation (HS) method is amongst the simplest ones for estimating the VaR for various assets and portfolios. HS uses the past history of returns to generate the distribution of possible future returns; in contrast to MC simulation which follows a certain stochastic process. In addition, the time series data used to run the HS are not used to estimate future variances and covariances, as is the case in the V&C approach; the assets returns over the time period examined provide all necessary information for computing the VaR. As with other methodologies for calculating VaR, there are various modifications of the HS method suggested, such as weighing the recent past more (Boudoukh et al., 1998), combining the HS with various time series models (Cabedo and Moya, 2003), and updating historical data for shifts in volatility (Hull and White, 1998).

Under the HS methodology, the VaR with coverage rate, a, is then calculated as the relevant percentile of the sequence of past returns, obtained non-parametrically from the data. The mathematical expression of the one-day-ahead VaR using the HS method is the following:

$$VaR_{t+1} = Percentile\left\{\left\{r_i\right\}_{i=t-T}^{t}, a\right\};\ T = 1827 \text{ days} \tag{14}$$

where T is the window width of past observations used. The window width of historical data used in the estimations plays a crucial role in the efficiency of the HS methodology. Having sufficient history of the relevant returns makes the HS method very attractive to use, mostly due to its simplicity, intuitive and straight forward implementation, and also its wide applicability to all instruments and market risk types. The HS method takes into account fat tails and skewness as it is based on past historical data. One of the method's drawbacks is that it is computationally demanding, and also the fact that the assumed returns distribution is based on the historical distribution over the time period selected, which can lead to significant variations in the VaR estimate when different time periods are used. This becomes even more important for the energy markets where risks are volatile and of big magnitude, and structural shifts occur at regular intervals.

Following, the VaR based on the Filtered Historical Simulation (FHS) is also calculated, using the following mathematical expression:

$$VaR_{t+1} = \sigma_i Percentile\left\{\left\{z_i\right\}_{i=t-T}^{t}, a\right\};\ T = 1827 \text{ days} \tag{15}$$

where $z_i = r_i / \sigma_i$ are the standardised residuals and $\sigma_i$ is the volatility of the 1827 historical observation window. The term filtered refers to the fact that the raw returns are not used to simulate,

[6] The "leverage effect" terminology is first used by Black (1976) who suggests that negative shocks on stock prices increase volatility more than positive ones. The intuition behind it is that a lower stock price reduces the value of equity relative to debt, thereby increasing the leverage of the firm and thus making it a more risky investment.

but instead a set of return shocks $z_i$, which are the returns filtered by the historical volatility of a window width T, are used. Thus, the FHS is a combination of the non-parametric HS and a parametric model. This combination is more likely to improve the HS VaR estimates as it continues to accommodate the dynamics of the empirical distribution, such as skewness, fat tails and volatility clustering. Also, the FHS method has the advantage that no assumptions need to be made for the distribution of the return shocks, and offers the flexibility of allowing the computation of any risk measure and for any investment horizon. Finally, one of the disadvantages that both the HS and FHS methods share is that each observation in the time series used for the simulation carries an equal weight for measuring VaR, which can be a problem when there is a trend identified in the series.

### *E. Hybrid Monte Carlo – Historical Simulation*

The Hybrid MC-HS approach developed in this paper can be the most appropriate methodology for calculating the VaR in the energy markets as it combines all the advantages of using two of the most popular and efficient existing methods, the MC simulations and the Historical Simulation. The HS methodology and all the proposed variations in the literature are mostly designed to capture any shifts in the recent past that are usually underweighted by the conventional approach. All of these proposed variations fail to bring in the risks that are not already included in the sampled historical period or to capture any structural shifts in the economy and the specific market examined. In contrast, the Hybrid MC-HS approach gives an accurate picture of the asset's risk as it allows for the incorporation of jumps and fat-tails in the returns' distribution, due to the flexibility provided by the MC simulations.

Both, the MC simulations and the Historical Simulation approaches are very popular amongst practitioners for calculating the VaR of their portfolios because of their flexibility, ease of use, and estimation performance. Perignon and Smith (2010) find that amongst the banks in their global sample that disclose their VaR methodology, 73% use the HS methodology or any of its variations, whereas the MC simulations methodology is the second most frequently applied VaR method, used by 22% of the banks. As mentioned previously, there have been many variations proposed in the literature for the MC simulations and the HS approaches, but only looking at each approach separately. However, to the best of our knowledge, is the first time that the MC simulation approach is combined with the HS in order to produce a Hybrid approach for calculating the VaR of energy assets.

Zikovic and Filer (2009) introduce a hybrid approach based on a combination of nonparametric bootstrapping and parametric GARCH volatility forecasting. They test the model using daily returns from sixteen market indexes, half from developed and the other half from emerging markets. The authors find that only the proposed hybrid model and the EVT-based VaR models can provide adequate protection in both developed and emerging markets. Lambadiaris et al. (2003) calculate the VaR in the Greek bond and stock markets using separately the HS and MC simulations approaches, and they find that for the linear stock portfolios the MC simulations approach performed better, as the HS approach overstated the VaR, whereas in the case of the non-linear bond portfolios the results are mixed. Vlaar (2000) investigates the Dutch interest rates term structure and applies the historical simulation, variance-covariance, and Monte Carlo simulation methods for estimating the accuracy of the VaR. He finds that the best results are obtained for a combined variance-covariance MC method that uses a term structure model with a normal distribution and a GARCH specification. Moreover, Hendricks (1996) compares the VaR estimates from the V&C and HS approaches, applied on foreign exchange portfolios, and concludes that both approaches have difficulties in capturing extreme outcomes and shifts in the underlying risks. Thus, it can be argued that in case of computing the VaR for non-linear assets over long time periods, where data are more volatile, with the non-stationarity and the normality assumptions being debatable, the MC simulations approach performs better than the HS approach.

Using each time the relevant underlying process 100,000 simulations are run, forecasting the spot prices 623 days ahead[7]. Then, using the average simulated path, the daily VaR is estimated using a 1 day ahead rolling window method as it is the case with the HS method. The estimation window is the first 1,827 daily forecasts, rolled one step forward for the next 623 days. The mathematical expression for calculating the VaR using the Hybrid model is the following:

$$VaR_{t+1} = Percentile\left\{ \left\{ \overline{r}_i \right\}_{i=t-T}^{t}, a \right\};\ T = 1827 \text{ days} \tag{16}$$

where $\overline{r}_t^{\,s} = \sum_{\omega=1}^{n} \frac{r_{t,\omega}^{s}}{n}$ is the average per time-step simulated return at time t, rt,ω is the return of the simulated spot price of path ω at time t, $n$ is the number of MC simulations, and T is the estimation window of 1827 observations.

