MPRA
Munich Personal RePEc Archive

Effective energy commodities' risk management: Econometric modeling of price volatility

Halkos, George and Tzirivis, Apostolos

Department of Economics, University of Thessaly

December 2018

Online at https://mpra.ub.uni-muenchen.de/90781/
MPRA Paper No. 90781, posted 22 Dec 2018 12:55 UTC

Effective energy commodities' risk management: Econometric modeling of price volatility

George Halkos and Apostolos S. Tsirivis
Laboratory of Operations Research
Department of Economics, University of Thessaly

Abstract
The current study emphasizes on the importance of the development of an effective price risk management strategy regarding energy products, as a result of the high volatility of that particular market. The study provides a thorough investigation of the energy price volatility, through the use of GARCH type model variations and the Markov-Switching GARCH methodology, as they are presented in the most representative academic researches. A large number of GARCH type models are exhibited together with the methodology and all the econometric procedures and tests that are necessary for developing a robust and precise forecasting model regarding energy price volatility. Nevertheless, the present research moves another step forward, in an attempt to cover also the probability of potential shifts in the unconditional variance of the models due to the effect of economic crises and several unexpected geopolitical events into the energy market prices.

Keywords: Energy commodities, WTI oil, Brent oil, electricity, natural gas, gasoline, risk management, volatility modeling, ARCH-GARCH models, Markov-Switching GARCH models.

JEL Classifications: C01; C58; D8; G3; O13; P28; Q43; Q47; Q58.

1. Introduction

Energy risk always composed a major factor for most firms involved in almost every key industrial sector in both developed and developing countries. However, the severe effects in energy market stemming from the existence of several unexpected economic and geopolitical events, have made risk management of energy products a crucial issue for any industrial firm, as it can seriously affect its competitiveness, viability and future profitability. Taking a close look at a number of such events it does worth noticing the extreme increase in price volatility of energy sources.

Especially during the last two decades there was a significant rise in the number as well as the intense of which these events affected the energy commodity market. With some of the most characteristic examples being the massive production and growth rates achieved by two gigantic developing economies like China and India, the military conflicts and the political and general instability in many oil producing countries in the Middle East, and most importantly the extended financial and debt crisis of 2007.

Each of the above factors affected in a different but equally essential way the need and supply capability for some of the most widely used energy commodities such as oil, natural gas and coal. From the early 2000's the rising necessity for energy sources, by developing countries maintaining a high annual growth, as well as the future expectation for high growth rates and the global rising consumption levels, steadily increased the demand and hence both the actual prices as well as the prices of derivative contracts (futures, forward, options, swaps) regarding energy sources and particularly oil.

In addition, the so called ‘Arabic spring’ led to continuous military conflicts and violent changes in the established political regimes and the general status-quo in many oil supplying countries in South-East Mediterranean and the Middle-East, destroying valuable infrastructure and significantly hurt the global oil supply capacity. Consequently, the cost of energy further boosted, with the prices of oil and its refinery products and a number of other main energy products reaching historical high levels.

Additionally, the longest economic crisis in modern history had severe side effects on the worldwide economy, considerably lowering the tendency to consume and thus industrial production. As a result, the growth rate for the vast majority of the countries reached zero levels or even turned negative, while that of countries with gigantic mass production like China was noticeably moderated. This led to a steep decline in the demand and hence the prices of energy products, reaching a record low and remained close to these levels until nowadays.

Under normal conditions, in an efficient and rational energy market the size of the imbalance between supply and demand would be the only factor influencing the level of price change. However, in real life there are several other factors disturbing market normality. The most crucial factors driving energy price volatility are structural and it is most likely to have a long term impact. Nevertheless, a number of significant unexpected events, such as natural disasters (Halkos and Zisiadou, 2018a, b), political uncertainty and corporate misconduct, intensify the uncertainty in energy market. As a result, this highly volatile and complex environment becomes attractive to financial speculators and hedge funds, magnifying the danger deriving from the main risk sources related to energy products and most importantly the risk from market price changes (*Price risk*).

The aim of this study is to focus on revealing the vital importance of price risk management strategies regarding energy commodities, while at the same time present the most appropriate methods and models to cope with this essential issue according to a large number of relative academic studies. Specifically, a lot of comprehensive research has been done in the field of risk management in the energy commodities markets especially during the last twenty years, due to its significance for both energy producers and heavily dependent industries and consumers, as well as the increase and in the volatility and the complexity of that particular market throughout that period. Energy commodity producers, large energy consuming industrial firms, traders, refiners and energy investors focused on developing the essential technical tools to regularly monitor and minimize their overall price risk market exposure, while at the same time build the optimal strategy which would allow them to maximize their profitability given a certain acceptable amount of risk.

As a result, many financial consulting firms and researchers were motivated to get closely involved and find ways to cope with the certain issue. The basic approach behind all research done in this field was to appropriately modify traditional financial risk management tools, in order to take account of the unique characteristic features of the energy commodities' market.

Another key issue is also the econometric methodology, which is going to be implemented in order to lead in accurate and efficient estimates of the volatility, with the largest part of academic researches relying on the ARCH (Engle, 1982) and GARCH family models (Bollerslev, 1986) and its expansions. Conventional econometric models are less capable to model and forecast volatility related to assets that show significant time-varying volatility. In contrast, ARCH based models by definition allow for time-varying volatility, enabling them to better fit time series data of energy commodities (Halkos and Sepetis, 2007). As a result, a significant number of academic papers are based on these models and especially the GARCH model, as in most cases the authors try to build a new improved model for predicting the risk exposure by creating an alternation of this basic model.

During the last twenty years a series of academic researches produced several GARCH based models showing remarkable results in estimating the risk exposure of the position or the portfolio of a participant in the energy market. Every single model differentiates in modeling and forecasting volatility, in terms of its ability to take into account special characteristics and phenomena that are present in the energy commodity market, such as the asymmetry and the long memory volatility effect, as well as its potential to examine the interactions between different markets affecting the price of a specific energy commodity and the correlations between several energy products in a specific portfolio.