## III. VaR Back-testing procedure

Having presented previously the various risk management techniques, this section sets forth a model selection process including all aforementioned models despite the major drawbacks and obvious limitations that some may have. This is done because it is expected that the tests of VaR models used, and the selection process proposed, will effectively reject the weakest models, knowing that some of them are widely used in practice. That makes the results of this paper even more important as useful feedback will be provided about the models' quality and efficiency.

To select the best model in terms of its VaR forecasting power, a two stage evaluation framework is implemented. In the first stage, three statistical criteria are used to test for unconditional coverage, independence, and conditional coverage, as proposed by Christoffersen (1998). A VaR model successfully passes the first stage evaluation only when it can satisfy all three statistical tests, at the 5% or higher significance level. In the second stage, a loss function is constructed in line with Lopez (1999) and Sarma et al. (2003) to test the economic accuracy of the VaR models that have passed the first evaluation stage. Then, the model that delivers that lowest loss function value is compared pair-wise with all remaining models that have passed the first evaluation stage, using the modified Diebold-Mariano (MDM) test as proposed by Harvey et al. (1997). Thus, the benchmark model is tested against the remaining models to choose the VaR calculation methodology which generates the least loss for each energy market. In general, it is worth noting that when choosing between VaR models the modeller should view the selection process as being more valuable and useful than the actual VaR number obtained.

To perform the proposed back-testing procedure a long period of historical data needs to be used. According to Alexander (2008), about 10 years of daily frequency data are needed for the results to be more powerful and to be able to reject any inaccurate VaR models. In this paper, 2,450 daily observations are used, representing almost 10 years of history of which 1,827 are used as the in-sample (estimation sample) and 623 as the out-of-sample period. Then, using the rolling window approach, the estimation sample is rolled over the entire data period, for a fixed length of 1 day as the risk horizon.

### *A. Statistical evaluation*

Statistical tests are used to back-test risk management models and access how well they can capture the frequency, independence, and magnitude of exceptions, defined as losses (gains) that exceed the

[7] For estimating the inputs for the MC simulations all 2,450 daily observations from the sample are used.

VaR estimates. Most of these tests rely on the assumption that the daily returns are generated by an i.i.d. Bernoulli process. Thus, the "hit sequence" or "failure process" of VaR violations is defined using an indicator function $I_{a,t}$ as:

$$I_{a,t+1} = \begin{cases} \begin{cases} 1, if\ r_{t+1} < VaR_{a,t+1} \\ 0, otherwise \end{cases} ; \text{for long positions} \\ \begin{cases} 1, if\ r_{t+1} > VaR_{a,t+1} \\ 0, otherwise \end{cases} ; \text{for short positions} \end{cases} \tag{17}$$

where rt+1 is the realised daily return from time t, when the VaR estimate is made, to time t+1. The hit sequence returns a 1 on a day t+1 if the loss on that day is larger than the VaR number forecasted. If there is no violation then the hit sequence returns a 0. In order to statistically back-test the VaR models, a sequence of $\{I_{t+1}\}_{t=1}^{T}$ across T needs to be constructed, indicating the past violations. In a sample with n observations, if the "hit" series $I_{a,t}$ follows an i.i.d. Bernoulli process, an accurate VaR model should return a number of "hits" equal to $n * a$.

Then, based on this hit sequence the VaR evaluation framework, as developed by Christoffersen (1998), is applied. Three tests for unconditional coverage, independence, and conditional coverage (which combines the unconditional coverage and independence into one test) are applied on the hit sequence, using in all cases a likelihood ratio statistic. Also, the P-values associated with the test statistic are calculated, using a 5% significance level. The two types of errors associated with the significance level chosen when testing a certain hypothesis in statistics, are the Type I (rejecting a correct model) and Type II (failing to reject an incorrect model) errors. The higher the significance level is, the larger the possibility for a Type I error. Thus, in line with common practice in risk management applications, and because Type II errors can be quite costly, a high enough threshold should be imposed for accepting the validity on any VaR model, and as such a 5% significance level is chosen in this paper[8].

First, the unconditional coverage test, introduced by Kupiec (1995) is applied, to test whether the indicator function has a constant success probability equal to the VaR significance level, a. The null hypothesis tested with $LR_{UC}$ is that the average number of VaR violations forecasted is correct. Therefore, a VaR model is rejected in either case that underestimates or overestimates the actual VaR. The likelihood ratio statistic $LR_{UC}$ is given by:

$$LR_{UC} = -2ln\left[\frac{(1-a)^{T_0} a^{T_1}}{(1-T_1/T)^{T_0}(T_1/T)^{T_1}}\right] \sim X^2(1) \tag{18}$$

where T is the out-of-sample days, $T_0\ and\ T_1$ are the number of 0s and 1s in the sample, and χ2 is the chi-squared distribution with one degree of freedom.

Second, the independence test is applied, to control for any clustering in the hit sequence which would indicate that the VaR model is not adequate in responding promptly to changing market conditions. The null hypothesis tested with $LR_{ind}$ is that the VaR violations forecasted are independent. To this end, the test should be able to reject a VaR model with clustered violations. The likelihood ratio statistic $LR_{ind}$ is given by:

$$LR_{ind} = -2ln\left[\frac{(1-T_1/T)^{T_0}(T_1/T)^{T_1}}{(1-\pi_{01})^{T_{00}}\pi_{01}{}^{T_{01}}(1-\pi_{11})^{T_{10}}\pi_{11}{}^{T_{11}}}\right] \sim X^2(1) \tag{19}$$

8 The smaller the significance level for the VaR estimates, the fewer the number of violations will be. Therefore, by choosing a 5% significance level more VaR violations can be observed than using a 1% level, leading to a better test for the accuracy of the VaR model.

where $T_{ij},\ i,j = 0,1$ is the number of observations with a j following an i. Also, $\pi_{01}$ and $\pi_{11}$ are given by the following equations:

$$\pi_{01} = \frac{T_{01}}{T_{00}+T_{01}} \tag{20}$$

$$\pi_{11} = \frac{T_{11}}{T_{10}+T_{11}} \tag{21}$$

Third, the conditional coverage test is applied, to simultaneously test whether the VaR violations are independent and that the average number of those violations is equal to n*a. The null hypothesis tested with $LR_{CC}$ is that both the average number of VaR violations forecasted is correct, and that the VaR violations are independent. It is important to test for conditional coverage because many financial and commodity time series exhibit volatility clustering. So, VaR estimates should be narrow (wide) in times of low (high) volatility, so that VaR violations are not clustered but spread-out over the sample period. The joint test of conditional coverage can be calculated as the sum of the two individual tests, so the likelihood ratio statistic $LR_{CC}$ is given by:

$$LR_{CC} = LR_{UC} + LR_{ind} \sim \mathcal{X}_1^2 \tag{22}$$

### *B. Economic evaluation*

In the second stage of the VaR models evaluation procedure the risk manager can work with fewer models, only those that pass all three statistical tests. However, because usually more than one model pass the first evaluation stage and the risk manager cannot choose a single VaR model as the most effective, an economic evaluation framework is needed to rank the models. Lopez (1999) and Sarma et al. (2003) set-forth such an evaluation approach by creating a loss function that measures the economic accuracy of the VaR models that pass the statistical tests. In this paper the approach introduced in Lopez (1999) and Sarma et al. (2003) is used, developing a loss function based on the notion of Expected Shortfall (ES), also termed Conditional VaR (CVaR), which measures the difference between the actual and the expected losses when actually a VaR violation occurs. A similar approach is also followed by Angelidis and Skiadopoulos (2008). Using this loss function the statistically accurate models are ranked and an economic utility function able to accommodate the risk manager's needs is specified as follows:

$$LF_i = \frac{1}{T}\sum_{j=1}^{T}\left[r_j - ES_i(a)\right]^2 \tag{23}$$

$$ES_a = \begin{cases} E[r_t|(r_t \leq -VaR_t(a))]; & \text{for long positions} \\ E[r_t|(r_t \geq VaR_t(a))]; & \text{for short positions} \end{cases} \tag{24}$$

where the ith ES is defined as the average loss over the VaR violations from the N out-of-sample violations that occurred for the ith VaR model, under the following conditions:

$$r_j - ES_i(a) = \begin{bmatrix} \begin{cases} 0, & if\ ES_i(a) \leq r_j \\ r_j - ES_i(a), & if\ r_j < ES_i(a) \end{cases}; & \text{for long positions} \\ \begin{cases} 0, & if\ ES_i(a) \geq r_j \\ r_j - ES_i(a), & if\ r_j > ES_i(a) \end{cases}; & \text{for short positions} \end{bmatrix} \tag{25}$$

The proposed LF uses the ES and not the VaR measures to compare with the actual returns, as the VaR returns do not give an indication about the size of the expected loss when a violation occurs. The model that minimizes the total loss, hence returns the lowest LF value, is preferred relative to the remaining models. Evidence in the literature shows that the ES is a more coherent risk measure than the VaR (Acerbi, 2002; Inui and Kijima, 2005). In addition, Yamai and Yoshiba (2005) argue that VaR is not as reliable as the ES measure, especially during market turmoil, and that it can be misleading for risk managers. However, the authors also suggest that the two measures should be combined for better results, as the ES estimations need to be very accurate in order to increase efficiency in the risk management process.

### *C. Selection process: Modified Diebold Mariano & Bootstrap Reality Check*

Amongst all VaR models that passed the first evaluation stage, the model with the lowest LF, calculated during the second evaluation stage, is used as the benchmark model in order to examine whether it statistically performs better than the competing models. First, the pair-wise model comparison methodology employed is the modified Diebold Mariano (MDM) test proposed by Harvey et al. (1997). This approach overcomes the limitation of the Diebold-Mariano (1995) test of frequently rejecting the null when it is true. Then, the values of the modified DM test are compared with the critical value of the Student's t-distribution with (T-1) degrees of freedom.

The null hypothesis of the MDM test is that both the benchmark and the competing models are equally accurate in their VaR forecasts. That is, $H_0: E(d_t) = 0$ with $d_t = LF_{1,t}^{MDM} - LF_{2,t}^{MDM}$. The MDM statistic and the loss function used to evaluate the models under this framework are the following:

$$MDM = \sqrt{\frac{T-1}{T}} \frac{\bar{d}}{\frac{\sum_{t=1}^{T}(d_t-\bar{d})^2}{T^2}} \tag{26}$$

$$LF_{i,t}^{MDM} = r_{i,t} - ES_{i,t} \tag{27}$$

where $t = 1, \ldots, T$, and $\bar{d} = \frac{\sum_{t=1}^{T} d_t}{T}$.

Second, in addition to the MDM evaluation method, to minimise the possibility that the performance amongst the competing VaR methodologies could be due to data snooping bias, the bootstrap version of White's (2000) Reality Check (RC) is implemented. According to Sullivan et al. (1999) and White (2000), data snooping occurs when a single data set is used for model selection and inference. While testing different models there is a probability of having a given set of results purely due to chance rather than these being truly based on the actual superior predictive ability of the competing models. In doing so, a relative performance measure is first constructed that can be defined as:

$$f_{k,n} = LF_{n,0} - LF_{n,k}; \quad k = 1,..,l; \ \mathrm{n} = 1,..,623 \tag{28}$$

where model 0 is the benchmark and k represents the kth model, n denotes the out-of-sample testing period, and LF is the loss function of equation (23) chosen in the previous section. Next, for each value of k and LF, pair wise comparisons are made between each portfolio and the remaining ones. Mathematically the null hypothesis for the reality check can be formulated as:

$$H_0 : \underset{k}{max}\left\{E\left(f_k\right)\right\} \leq 0. \tag{29}$$

The null hypothesis states that none of the models is better than the benchmark, i.e. there is no predictive superiority over the benchmark itself. Hence, whenever the null hypothesis is accepted it means that there is no competing model that performs better in terms of its VaR forecasting ability than the benchmark model. Following White (2000), the null hypothesis is tested by obtaining the test statistic of the reality check as $T_n^{RC} = \underset{k}{max}\left(n^{1/2}\bar{f}_k\right)$, where $\bar{f}_k = n^{-1}\sum_{t=1}^{n} f_{k,t}$ and n is the number of days of the out-of-sample period. To construct the test statistic, the stationary bootstrap technique of Politis and Romano (1994) is employed and B=1,000 random paths of VaR models' Loss Functions are generated. A similar approach is used by Alizadeh and Nomikos (2007) who applied the stationary bootstrap to approximate the empirical distribution of Sharpe ratios and test different trading rules in the sale and purchase market for ships.