Consequently, these models are categorized by researchers into several different groups, such as linear and non-linear, symmetric and asymmetric and univariate and multivariate models. Furthermore, specific tests have been developed in order to check for the precision and efficiency of these models, revealing which one of the examined models is the most appropriate to investigate the risk exposure using the data sample of certain energy commodities. On the other hand, in case that the investigated period is characterized by extreme uncertainty due to an economic turmoil or a serious geopolitical event, researchers turn to the Markov-Switching GARCH models to account for possible regime shifts of the variance of the examined energy products, as in that case there would be an upward bias in the volatility persistence parameters, hence in the overall forecasted volatility of the model.

The current study tries to provide the reader with a comprehensive examination of the academic research that has been done in the field of risk management and risk modeling regarding the energy market. However, in harmony with the majority of the researches it mainly focuses on three key energy commodities due to its vital economic importance, which are the crude oil, natural gas and electricity. In the present work, an analytic presentation on the most representative models and methodologies in this specialized scientific field will be attempted.[1]

2. Econometric modeling of energy commodity price volatility

A large econometric model framework has been developed by researchers in an attempt to create a series of econometric models, which are most suitable for fitting energy price volatility for as many energy commodities as possible and provide trustworthy future predictions in this highly volatile market. This review will focus on the most widely acceptable and most commonly examined volatility models by academic researchers in this field of economics and particularly the GARCH class models.

Generally, in financial markets the asset return volatility is specified as the conditional variance. However, because time series data of financial assets show significant time-varying volatility traditional econometric models are less appropriate to model and forecast volatility compared to the ARCH family models (Halkos and Papadamou 2006, 2007). As a result, these models are even more appropriate to model the volatility of energy commodities price data, which exhibit an even more intensive time-varying and periodic volatility. Panas and Ninni (2000) confirm this statement, as they discovered that the returns of an energy commodity of key

[1] To account for various stylized facts in volatility modelling and forecasting in finance (fat tails, volatility clustering, leverage effects, long memory and co-movements in volatility) apart from ARCH and regime switching models, stochastic volatility and autoregressive moving average models (ARMA) are also used (Xiao and Aydemir, 2007).

economic importance such as oil and its refined products, show strong evidence of time-varying volatility, skewness and kurtosis, clearly favoring the use of ARCH and GARCH type models for modeling and forecasting price and return volatility.

Specifically, regarding oil Morana (2001) and Giot and Laurent (2003) using price data from the oil markets fount clear signs of volatility clustering combined with fat tails and negative skewness in the return distribution. Furthermore, Khindanova and Atakhanova (2002) analyzing data concerning one month futures returns for WTI oil, Brent oil, natural gas, gas oil and electricity revealed that for the total of the examined energy commodities there is a clear indication for fat tails and negative skewness in their return distributions. Finally, Lanza et al. (2004) processing a data sample for forwards' and futures' WTI oil returns from 1985 to 2004, they confirmed the persistent presence of both the asymmetric and long memory volatility effects.

As a result, researchers tried to further develop the original GARCH model in ways so that to produce more efficient and reliable models that would be better able to deal with the growing uncertainty and the unique characteristics of the energy commodity market. This highly volatile environment was the result of multiple political and economic unexpected events, crucially affecting the crude oil prices, as well as the prices of its substitute products like natural gas. Additionally, the deregulation of the electricity market further intensified this phenomenon, due to the vital importance of electricity for millions of households and various industries which use electricity as a basic energy input.

There are several ways to categorize GARCH class model further into groups, relative to their main alternation from the basic GARCH model, as well as their unique characteristic features and the advantages that these enhance to the model. A common classification that is made by most researchers is between linear and non-linear GARCH models. Linear models, such as GARCH and IGARCH, propose a linear model setting of the conditional variance, assuming that the conditional variance is a linear function of past squared returns. Nevertheless, linear GARCH models are unable to take into account significant features of the financial markets, such as the asymmetric volatility effect. The leverage effect appears to be present in many financial assets, as negative shocks tend to have a larger impact on the conditional volatility. Consequently, a number of non-linear GARCH models has been developed by researchers, such as the EGARCH, the GJR GARCH, the FIGARCH, the FIAPARCH, the HYGARCH models and others, which are better able to capture the above phenomenon as well as the long and short memory volatility effects.

Another typical categorization of GARCH class model is between univariate and multivariate models. Traditional univariate GARCH models like the above mentioned may have the ability to detect and estimate the volatility dynamics of a single financial asset, yet they lack the capacity to reveal any potential correlations between different energy markets or interaction between energy and financial markets. For this reason, researchers targeting the above weakness of univariate models, they have developed a series of multivariate models such as the VEC and BEKK and the CCC and DCC models that can deal with market correlations and volatility spillovers, while providing important information regarding energy portfolio volatility and successful hedging strategies.

Finally researchers also separate GARCH models between symmetric and asymmetric models. Symmetric models are alternations of the basic GARCH model in terms of the distributional assumption and were developed in an attempt to mitigate the weaknesses from the normal error distribution assumption of the basic model.

Some characteristic symmetric models are the GARCH-t model (Bollerslev, 1986) in which the normal distribution was replaced by the student-t distribution and the GARCH-HT model (Politis, 2004) which uses the Heavy-Tailed distribution instead of the normal distribution. In contrast, as asymmetric models can be considered all the models that are able to capture the asymmetric volatility or leverage effect. This group of GARCH based models basically has the same characteristics and consists of the same models that comprise the non-linear GARCH models group.

2.1 GARCH family volatility models

2.1.1 GARCH model

The fact that most energy commodity time series data sets (both actual product price and especially its financial derivative prices) show significant autocorrelation and volatility clustering, makes the generalized autoregressive conditional heteroskedasticity model (GARCH) perhaps the most suitable and commonly used energy commodity volatility model among researchers. The model came as a development from Bollerslev (1986) to the original autoregressive conditional heteroskedasticity model (ARCH) of Engle (1982) and since then it sets the model framework foundation upon which several other volatility models were developed. GARCH models can be very effective and best fit energy price data as they are able to provide jointly estimates of the mean and the conditional variance (Halkos 2006, 2011).

The conditional mean for the GARCH (1,1) is given by the following equation:

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_{t*} z_t, \quad z_t \sim NID(0\ ,\ 1) \tag{1}$$

The conditional variance is given by:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{2}$$

While the one day future variance is estimated by:

$$\hat{\sigma}_{t+1}^2 = \hat{\omega} + \hat{\alpha}\hat{\varepsilon}_t^2 + \hat{\beta}\hat{\sigma}_t^2 \qquad (3)$$

Where r_t symbolizes the rate of return, the μ_t the conditional mean, σ_t^2 the conditional variance and z_t is the standardized error term. Additionally α, β, and ω are positive parameters, so that to ensure a positive conditional variance, with ω being a constant parameter. For α and β applies the restriction $\alpha + \beta < 1$, with $\alpha + \beta$ capturing the gradually reduced future effect of possible current volatility shocks.

2.1.2 IGARCH model

An interesting alternation of the basic GARCH(1,1) model is the IGARCH(1,1) model, which is a later development by Engle and Bollerslev (1986). The model differs to the GARCH(1,1) as it incorporates the parameter restriction of $\alpha+\beta=1$. As a result equation (2) can be written as:

$$\sigma_t^2 = \omega + \sigma_{\tau-1}^2 + \alpha\,(\varepsilon_{t-1}^2 - \sigma_{\tau-1}^2), \qquad 0 < \alpha \leq 1 \qquad (3)$$

Given that, $\varepsilon_t^2 = \sigma_t^2 * z_t^2$

$$\sigma_t^2 = \omega + \sigma_{t-1}^2\,[(1-\alpha) + \alpha z_{t-1}^2] \qquad (4)$$

As a result, when $\omega = 0$, σ_t^2 works basically as a martingale. In general, IGARCH(1,1) is a restricted special case of GARCH(1,1) model able to capture the persistent long memory effect of volatility shocks in volatility process. However, its biggest disadvantage in modelling the long memory volatility effect is that in IGARCH(1,1) the effect of volatility shocks remains as it is and does not gradually disappear like in the real market.

2.1.3 Exponential GARCH (EGARCH) model

A popular model for its ability to portray the asymmetric volatility effect is the EGARCH(1,1) model, which was first introduced by Nelson (1991) and was further developed by Bollerslev and Mikkelson (1996). The EGARCH model has the advantage not to be limited by the nonnegative parameter restriction of the linear GARCH(1,1) model, as the logarithm allows for the parameters to turn negative without the conditional variance becoming negative, with the α parameter measuring the volatility clustering effect, while an extra variable (γ) is added to control for the leverage effect.

The EGARCH(1,1) model is estimated based on the following equation:

$$\log(\sigma_t^2) = \omega + \alpha z_{t-1} + \gamma(|z_{t-1}| - E|z_{t-1}|) + \beta\log(\sigma_{\tau-1}^2) \qquad (5)$$

The EGARCH model provides the ability to the researcher to account for volatility persistence and mean reversion, while allowing for negative news to have a different impact on the volatility compared to positive news, enhancing the model with its most important advantage over the basic GARCH model.

2.1.4 GJR GARCH model

Another similar model to EGARCH is the GJR model, which is also capable of capturing the asymmetric volatility effect and was first developed by Glosten, Jagannathan and Runkle (1993). The two models have little differences and are difficult to set a rule under which it is better to choose the one over the other.

The conditional variance in the GJR(1,1) model is given by:

$$\sigma_t^2 = \omega + [\alpha + \gamma I(\varepsilon_{t-1} < 0)]\,\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \qquad (6)$$

Where $I(.)$ is an indicator function taking the value of 1, when the condition in ( ) is met or becomes 0 otherwise. However, the model has several parameter restrictions in order for the conditional variance to be positive, with ω being a strictly positive parameter and $\alpha \geq 0$, $\beta \geq 0$, $\gamma \geq 0$ and $\alpha+\gamma \geq 0$, with γ accounting for the asymmetric volatility effect.

2.1.5 Fractional Integrated GARCH (FIGARCH) model

In contrast with other univariate models like the IGARCH and the EGARCH in which it is assumed infinite persistence of volatility shocks and volatility autocorrelation decay at an exponential rate respectively, the fractionally integrated model (FIGARCH) which was developed by Baillie et al. (1996) is a more flexible and realistic model. In this model volatility shocks are considered to have a finite effect on the conditional volatility, allowing for volatility shocks to decay at a slow hyperbolic rate, while taking into account the individual long memory effect of every single shock.

The conditional variance in the FIGARCH model is estimated by:

$$\sigma_t^2 = \omega + \beta\sigma_{\tau-1}^2 + [1- (1-\beta L)^{-1} (1-\varphi L) (1-L)^d]\, \varepsilon_t^2, \tag{7}$$

Where d is the fractionally integrated parameter for which applies $0 \leq d \leq 1$, while L is the lag operator. For φ, ω, β, applies $\omega > 0$, and φ, $\beta < 1$ respectively. The FIGARCH(1, d, 1) model basically becomes similar to a GARCH(1,1) model for d=0 (Geometric decay) and equal to an IGARCH model when d=1 (Infinite persistence). Parameter d measures the long memory volatility effect of a shock giving the advantage to the model to allow and examine for intermediate ranges of persistence for $0 \leq d \leq 1$. Finally, according to Baillie et al. (1996) and later academic researchers who incorporated the model in to their researches, the model requires a minimum of 1000 observations to provide robust and reliable results. This limitation derives from the truncation order of the fractional differencing operator $(1-L)^d$.

2.1.6 Fractionally integrated asymmetric power (FIAPARCH) model

The FIGARCH model was further improved by TSE (1998), developing the fractionally integrated asymmetric power ARCH model **(**FIAPARCH), which contains all the features and advantages of a FIGARCH model, while at the same time has the extra ability to account for asymmetry in the conditional variance. TSE basically modified the original FIGARCH model by incorporating the $(|\varepsilon_t| - \gamma\varepsilon_t)^{\delta}$ function of the APARCH model.