The stationary bootstrap re-samples blocks of random length from the original data, to accommodate serial dependence, where the block length follows a geometric distribution and its mean value equals $1/q$. In this paper, similarly to Sullivan et al. (1999)[9], $q=0.1$ which corresponds to a mean block length of 10; for $q=1$ the problem is reduced to the ordinary bootstrap which is suitable for series of negligible or no dependence. Finally, the bootstrap loss function and thus the performance measure, is constructed by using the simulated loss functions, whereas the Bootstrap RC p-value is obtained by comparing $T_n^{RC}$ directly with the quantiles of the empirical distribution of $T_n^{RC*}$ using the following expression:

$$T_n^{RC*} = \max_k \left\{ n^{1/2} \left( \bar{f}_k^*(b) - \bar{f}_k \right) \right\} \tag{30}$$

where $\bar{f}_k^*(b)$ represents the sample mean of the relative performance measure calculated from the bth bootstrapped sample, with $b=1,...,B$.

With the proposed back-testing procedure, VaR forecasts can be more accurate, reducing the probability of accepting flawed models, and thus satisfying the requirements of stringent risk management control procedures. In addition, using the proposed economic utility function, the risk manager is able to rank a range of candidate VaR models and select the best performing one amongst them. Finally, the market players can be better informed, and thus well prepared to withstand any future losses, should the market moves to the opposite direction, by forecasting the ES measure more accurately.

## IV. Data and the spot energy index

This section describes the data used for the VaR models assessment. For each model, in total 2,450 daily observations are collected from DataStream for the period 12/09/2000 to 1/02/2010. From the total sample, 1,827 observations are used in estimation to forecast the next day's VaR. Using this "rolling window" method, for a fixed length of 1 day, the estimation sample is rolled over the entire data period generating 623 daily out-of-sample VaR forecasts. The spot prices collected are from eight energy markets that trade futures contracts on NYMEX, and the Spot Energy Index, as explained below:

Heating Oil, New York Harbour No.2 Fuel Oil, quoted in US Dollar Cents/Gallon (US C/Gal); hereafter named as "HO";

Crude Oil, West Texas Intermediate (WTI) Spot Cushing, quoted in US Dollars/Barrel (US$/BBL); hereafter named as "WTI";

Gasoline, New York Harbour Reformulated Blendstock for Oxygen Blending (RBOB), quoted in US C/Gal; hereafter named as "Gasoline";

[9] For more technical details on the implementation of the stationary bootstrap RC the reader is referred to Sullivan et al., 1999; Appendix C, pp 1689-1690.

1-1 Crack Spread of Gasoline with WTI, quoted in US $/BBL; hereafter named as "CS_Gasoline_WTI"[10];

1-1 Crack Spread of Fuel Oil with WTI, quoted in US $/BBL; hereafter named as "CS_HO_WTI";

Natural Gas, Henry Hub, quoted in US Dollars/Milion British Thermal Units (US$/MMBTU); hereafter named as "NG";

Propane, Mont Belvieu Texas, quoted in US C/Gal; hereafter named as "Propane";

PJM, Interconnection Electricity Firm On Peak Price Index, quoted in US Dollars/Megawatt hour (US $/Mwh); hereafter named as "PJM".

Geometric average Spot Energy Index, quoted in index points and constituted by daily prices of WTI, HO, Gasoline, NG, Propane, and PJM; hereafter named as "SEI".

The Spot Energy Index (SEI) is constructed as an un-weighted geometric average of the individual commodity ratios of current prices to the base period prices, set at September 12, 2000. The index's construction methodology is similar to that of the world-renowned CRB Spot Commodity Index. The SEI is designed to offer a timely and accurate representation of a long-only investment in energy commodities using a transparent and disciplined calculation. Geometric averaging provides a broad-based exposure to the six energy commodities, since no single commodity dominates the index. Also, through geometric averaging the SEI is continuously rebalanced which means that the index constantly decreases (increases) its exposure to the commodity markets that gain (decline) in value, thus avoiding the domination of extreme price movements of individual commodities. Moreover, the risks that other types of indexes are subject to can be avoided, like potential errors in data sources for production, consumption, liquidity, or other errors that could affect the component weights of the index.

The mathematical expression used to calculate the geometric average Spot Energy Index (SEI) is the following:

$$\mathrm{SEI}_t = \left( \prod_{i=1}^{n} \frac{S_t^{\,n}}{S_0^{\,n}} \right)^{\frac{1}{n}} \times 100 = \sqrt[n]{\frac{S_t^{\,1}}{S_0^{\,1}} \times \frac{S_t^{\,2}}{S_0^{\,2}} \times \ldots \frac{S_t^{\,n}}{S_0^{\,n}}} \times 100, \ \mathrm{n} = 1, 2, ..6. \tag{31}$$

where, $\mathrm{SEI}_t$ is the index for any given day, n represents each one of the six commodities comprising the index, $S_t^{\,n}$ is the price of each commodity for any given day, and $S_0^{\,n}$ is the average (geometric) price of each commodity in the base period.

All the commodity prices and the Spot Energy Index chosen represent a barometer of the energy market trends worldwide. Figure 1 shows the evolution of the logarithmic price series and their returns, over the whole period examined from 12/09/2000 to 1/02/2010. It is observed that all series follow an upward sloping time trend until the end of June, 2008 (WTI reached $145/barrel), followed by a steep downward slope until the end of December of the same year (WTI fell to $31/barrel). Then, for the remainder of the sample a small recovery of the prices is witnessed with WTI prices recovering and staying at the range of $70 - $80/barrel. In general, from the figure it can be inferred that all spot energy prices are quite volatile, with the two crack spreads with WTI, the Natural Gas and the PJM markets exhibiting more distinct price jumps. Furthermore, all series vary with time as it can be

[10] The spot series of the two 1-1 crack spreads with the WTI have been constructed after converting the Fuel Oil and Gasoline spot prices that are quoted in US C/gallon into US $/Barrel, taking into account that there are 42 gallons in one barrel and 100 cents per dollar. Then, the two series are rebased to 100 so they can later be transformed to logarithmic prices and apply our modelling methodology.

observed by the log-price differences, also forming clusters, both signs that indicate the presence of time-varying volatility.