The conditional variance in the FIAPARCH(1, d, 1) model can be estimated by the following equation:

$$\sigma_t^{\delta} = \omega\,(1-\beta)^{-1} + [1-(1-\beta L)^{-1}(1-\varphi L)(1-L)^{d}]\,(|\varepsilon_t| - \gamma\varepsilon_t)^{\delta}, \qquad (8)$$

For parameters d, ω, δ, φ, β applies that $0 \leq d \leq 1$, $\omega, \delta > 0$, $\varphi, \beta < 1$ and $-1 < \gamma < 1$. Additionally, when $0<d<1$ the long memory volatility effect is present, while when $\gamma>0$ negative news have a larger impact on volatility relative to positive news. Finally, for values $\gamma=0$ and $\delta=2$ the FIAPARCH models becomes equal to a FIGARCH model. Taking into consideration all the advantages of the model the FIAPARCH model appears to have all the potential to fit the unique characteristics that are present in most time series energy commodity data.

2.1.7 Hyperbolic GARCH (HYGARCH) model

One last univariate model that was more recently developed and is widely used by academic researchers is the hyperbolic GARCH (HYGARCH) model. The model was first introduced by Davidson (2004) and is basically an extension of the FIGARCH process.

The conditional variance in the HYGARCH(1, d, 1) model is estimated by:

$$\sigma_t^2 = \omega + \{1-(1-\beta L)^{-1}\,\varphi L\{1+\alpha[(1-L)^{d} - 1]\}\}\,\varepsilon_t^2, \qquad (9)$$

Where $0 \leq d \leq 1$, $\omega > 0$, φ, $\beta < 1$, and $\alpha \geq 0$. The HYAGARCH model becomes equal to a FIGARCH model for $\alpha=1$ and equal to a stable GARCH model for $\alpha=0$. Despite the model's ability to take into consideration both volatility clustering and long memory volatility effect, HYGARCH model has a significant shortcoming, which has to do with its weakness to account for asymmetry.

2.1.8 VEC GARCH model

A rather interesting type of GARCH class models is the multivariate models like the VEC GARCH model, which allows to measure and forecast the volatility of a portfolio of energy commodities as well as to investigate the interactions between different energy markets and their contribution to the price volatility of a particular energy commodity. The VEC model was first introduced by Bollerslev, Engle, and Wooldridge (1988) and is basically the most general multivariate model, as it can be considered as a straightforward extension of the basic univariate GARCH model.

The VEC model is estimated as follows:

$$\mathrm{VECH}\,(H_t) = C + \sum_{i=1}^{q} A_i \,\mathrm{vech}\,(\varepsilon_{t-i}\,\varepsilon'_{t-i}) + \sum_{j=1}^{p} B_j \,\mathrm{vech}\,(H_{t-j}), \tag{10}$$

All conditional variances and covariances are functions of their own lagged values, along with lagged squared returns and cross-products of returns. Vech(.) denotes an operator, stacking the columns of the lower triangular elements of its suggested square matrix, while C is an $N(N + 1)/2 \times 1$ vector and the A_i, B_j are $[N(N + 1)/2 \times N(N + 1)/2]$ parameter matrices.

The model has the advantage to be rather simple and flexible, however it is accompanied by some serious drawbacks and limitations, as the H_t is necessary to remain positive for all ε_t, in order to reasonably estimate all the parameters that are specified by the model, however this can be rather difficult to investigate.

Furthermore, the large number of the required parameters, as well as the rather demanding computational time critically limits the model not to be easy to consider more than two basic factors and as a result it is limited to a bivariate model.

Researchers developing the VECH model, realized both its difficulty to implement and its various limitations tried to improve the model by developing the Diagonal VECH model, which differentiated from the initial model by incorporating the assumption that A_i and B_i in (10) are diagonal matrices. The Diagonal VECH provides the ability to ensure that H_t is positively defined for all ε_t, while it is less demanding to estimate as the two equations can be calculated separately from one another, however it shows a significant weakness as it does not any interaction between each conditional variance and covariance.

2.1.9 BEKK GARCH model

The BEKK model was named after its developers Baba, Engle, Kraft and Kroner (BEKK) (1995) and can be considered as a special case of a VEC model. The model was built in such way to deal with two major drawbacks of the VEC model, which are the difficulty to verify whether Ht is always taking positive values and the large number of required parameters that need to be estimated.

The BEKK model is defined as follows:

$$H_t = CC' + \sum_{k=1}^{K} \sum_{i=1}^{q} A_{ik}\, \varepsilon_{t-i} \varepsilon'_{t-i} A'_{ik} + \sum_{k=1}^{K} \sum_{i=1}^{p} B_{ik}\, H_{t-i} B'_{ik}, \qquad (11)$$

Where A_{ik} , B_{ik} , and C are N × N parameter matrices, with C being decomposed in to two triangular matrices in order to guarantee that H_t will be positive at all times. However, the number of parameters in the full BEKK model still remains high with $(p+q)KN^2+N*(N+1)/2$ parameters required to be estimated. Nevertheless, by assuming that the A_{ik} and B_{ik} are diagonal matrices, the diagonal BEKK model can be estimated instead, decreasing the number of estimated parameters to (p+q)KN +N*(N+1)/2.

An additional alternative to decrease the number of estimated parameters is the Scalar BEKK model, which is created by implementing the assumption that A_{ik} and B_{ik} are equal to the result of a multiplication of a scalar, times the elements of their matrixes. The BEKK model is covariance stationary if and only if all $A_{ik} \otimes A_{ik} + B_{ik} \otimes B_{ik}$ are less than one in modulus, where ⊗ is the Kronecker product of two matrices. Both VEC and BEKK models are quite general, yet the BEKK model is considered the most applicable general multivariate model as all the VEC models that are not representable by the BEKK model are difficult or practically impossible to be formulated.

2.1.10 <u>Constant Conditional Correlation (CCC) model</u>

A rather convenient and computationally functional model that reduces the number of the required parameters that need to be estimated is the Constant Condition Correlation (CCC) model. The model was developed by Bollerslev (1990) and in contrast with the above two multivariate models, it offers the advantage to the researcher to efficiently examine the correlation between several financial assets. The correlation between the examined variables in the CCC model is assumed to be constant, which significantly decreases the number unknown of parameters providing a more simplified estimation.

The conditional covariance matrix in the CCC model is defined as follows:

$$H_t = D_t * P D_t, \qquad (12)$$

Where $D_t = \mathrm{diag}(h_1^{1/2},\ldots..,h_m^{1/2})$ with m denoting the total of model variables, $P = [\rho_{i,j}]$ represents the matrix of constant conditional correlations. The CCC model in its basic form is expressed by incorporating a GARCH(1,1) specification for every conditional variance in D_t. In that case the model encloses N* (N+5)/2 parameters. The conditional covariance in the CCC is positive definite, if and only if the all the conditional variances are positive and P is positive definite.

2.1.11 Dynamic Conditional Correlation (DCC) model

The basic assumption in the CCC model for constant correlations between the examined variables may provide the advantage of lower number of parameters, however it also constitutes one of the main weaknesses of the model, as this assumption was found to be rather unrealistic and quite restrictive by a number of empirical studies and academic researches such as Chang et al. (2009) and Chang et al. (2011). Engle (2002) trying to overcome this drawback developed the Dynamic Conditional Correlation model (DCC), which allows for the conditional correlation matrix to vary over time.

$$R_t = (1 - \lambda_1 - \lambda_2)\bar{\rho} + \lambda_1 \varepsilon_{t-i}\varepsilon'_{t-i} + \lambda_2 R_{t-1}, \tag{13}$$

Where λ_1 is a positive scalar parameter and λ_2 is a non-negative scalar parameter for which it applies that $\lambda_1 + \lambda_2 < 1$. Additionally, when $\lambda_1 = \lambda_2 = 0$ DCC model becomes equal to a CCC model. $\bar{\rho}$ is the unconditional correlation matrix of the standard errors (ε_t), with R_0 being positive definite.

Both the CCC and the DCC models require a two stage methodology in order to be estimated, which includes using a univariate model to calculate the necessary parameters for each series at a first stage and then at a second stage to estimate the correlations.

2.2 GARCH models and innovations distribution

A crucial point regarding the GARCH model's conditional variance is the assumption relative to the z_t ($\varepsilon_t = \sigma_t * z_t$) innovations distribution in equation (2). There are several examined types of distribution for z_t, from which the most commonly used are the Gaussian (Normal) distribution, the Student-t distribution, the Skewed Student-t distribution (SSD) introduced by Hansen (1994), the Generalized-Error-Distribution of Nelson (1990), the Skewed-Generalized-Error-Distribution (SGED) and the Skewed-t Generalized Distribution introduced by Theodossiou (1998, 2001) and the heavy-tailed (HT) distribution, which was first incorporated in a GARCH model by Politis (2004).

All GARCH class models can be estimated for different assumed distributions for z_t, taking into consideration of the innovations potential asymmetric characteristics (e.g. positively, negatively skewed) and tail distribution alternations (e.g. fat tail). As a result, by replacing the basic model's normal distribution with a number of different innovations distributions that can capture the above characteristics, which are typically present in most time series energy data sets, a new GARCH based model is created. The produced models which differentiate from the initial GARCH model and the assumed normal innovations distribution are more likely to be capable of better fitting and processing the examined data set, providing more accurate and robust volatility forecasts. Finally, Gao et al. (2012) by running several tests using 10 years data of two financial indices, they discovered that in all cases the GARCH model assuming GED innovations distribution, outperformed the GARCH model with t-distribution, while the latter was always found to be superior to the basic GARCH model using a normal distribution.

Latest academic studies in risk management confirm this assertion with Christoffersen et al. (2006) claiming that GARCH class models with non-normal return distributions outperform more simple model versions assuming a normal innovations distributions. The authors used long time series data of the S&P500 return to run several GARCH type models, concluding that all models showed a relatively small but evident and persistent improvement when implementing a non-normal innovations distribution, relative to the initial models which used the standard normal distribution.

Furthermore, there are a significant number of studies which find strong evidence to support and underline the superiority of the GARCH models that differentiate from the normal distribution assumption regarding innovations, in terms of financial assets' option pricing. Menn and Rachev (2005), Christoffersen et al. (2006), Badescu and Kulperger (2008) and Barone-Adesi, Engle and Mancini (2008) are some indicative papers comparing the ability of normal and non-normal innovations GARCH models in option modeling and valuation , with all of them clearly arguing in favor of non-normal innovations models.

Additionally, McNeil and Frey (2000) further suggest that normal innovations models tend to persistently undervalue the possible downside risk regarding returns of financial assets or indexes and that GARCH models using a heavy-tailed innovation distribution are more capable of calculating and predicting this risk.

2.2.1 Choosing the most appropriate innovations distribution

The most common scientific methodology in deciding the most appropriate innovations distribution for a particular GARCH class model while examining a specific time series data of financial returns involves three basic steps. At the beginning, the Quasi Maximum Likelihood Estimators (QMLE) [2] of the GARCH model coefficients is calculated subtracting the innovations. Next, using a relevant diagnostic tool, the innovations are being examined whether they fit a standard normal distribution or not. In the final stage, assuming that the innovations do not fit a standard normal distribution, all the information derived at the second step regarding the most appropriate type of innovations distribution is used and the best possible distribution is selected to fit and approximate the tail shape of the particular case innovations distribution.

However, this method shows a rather confusing paradox, as when applying the specific methodology to normal innovations GARCH processes, which approaches that of an IGARCH model, contrary to the original assumption of standard normal distribution it is quite often being realized that the innovations follow a heavy tailed distribution. This effect is particularly evident in financial returns and as a result it makes the above method less suitable for examining time series of energy data. Klar et al. (2012) extending the traditional Goodness-of-fit tests, checking normality in GARCH models' innovations distribution, as well as creating a relative symmetry test they support that they built a reliable methodology for testing the Gaussian distribution assumption in GARCH models, which can be safely extended to test the appropriateness of other types of assumed distributions, such as the student-t distribution.