Next, the descriptive statistics for the natural logarithm of the spot prices of all series are also estimated. To identify whether the series are mean reverting, a comparison procedure known as “confirmatory data analysis” is performed, where two tests for unit root non-stationarity, the Augmented Dickey-Fuller (ADF; Dickey and Fuller, 1979) and the Philips-Perron (PP; Phillips and Perron, 1988), and one test for stationarity, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS; Kwiatkowski et al., 1992), are employed. For the results to be robust, all three tests should give the same conclusion. Table 2 shows the descriptive statistics of the spot price series in logarithmic levels (Panel A) and their first differences (Panel B). As can be seen in panel B, the annualized volatility (as measured by the standard deviation of log-returns) of most energy markets ranges from 16% for the Heating Oil – WTI crack spread to 236% for PJM, which is significantly larger than the typical volatility observed in financial markets (e.g. the historical annualised volatility for the S&P500 is in the range of 20%-25%). As for the SEI, being an index, by construction its annualised volatility (48.5%) is in the same range as for the remaining fuel markets, WTI (41.9%), HO (42.4%), and Gasoline (50.5%), and significantly smaller than the highly volatile NG (75.4%). Overall, the two crack spreads have lower volatility than the outright series due to the high correlation between the prices of their constituent contracts.

Looking at panel A of table 2, is observed that for all energy markets, with the exception of NG, Propane, and the SEI, the skewness is positive, indicating that extreme high values are more probable than low ones. Turning next to the log-price changes, the results regarding the coefficients of skewness are different since only the Heating Oil, Gasoline, and the crack spread of WTI with Gasoline are negatively skewed, whereas the rest of the energy markets are positively skewed (see panel B, table 2). Also looking again in Panel B of table 2, the coefficient of kurtosis which gives an indication of the probability of extreme values, is above three for all energy markets, implying that log-returns are leptokurtic; this suggests that the probability of extremely high or low returns is much higher than that assumed by the normal distribution. This effect is more obvious for the PJM, NG, the two crack spreads, and Propane in which case the high values of the coefficient of kurtosis (between 12.07 and 34.53) is indicative of spikes in the price series. It is also found that normality is overwhelmingly rejected in the first difference series for all the energy markets and the SEI, on the basis of the Jarque-Bera (1980) test which is significant at the 1% level. It is obvious that non-normality occurs mostly due to the large price movements and spikes in all logarithmic price series that eventually lead to fat tails.

Moreover, from panel A in table 2 it is observed that the average logarithmic price for most energy markets is reduced when the filtered series is examined (i.e. when jumps are excluded) indicating that jumps have a positive impact on log-prices[11]. The only exceptions are the WTI and Gasoline markets where jumps have a negative impact on log-prices. It can also be inferred that the price-levels of most energy markets are not stationary, a conclusion confirmed by all three tests; the only exceptions are, as expected, the two crack-spreads and the PJM markets where price levels appear to be stationary on the basis of the ADF and PP tests. On the other hand, from Panel B of table 2 it can be seen that the first differences of the spot log-price series are strongly stationary for all energy markets, indicating the presence of mean reversion in the series. This conclusion, although it may not have been expected due to the presence of jumps in most of the energy series, can be justified by the fact that these jumps do not seem to affect the stationarity of the series because they are short-lived and price levels eventually revert to their mean after a jump has occurred. Panel B also reports the Ljung-Box (1978) Q(k)-statistic and Engle’s (1982) ARCH test (Q2(k)-statistic) to test the significance of autocorrelation in the returns and squared returns for lags one and 20, respectively. From the reported values there is evidence of serial correlation for all the log-return series, and for both time lags, at conventional

[11] A detailed discussion on how the filtered series is estimated is given in the following section.

significance levels; the only exception is for the Gasoline market for 20 lags. Finally, based on Engle's ARCH test significant serial correlation in the squared log-returns of all energy markets and the SEI is found, which indicates the presence of time-varying volatility in the return series.

## V. Empirical analysis

To evaluate the efficiency of all available VaR models, out-of-sample 99%[12] one-day VaR forecasts are generated for each one of the energy commodities examined and the SEI. The period used to estimate the parametric VaR models is the 12/09/2000 to 12/09/2007 consisting of 1827 observations, whereas the period used for the 623 out-of-sample forecasts is the 13/09/2007 to 1/02/2010. This research contributes to the relevant literature by testing all the VaR models for both long and short trading positions undertaken by the energy market players. As Angelidis and Degiannakis (2005) argue, it is imperative that a risk manager is able to forecast accurately the VaR for both long and short trading positions. In total, twenty two VaR models are implemented on the energy spot price series and the Spot Energy index, as described previously.

The VaR results for all applied models and for all energy commodities and the SEI are shown in tables 3 to 11, for the 1% significance level. Each table reports, for both long and short positions, the average VaR or Expected Tail Loss in percentage points, the frequency of violations or number of hits in percentage points, alongside the p-values for Christoffersen's three statistical tests for unconditional coverage, independence, and conditional coverage. The models that pass each test at the 5% significance level, and thus do not reject the null hypothesis, are indicated in bold. A 5% significance level is chosen in this paper as the acceptance threshold for the three tests, because the smaller the significance level the fewer the number of violations is, which leads to larger Type II errors that can be very costly for the risk manager. In addition, the results from the second evaluation stage, i.e. the Expected Shortfall, and the Loss Function that measures the economic accuracy of the models, are reported for both the short and long positions. The model that minimizes the total loss, hence returns the lowest LF value, is preferred relative to the remaining models. The numbers indicated in bold represent the models that have successfully passed all three statistical tests, whereas an asterisk indicates in each case the model that provides the smallest LF value and that is later used in the MDM pair-wise comparison as the benchmark model. The economic evaluation framework that uses the proposed LF can provide useful information for evaluating the VaR estimates for regulatory purposes. That is because by using the ES measure in the LF, the additional information on the magnitude of a loss that exceeds the estimated VaR is incorporated into the evaluation process. In addition, with the use of the proposed LF, the risk manager is able to rank all the candidate VaR models and distinguish the best performing one amongst them.