Finally, Sun and Zhou (2014) in an attempt to overcome the above mentioned troubleshooting developed a test for the innovations distribution by comparing the examined tail indices with the Hill tail index estimator. The test, when examined

[2] For more information regarding the Quasi Maximum Likelihood Estimators (QMLE) see Elie and Jeantheau (1995).

using real, long, time-series return data from several financial indices, was found competent to provide robust results that can lead to a more detailed and safe estimate of the GARCH model's tail index.

2.3 GARCH class models performance criteria and model testing

2.3.1 Preliminary statistic tests for energy data sets indicating GARCH type modeling appropriateness

The vast majority of the analyses regarding energy commodities follow a specific procedure. At an initial stage, the long time series data of the examined energy commodities are being tested for several characteristic features that would encourage the use of GARCH class models for further analysis of the sample, including volatility modeling and forecast. The most common tests at this early level include skewness, kurtosis, normality, stationarity and serial correlation tests. It is most likely and usual for energy datasets to exhibit signs of non-zero skewness and fat tails combined with rejection of the hypothesis that the data follow a Gaussian type (Normal) distribution. In this case, a particularly popular test for checking normality is the Jarque – Bera (1980) test.

Furthermore, an important test that could decisively encourage the use of GARCH class models is the serial correlation test, which is normally being conducted with the use of the Ljung-Box (1978) Q test. The test is normally used to check for autocorrelation up to the 10^{th} and 20^{th} order for both actual and square energy price returns, supporting the use of GARCH class models whenever the null hypothesis of no serial correlation is being rejected, as in this case the examined observations are serially correlated and exhibit time-varying volatility.

Additionally, the data are also being tested for stationarity with the help of a series of relative tests such as the Augmented Dickey-Fuller (ADF) (1979), the Phillips-Perron (PP) (1988) and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) (1992). The first two models assume a null hypothesis of a unit root in the particular energy commodity price or return dataset, contrary to the KPSS test which uses a null hypothesis of data stationarity. Checking for stationarity is a crucial point for such research, as a stationarity data set provides the ability to the researcher to proceed in modeling the examined dataset without any restrictions or any further need for processing and transformation.

2.3.2 Goodness of fit tests and in-sample performance for GARCH class models

After the preliminary statistic results confirm that GARCH type models are appropriate to model and forecast volatility of the examined energy commodity price or return dataset, the investigation for the various GARCH models' ability to better fit the data follows. Due to the large number of differentiation and developments on the basic GARCH models, it is rather often for researchers to test the goodness of fit of a number of different GARCH type models in order to determine which model is the most appropriate to base on their research. The in-sample model performance is then examined, estimating model parameters of the selected research models that are able to capture phenomena like the asymmetric and long memory volatility effect and run several tests regarding distribution fit and any remaining unabsorbed serial correlation and ARCH effects. The models' pre-specified innovations distribution is being tested on whether it fits the examined data set, using the Maximum Likelihood Estimation (MLE) method. The logarithmic likelihood function value (Log(L)) is estimated and the model with the highest estimate is preferred over the others.

Nevertheless, because it is most likely that more complex models including a larger number of parameters yield higher log-likelihood values, it is of vital importance to further test the models on equal terms. This task is fulfilled with the use of one or both the Akaike (1974) information criterion (AIC)[3] and the Bayesian information criterion (BIC)[4], which allows the comparison of all investigated models' fit for specific dataset, penalizing the more complex models for its increase number of parameters. Under these two criteria, the model with the best in-sample fit is the one with the lowest criterion value.

Another crucial part of the in-sample model testing performance consists of the models control for any unabsorbed serial correlation and ARCH effects. At this stage models are being tested using again the Ljung-Box Q statistics, as in the initial preliminary statistics tests, to determine if there is any remaining serial correlation. In this case, the most appropriate and effective GARCH type model is the one which its P-value in the Ljung-Box test is well above 5%, indicating that the model is capable to absorb the autocorrelation which was initially detected in the data sample. Finally, the models are also tested for their ability to successfully incorporate ARCH effects up to a pre-determined order, using the Engle's (1982) Lagrange Multiplier ARCH test. Similarly, the most appropriate model is considered the one that its P-value is larger than 5% and higher than that of other models, accepting the null hypothesis of no persistent signs of heteroskedasticity.

---

[3] The AIC criterion is described as follows: $AIC = -2\log L(\theta) + 2k$, $k=1,2,\ldots,K$ where the $\log L(\theta)$ is the maximized log likelihood function of the examined model and k is the sum of the model parameters.

[4] The BIC criterion is described as follows: $BIC = -2\log L(\theta) + k\log(N)$, with N denoting the number of observations. The BIC criterion contrary to the AIC criterion, contains a more constricting penalty favoring less complex models with a smaller number of parameters.

2.3.3 Out of sample GARCH models' performance criteria

Although during the first two stages of tests regarding the suitability and appropriateness of the examined GARCH type models can provide the researcher with a relatively safe intuition, it is the out-of-sample performance and forecasting ability that is most critical for selecting a particular model over others. The reason is that market participants, whether these are buyers or sellers of specific energy commodities or they are involved in the derivative market, are only interested in researches from which they could extract valuable information that could be practically used and incorporated and hence help them improve their future buying, selling or investment strategies.

As a result, researchers put extra weight on this crucial part of their study trying to reveal the models which are more suitable and effective in modelling and making safe forecasts relative to a particular energy commodity's price and returns, or a portfolio of investments related to energy products. Several procedures and tests have been developed to deal with this issue, however the most popular among researchers in the field of energy economics are the direct evaluation of the models' results based on one or a number of approved loss functions for the particular type of study, as well as comparing the examined models relative to a predefined benchmark model for a predetermined loss function like in the Superior Predictive Accuracy (SPA) test of Hansen (2005) and Koopman et al. (2005) and Diebold and Mariano (1995) DM test. The benchmark used in these tests is usually one of the investigated models, with all of the models eventually being tested as benchmarks one by one.

2.3.3.1 Using loss functions to evaluate forecasting performance

A rather common approach that it has been used to perform a basic evaluation of the examined GARCH class models' performance is via a specific loss function.

The model that produces the minimum loss function of all models is considered in this case to be the superior one. Nevertheless, the fact that a model is found to be superior to others based on only one loss function cannot be considered as a safe measure, as a result researchers tend to use a number of different loss function to verify it. According to Bollerslev et al. (1994), Diebold and Lopez (1996) and Lopez (2001), another reason to confirm the model's superiority using a variety of different loss functions, is the fact that it is fairly difficult to decide with certainty which of the available loss functions is the best criterion for evaluating the forecasting accuracy of volatility models.

The most widely used loss functions as forecasting criteria in the field of energy economics are the following:

1. Mean Square Error (MSE) = $\frac{1}{N}\sum_{t=1}^{N}(\sigma_t^2 - \hat{\sigma}_t^2)^2$, (14)

2. Root Mean Square Error (RMSE) = $\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\sigma_t^2 - \hat{\sigma}_t^2)^2}$, (15)

3. Mean Absolute Error (MAE) = $\frac{1}{N}\sum_{t=1}^{N}\left|\sigma_t^2 - \hat{\sigma}_t^2\right|$, (16)

4. Mean Square Error adjusted for heteroskedasticity (HMSE)

   HMSE = $\frac{1}{N}\sum_{t=1}^{N}(1 - \sigma_t^2/\hat{\sigma}_t^2)^2$, (17)

5. Mean Absolute Error adjusted for heteroskedasticity (HMAE)

   HMAE = $\frac{1}{N}\sum_{t=1}^{N}\left|1 - \sigma_t^2/\hat{\sigma}_t^2\right|$, (18)

6. Quadratic Loss Function (QLIKE) = $\frac{1}{N}\sum_{t=1}^{N}[\ln(\hat{\sigma}_t^2) + \sigma_t^2/\hat{\sigma}_t^2]$, (19)

7. Logarithmic Loss Function (R^2LOG) = $\frac{1}{N}\sum_{t=1}^{N}[ln(\sigma_t^2/\hat{\sigma}_t^2)]^2$, (20)

With N symbolizing the number of forecasts, the σ_t^2 the actual volatility and $\hat{\sigma}_t^2$ the volatility forecast at time *t*. The HMSE and HMAE loss functions are basically the

MSE and MAE adjusted for heteroskedasticity, while the QLIKE loss function represents the loss indicated by a Gaussian likelihood and the R^2LOG providing analogous results to the R^2 of the Mincer-Zarnowitz (1969) regression. These specific loss functions may be incorporated in different studies serving as criteria for several examined cases. Perhaps the most representative example is the MSE loss function, which is considered as the most appropriate for risk management researches, such as Value-at-risk applications which focus more on accurately predicting higher levels of volatility than lower ones.

2.3.3.2 Superior Predictive Accuracy (SPA) test

An obvious weakness of the above approach is the fact that the models are not being compared together at the same time with each other under a specific loss function. Consequently, in this way merely the nominal values of the models' loss functions are being matched, neglecting to verify if the models' differentials among forecasting errors are significant or not. This problem is often being addressed by researchers by applying the SPA test on their models' loss functions results. The test is an extension of White's (2000) relative framework and provides the ability to the examiner to compare the forecasting performance of a specific benchmark model with all the other contestant models, using a predetermined loss function.

In this case also, the model with superior forecasting performance is the one producing the minimum expected loss. In the SPA test the loss differential between the model that is currently used as a benchmark and its K competitors is given by:

$$X_{t,l}^{(0,i)} = L_{t,l}^{(0)} - L_{t,l}^{(i)}, \qquad i = 1,2,\ldots,K \tag{21}$$

Where the $L_{t,l}^{(0)}$ is the price of loss function *l* at time *t* for a specific benchmark model M_0 and $L_{t,l}^{(i)}$ is the price for the same loss function and time for a competitive model

M_i. The null hypothesis of the SPA model is that none of the other examined model is able to outperform the model that is used as benchmark.

Specifically the null hypothesis in the SPA test can be described as follows:

H_0: $\text{Max}_{i = 1,2,\ldots,K}$ $E(X_{t,l}^{(0,i)}) \leq 0$,

Which is being tested using the following statistic:

$$T_l^{SPA} = \text{Max}_{i = 1,2,\ldots,K} \left(\sqrt{n}\bar{X}_{i,l} * \left(\lim_{n \to \infty} var\left(\sqrt{n}\bar{X}_{i,l}\right)\right)^{-1/2}\right), \quad (22)$$

With n symbolizing the number of volatility forecasting data and $\bar{X}_{i,l} = \frac{1}{n} \sum_{t=1}^{n} X_{t,l}^{(0,i)}$. The estimation of $\lim_{n \to \infty} var\left(\sqrt{n}\bar{X}_{i,l}\right)$ and the relative P-Value respectively is being made using the stationary bootstrap procedure proposed by Politis and Romano (1994). In the SPA test, the P-Value specifies the potential superiority of the benchmark model relatively to the other competitive models, with a large P-Value confirming the initial hypothesis that none of the examined models can perform better than the benchmark model. The SPA test is proved to be robust by a large number of researches, providing an accurate and simultaneous comparison regarding the potential forecasting ability of all volatility models included in the examined set. Therefore, it is the most widely applied test regarding the out-of-sample model comparison in energy commodity volatility studies.

2.3.3.3 The Diebold and Mariano (DM) test

An alternative predicting ability test is the DM test, which is also used to investigate whether the difference in loss function values between alternative volatility models is statistically significant or not. The test provides the researcher with a comparison between different volatility models regarding their forecasting accuracy based on a specific loss function. Nevertheless, the DM test examines only the predictive ability between just two competing models assuming initially equal predictive ability, contrary to the SPA test which tests the forecasting superiority of the entire investigated set of volatility models at the same time. As a result, the null hypothesis of equal predicting ability in the DM test can be described by $E(d_t) = 0$, with E symbolizing the mathematical expectation operator and $d_t = g(e_{1,t}) - g(e_{2,t})$, where $g(e_{1,t})$, $g(e_{2,t})$ is the pre-specified loss function written as a function of the forecasting errors $e_{1,t}$ and $e_{2,t}$ of the two compared models.