From tables 3 to 11 it can be seen that for all commodities and the SEI there is always at least one model that passes all three statistical tests at the 1% significance level, for both long and short trading positions. In the majority of cases, it is the MC simulation and the proposed Hybrid MC-HS models that successfully pass the first evaluation stage, thus overall prevailing against the more traditional ARCH type and Historical Simulation methodologies. Even though in some cases the MC simulation models do not pass all three statistical tests, they tend to produce the lowest LF values, followed by the Hybrid MC-HS models. Due to the economic importance of the LF for the risk manager, it can be argued that even for those energy commodities that the simulation-type models do not pass the statistical tests, they can still be considered as good alternative methodologies for estimating VaR. When the frequency of hits is zero the respective models are unsuitable candidates for the application of both the statistical and the economic evaluation tests; these cases are indicated by a dash line in all tables. In addition, in those cases where the frequency of hits is too high, above 20%, the respective

---

[12] 95% one-day VaR forecasts are also calculated but are not reported because results are very similar with the 99% forecasts that are reported in the tables.

models are unsuitable candidates for the application of the two statistical tests for unconditional and conditional coverage; in these cases a dash is also inserted. However, this does not mean that these models should be immediately rejected but it should be noted that consistently overestimate in the former case, and underestimate in the latter case, the actual VaR. For the entire fuels complex, including the WTI, HO, Gasoline, and the crack spreads with WTI, and for both long and short positions, the MC simulations methodology under the MRJD specifications, is the one that manages to pass all three statistical criteria from the first evaluation stage, and at the same time to deliver the lowest LF at the second evaluation stage. The only exceptions are the WTI and the CS-HO-WTI just for the long trading positions, with the F-EGARCH and F-GARCH methodologies delivering the lowest LFs respectively. As for the PJM and the SEI, and for both the long and the short trading positions, it is the Hybrid MC-HS specifications that successfully pass the first evaluation stage and deliver the lowest LF values at the second evaluation stage. Finally, the VaR for both the NG and the Propane series, for the long positions, is best estimated by the F-EGARCH methodology, whereas for the short positions is best estimated with the Hybrid MC-HS and the GARCH methodologies respectively.

Next, table 12 reports the p-values for the pair-wise modified Diebold-Mariano (MDM) test, between the model that delivers the smallest LF and all those models that pass the first evaluation stage, for the long and the short trading positions, respectively. The null hypothesis of the MDM test is that both the benchmark and the competing models are equally accurate in their VaR forecasts. The null hypothesis is rejected whenever the reported p-value is less than 1%[13]. An asterisk indicates that the competing models are statistically performing equally well for predicting VaR, whereas a double asterisk indicates that the VaR hit series for both competing models is identical and so there cannot be any differentiation between the two. In such cases the p-value is equal to 1 as the null is accepted with 100% confidence. As far as the long trading positions are concerned, according to the reported p-values and for $\alpha$=1%, it is for the WTI and the Propane markets that the F-EGARCH is statistically superior as a stand-alone model relative to the competing models, and for the HO market that the MCS-MRJD-GARCH model stands out. For all remaining energy markets and the SEI, all the pair wise competing models perform statistically equally well with the model that delivered the lowest LF at the second evaluation stage. In some cases the two competing models are statistically identical, as is the case for example with PJM and the SEI where the benchmarks HMCS-MR-GARCH and HMCS-MR-EGARCH when compared with the HS and the HMCS-MR-GARCH, respectively, seem to be delivering exactly the same statistical accuracy. As far as the short trading positions are concerned and for all energy commodities and the SEI, according to the respective p-values, the null hypothesis cannot be rejected for all competing pairs of models. Again, there are many cases that the two competing models behave statistically the same. For example in the case of the SEI and for the benchmark HMCS-MR-EGARCH model, the null that the two competing models are the same, is respectively accepted with 100% confidence for the comparisons with the F-HS, HS, HMCS-MR-GARCH, and HMCS-MR-OLS models.

In addition, table 13 reports the p-values for the White's (2000) Reality Check (RC) test, between the model that delivers the smallest LF (benchmark) and all those models that pass the first evaluation stage, for both long and short trading positions. The null hypothesis states that none of the models is better than the benchmark, i.e. there is no predictive superiority over the benchmark itself. Hence, whenever the null hypothesis is accepted it means that there is no competing model that performs better in terms of its VaR forecasting ability than the benchmark model. The null hypothesis is rejected whenever the reported p-value is less than the conventional level of significance of 1%. For the long positions, the null cannot be accepted for Gasoline, the crack spread of HO with WTI, PJM, and the SEI as there can be at least one model that performs equally well or better than the benchmark model. For the WTI, NG, Propane, the crack spread of Gasoline with WTI, and HO markets there is

[13] The relevant t-stats (MDM-statistics) are also calculated but are not reported in the table because in every case, the outcome is identical to that of the p-values.

strong evidence that the benchmark model is indeed the best in terms of its VaR performance across the competing models; the F-EGARCH for the former four markets and the MCS-MRJD-GARCH for the latter. As for the short positions, the null cannot be rejected in all cases but three. It is only for the WTI, HO and Gasoline that the benchmark model is not the best performing one according to the reported RC p-values. On the other hand, based on the reported RC p-values, for the two crack spreads of Gasoline and HO with WTI, the Propane, NG, PJM, and the SEI, the benchmark model is indeed the best performing one; that is the MCS-MRJD-EGARCH, MCS-MRJD-GARCH, GARCH, and HMCS-MR-EGARCH for the latter three markets respectively. The results from the RC test indicate that for the long trading positions there is mixed evidence as to which model performs better in terms of its VaR forecasting ability. However, for the short trading positions it is clearer from the results that the proposed MC Simulation and the Hybrid MC-HS methodologies produce a better VaR performance compared to the more traditional ARCH type and Historical Simulation methodologies.

Finally, table 14 summarises the VaR models that have been shortlisted as being the best for predicting VaR for each energy market and the SEI, following the proposed back-testing methodology. Panels A and B show the results for the long and the short trading positions respectively. In both panels, the first two columns list all the models that have successfully passed all three statistical tests, i.e. the first evaluation stage. Next, the remaining columns in each panel report only those VaR models that deliver the lowest LF, alongside those models that the MDM test identifies that their hit series is identical. According to the implemented two stage back-testing procedure, at the 1% significance level and for the short positions, it is the MC simulation and the Hybrid MC-HS methods from which the preferred models for estimating the VaR are short-listed; this finding is consistent with all energy markets and the SEI. As for the long trading positions results are mixed. On the one hand, it is again the MC simulation and the Hybrid MC-HS methods that are the best choices for the HO, Gasoline, CS-Gasoline-WTI, PJM, and the SEI. On the other hand, it is the ARCH-type models, and more specific the F-GARCH and F-EGARCH models, that stand out as the best VaR modelling options for the WTI, CS-HO-WTI, NG, and Propane markets.

Therefore, whenever a risk manager wants to choose a single approach for calculating the VaR for all energy commodities that he/ she holds, as it is usually the case in practice, the results show that the MC simulations and the Hybrid MC-HS approaches proposed in this paper are the most reasonable, efficient, and consistent candidates. The findings of this research have important implications for regulatory and policy-making purposes as the decision making bodies can reconsider the commonly used VaR models and establish an industry-wide methodological approach for calculating and back-testing the VaR in the energy markets. The proposed MC simulation and the Hybrid MC-HS models, in combination with the proposed selection procedure, have the potential of becoming common practice in the energy industry.

## VI. Conclusion

This paper proposes and compares a set of models for estimating the VaR of eight spot energy markets that trade futures contracts on NYMEX, and of the constructed Spot Energy Index, for both long and short trading positions, at the 1% significance level. The two proposed VaR methodologies are a MC simulation approach, and a Hybrid MC with Historical Simulation approach, both assuming various processes for the underlying spot prices. Next, a two-stage evaluation and selection process is applied, combining statistical and economic measures, to choose amongst the competing VaR models. The results show that, at the 1% significance level, for all commodities and the SEI there is at least one model that passes all three statistical tests with the ARCH type, the MC simulation, and the Hybrid MC-HS models prevailing more. For the entire fuels complex, including the WTI, HO, Gasoline, and the crack spreads with WTI, and for both long and short positions, the MC simulations methodology under the MRJD specifications, followed by the Hybrid MC-HS models pass all three statistical criteria from the first evaluation stage, and at the same time deliver the lowest LF at the second evaluation stage. The only exceptions are the WTI and the CS-HO-WTI just for the long trading positions, with the ARCH-type methodologies delivering the lowest LFs respectively. Therefore, it is concluded that the two former approaches are the most reasonable, efficient, and consistent candidates for calculating the VaR of energy prices, for both long and short positions.

The accurate calculation of VaR measures in the volatile energy markets is important for all market players and for a variety of reasons. First, the spot energy price risk is quantified taking into consideration the occurrence of extreme volatility events and thus at the same time allowing managers to develop efficient hedging strategies to protect their investments. Second, with the proposed VaR model selection process, modelling risk can be minimised as it satisfies strict risk management requirements and control procedures, by reducing the probability of accepting flawed models. Third, quantifying the risk profile of the energy markets, as expressed by the individual spot price series and the SEI, is vital for many hedge fund managers and alternative investors that have recently been following closely and started expanding their presence in the energy markets. Finally, the proposed VaR estimates can be used for setting the margin requirements in the growing energy derivatives market, and more importantly for the energy forwards, futures, and options that are widely used for both hedging and speculation purposes by many industrial players, commodity and investment houses.

The latter can be achieved by adopting the proposed models for their derivative contracts' valuations which are able to describe the energy markets better, exhibiting better explanatory power and goodness of fit. These models incorporate mean-reversion and spikes in the stochastic behaviour of the underlying asset, allowing for a different speed of mean reversion once a jump is identified, while at the same time allowing for time-varying volatility in their specification modelled as a GARCH or an EGARCH process. While risk management clearly did not fully prevent a downside in investment portfolios during the recent economic recession, according to Briand and Owyong (2009) those organisations that had invested in risk management practices prior to the crisis, and acted on their findings, performed significantly better than those that did not.

## Appendix

All tables and graphs can be provided by the authors upon request.

## References

Acerbi, C., 2002. Spectral measures of risk: a coherent representation of subjective risk aversion, Journal of Banking and Finance, 26 (7), 1505-1518.

Alexander, C., 2008. Statistical models of operational loss. In Handbook of Finance, 1, F.J. Fabozzi (ed.), Wiley.

Alizadeh, A. and Nomikos, N., 2007. Investment timing and trading strategies in the sale and purchase market for ships. Transportation Research Part B: Methodological, 41 (1), 126-143.

Aloui, C. and Mabrouk, S., 2010. Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models, Energy Policy, 38, 2326-2339.

Angelidis, T. and Degiannakis, S., 2005. Modelling risk in three markets: VaR methods for long and short trading positions, Journal of Risk Finance, 6 (3), 226-238.

Angelidis, T. and Skiadopoulos, G., 2008. Measuring the market risk of freight rates: a value-at-risk approach. International Journal of Theoretical and Applied Finance, 11 (5), 447-469.

Bams, D., Lehnert, T., and Wolff, C.C.P., 2005. An evaluation framework for alternative VaR models, Journal of International Money and Finance, 24, 944-958.

Bierbrauer, M., Menn, C, Rachev, S.T., Truck, S., 2007. Spot and derivative pricing in the EEX power market. Journal of Banking and Finance 31; 3462-3485.

Boudoukh, J., Richardson, M. and Whitelaw, R., 1998. The best of both worlds, Risk, 11, 64-67.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics **31;** 307–327.

Black, F., 1976. The pricing of commodity contracts. Journal of Financial Economics 3, 167-179.

Briand, R. and Owyong, D., 2009. How to kill a black swan: Risk and asset allocation in crises, Journal of Indexes, July/ August, 10-17.

Brooks, C. and Persand, G., 2003. Volatility forecasting for risk management, Journal of Forecasting, 22, 1-22.

Cabedo, J.D. and Moya, I., 2003. Estimating oil price Value at Risk using the historical simulation approach, Energy Economics, 25, 239-253.

Carol, A. Market Risk Analysis IV: Value-at-risk models. John Wiley & Sons, 2008.

Chiu, Y.C, Chuang, I.Y., and Lai, J.Y., 2010. The performance of composite forecast models of value-at-risk in the energy market, Energy Economics, 32, 423-431.

Christoffersen, P.F., 2003. Elements of Financial Risk Management. Elsevier Science, USA.

Christoffersen, Peter F., 1998. Evaluating Interval Forecasts. International Economic Review, 39, 841-862.

Clewlow, L., Strickland, C. Energy Derivatives Pricing and Risk Management. Lacima Publications (Published in association with Enron Corp.); 2000.

Duffie, D. and Pan, J., 1997. An overview of value at risk, Journal of Derivatives, 4, 7-49.

Dickey, D.A., W.A. Fuller, 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association 74; 427–431.

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

Dixit, A. K., Pindyck, R. S., 1994. Investment under uncertainty, Princeton University Press.

Engle, R. F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation. Econometrica 50; 987-1008.

Engle, R., 2001. Garch 101: The use of ARCH and GARCH models in Applied Econometrics, Journal of Economic Perspectives, 15, 157-168.

Geman, H., Roncoroni, A., 2006. Understanding the fine structure of electricity prices. The Journal of Business 79 (3); 1225-1262.

Giot, P. and Laurent, S., 2003. Market risk in commodity markets: A VaR approach, Energy Economics, 25, 435-457.

Glasserman, P., Heidelberger, P., and Shahabuddin, P., 2000. Efficient Monte Carlo methods for Value at Risk, Working Paper, Columbia University.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13 (2), 281-291.

Hendricks, D., 1996, Evaluation of value-at-risk models using historical data, Federal Reserve bank of New York, Economic Policy Review, 2, 39-70.

Hibon, M. and Evgeniou, T., 2005. To combine or not to combine: selecting among forecasts and their combinations, International Journal of Forecasting, 21, 15-24.

Huang, A.Y., 2010. An optimization process in Value-at-Risk estimation, Review of Financial Economics, 19 (3), 109-116.

Huang, J.C., Lee, M.C., and Liu, H.C., 2008. Estimation of value-at-risk for energy commodities via fat-tailed GARCH models, Energy Economics, 30, 1173-1191.

Hull, J., White, A., 1998. Incorporating volatility updating into the Historical Simulation method for Value at Risk, Journal of Risk, 1, 5-19.

Inui, K. and Kijima, M., 2005. On the significance of Expected Shortfall as a coherent risk measure, Journal of Banking and Finance, 29, 853-864.

Jamshidian, F., Zhu, Y., 1997. Scenario simulation: theory and methodology, Finance and Stochastics, 1, 43-67.

Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters 6 (3); 255–259; doi:10.1016/0165-1765(80)90024-5.

Jorion, P., 2006. Value at risk: The new benchmark for managing financial risk, 3rd edition: McGraw-Hill.

JP Morgan, 1996. RiskMetrics, Technical Document, New York.

Kanamura, T., 2009. A supply and demand based volatility model for energy prices. Energy Economics; doi: 10.1016/j.eneco.2009.04.001.

Knittel, C.R., Roberts, M.R., 2005. An empirical examination of restructured electricity prices. Energy Economics 27; 791-817.

Krehbiel, T. and Adkins, L.C., 2005. Price risk in the NYMEX energy complex: an extreme value approach, Journal of Futures Markets, 25 (4), 309-337.

Kuester, K., Mittnik, S., and Paolella, 2006. Value-at-Risk prediction: A comparison of alternative strategies, Journal of Financial Econometrics, 4, 53-89.

Kupiec, P., 1995. Techniques for verifying the accuracy of risk management models. Journal of Derivatives, 3, 73-84.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics 54; 159–178.

Lambadiaris, G., Papadopoulou, L, Skiadopoulos, G., and Zoulis, Y., 2003. VaR: History or Simulation?, Risk, 123-126.

Lopez, J.A., 1999. Regulatory evaluation of Value-at-Risk models. Journal of Risk, 1, 37-64.

Lucia, J.J., Schwartz, E.S., 2002. Electricity prices and power derivatives: evidence from the Nordic power exchange. Review of Derivatives Research 5; 5-50.

Ljung, G.M., Box, G.E.P., 1978. On a Measure of a Lack of Fit in Time Series Models, Biometrika 65, 297-303.

Marimoutou, V., Raggad, B., and Trabelsi, A., 2009. Extreme Value Theory and Value at Risk: Application to oil market, Energy Economics, 31 (4), 519-530.

Merton, R.C., 1976. Option Pricing when Underlying Stock Returns are Discontinuous. Journal of Financial Economics 3; 125-144.

Mittnik, S. and Paolella, M.S., 2000. Conditional density and value at risk prediction of Asian currency exchange rates, Journal of Forecasting, 19, 313-333.

Nelson, D. B., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59; 347-370.

Nomikos, N., Pouliasis, P., 2011. Forecasting petroleum futures markets volatility: The role of regimes and market conditions. Energy Economics, 33 (2), 321-337.

Nomikos, N., Soldatos, O., 2008. Using affine jump diffusion models for modeling and pricing electricity derivatives, Applied Mathematical Finance, 15 (1), 41-71.

Perignon, C. and Smith, D. R., 2010. The level and quality of Value-at-Risk disclosure by commercial banks, Journal of Banking and Finance, 34 (2), 362-377.

Phillips, P., Perron, P., 1988. Testing for a unit root in time series regressions. Biometrika 75; 335-346.

Pindyck, R. S., 1999. The Long-Run Evolution of Energy Prices. The Energy Journal 20 (2).

Pilipovic, D., 1998. Energy risk: Valuing and managing energy derivatives, McGraw-Hill.

Politis, D. N., Romano, J. P., 1994. The stationary bootstrap. Journal of The American Statistical Association 89 (428), 1303–1313.

Poon, S.H. and Granger, C.W.J., 2003. Forecasting volatility in financial markets: a review, Journal of Economic Literature, 41, 478-539.

Rogoff, K. and Reinhart, C., 2008. Is the 2007 U.S. sub-prime financial crisis so different? An international historical comparison. American Economic Review, 98, 339-344.

Sadeghi, M. and Shavvalpour, S., 2006. Energy risk management and value at risk modelling, Energy Policy, 34, 3365-3373.

Sarma, M., Thomas, S., Shah, A., 2003. Selection of Value-at-Risk models. Journal of Forecasting 22, 337-358.

So, M., Chen, C., Lee, J., and Chang, Y., 2008. An empirical evaluation of fat-tailed distributions in modelling financial time series, Mathematics and Computers in Simulation, 77, 96-108.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. The Journal of Finance 54, 1647-1691.

Vlaar, P.J.G., 2000. Value at risk models for Dutch bond portfolios, Journal of Banking and Finance, 24, 1131-1154.

White, H., 2000. A reality check for data snooping. Econometrica 68 (5), 1097–1126.

Yamai, Y. and Yoshiba, T., 2005. Value at risk versus expected shortfall: a practical perspective, Journal of Banking and Finance, 29 (4), 997-1015.

Zikovic, S. and Filer, R., 2009. Hybrid historical simulation VaR and ES: performance in developed and emerging markets. CESIFO Working Paper (October).

**Authors contacts:**

**Kostas Andriosopoulos**

ESCP Europe Business School

527 Finchley Road, Hampstead

London NW3 7BG

UK

Email : kandriosopoulos@escpeurope.eu

**Nikos Nomikos**

Cass Business School

106 Bunhill Row

London EC1Y 8TZ

UK

Email: N.Nomikos@city.ac.uk

European
University
Institute

European University Press