The null hypothesis in the DM test is being checked with the following statistic:

$$DM = \left(\hat{V}\,(\bar{d})\right)^{-1/2}\bar{d}, \tag{23}$$

Where $\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t$ according to Diebold and Mariano (1995) denotes the mean loss differential, while the asymptotic variance is approximated by the following equation:

$$\hat{V}\,(\bar{d}) \approx \frac{1}{n}\left(\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k\right), \tag{24}$$

Where n, h are the generated steps ahead forecasts from the two examined models and γ_k is the k^{th} autocovariance of d_t.

In DM test the null hypothesis is rejected whenever the test statistic is negative and significant, indicating that the benchmark model outperforms the compared model. In general the DM test is less preferred to the SPA test, as it is more reasonable and valuable to test all the alternative volatility models simultaneously and not in pairs, while Harvey et al. (1997) also claim that the DM test can be relatively oversized for small samples and especially for wider forecasting horizons. Nevertheless, the DM test is really useful and often used to provide a relative comparison of the out-of-sample model forecasting performance and in most cases it is used together with the SPA test.

2.4 *The Markov Regime Switching GARCH model (MS-GARCH)*

All the GARCH type models that are being analyzed and presented so far in the paper refer to single-regime models. However, under the effect of extreme economic and geopolitical events there is high chance that the market of more volatile assets, such as that of energy commodities, will behave in a completely changed way than it would under normal conditions. As, a result researchers trying to capture this change in the market behavior of energy products, adopted the Markov-Switching approach as first presented by Hamilton (1994) and combined it with the traditional GARCH model of Bollerslev (1986), developing a new hybrid model that has all the advantage of GARCH type models, while further allowing for a probable regime switch in the estimated parameters of the variance process. Hence, the basic difference between the single-regime GARCH models and the multiple-regime Markov-Switching GARCH models (MS-GARCH), is that the latter incorporate a regime-switching variable which is allowed to switch over different regimes following the Markov process.

In that case, the probability for a jump from a regime *i* at a specific time period *t* −1 to a regime *j* at a time period *t* is estimated as follow:

$$P(s_t = j \mid s_{t-1} = i) = p_{i,j} \qquad (25)$$

Nevertheless, because the MS-GARCH models can prove to be rather computationally intensive, researchers in the relative academic literature[5] propose a representative two-regime MS-GARCH model, with one regime corresponding to periods with relative stability and market normality and another regime corresponding two periods with particularly excessive economic uncertainty and market price turbulences.

This model is specified by the following equation:

$$h_t = a_0 + a_{1(s_t)}\, \varepsilon_{t-1}^2 + \beta_{1(s_t)}\, h_{t-1} \qquad (26)$$

[5] See Gray (1996), Klaassen (2002)

In which, S=0 indicates a regime with relatively normal volatility, while S=1 indicates a jump to a regime where prices follow high volatility dynamics. However, much more complex models can be built allowing for three or more jumps, representing for example the pass from an upward to a downward trend period with positive and negative mean changes respectively, or to a trendless period with a mean change close to zero. Finally, it is important to mention that the MS-GARCH models are being checked relatively to their appropriateness and forecasting performance, based on the same utility functions and tests that are being implemented in the case of single-regime GARCH models.

3. Literature review

3.1 Energy commodity risk management via single regime GARCH type volatility models

The presence of the asymmetric as well as the long memory volatility effect together with skewness, kurtosis and fat tails are some of the elements that are empirically confirmed by many studies to be present and strongly characterize the energy commodity market. As result, it became a necessity for every party involved in this market to conduct risk management analyses to deal with the high risk coming from this excessively volatile market environment. Additionally, due to the importance of the market, the development of suitable modeling and analytical tools of energy commodity volatility, it has been a rather popular issue for research among academics.

GARCH based volatility models have been widely acceptable as one of the most appropriate and efficient ways to deal with energy commodity risk as well as to accurately model and forecast energy commodity volatility. This is mainly happening because these types of models offer researchers the ability to capture and account for the special features of the return volatility of these commodities. Hence a large number of studies has been developed trying to analyze the return characteristics and make volatility forecasts of various energy products with the use of several GARCH family models, which their predicting ability is then assessed using various loss functions and accuracy tests.

Sadorsky (2006) is one of the first academic papers which focus on modeling and forecasting energy commodity volatility. Specifically, this study emphasizes on oil as well as other commodities and using a random walk model as benchmark it compares the forecasting ability of several univariate and multivariate models such as the autoregressive model (AR), the basic GARCH(1,1) model, the TGARCH[6] model and the Vector Autoregression[7] (VAR) and BEKK models. The models were developed and assessed based on a 15-year (1988-2003) sample of daily futures returns for WTI oil, heating oil, gasoline and natural gas, with the relative loss functions used in the study and the DM accuracy test revealing that no particular model outperformed the others for all four commodities.

Nevertheless, the TGARCH model proved to be the more appropriate for the heating oil and natural gas and the GARCH(1,1) model to fit better for WTI oil and gasoline volatility, while none of the most sophisticated models (VAR,BEKK) managed to outclass the basic GARCH(1,1) model. After Sadorsky (2006) a plethora of relative studies started to be published highlighting the advantages of individual GARCH models and trying to determine the most suitable GARCH type models for each energy commodity and examined time period. Narayan et al. (2007) examining daily oil returns for a 15-year period (1991-2006) determined that the EGARCH model is capable of capturing both asymmetric and long memory volatility effects, while they also found strong signs of time varying volatility.

Choi and Hammoudeh (2009) relied on the FIGARCH model to successfully identify and capture the intensity of the long memory volatility effect in a 20-year (1986-2005) sample of daily futures returns for WTI oil and two refined products. Agnolucci

---

[6] See Engle and Ng (1993)
[7] See Hamilton (1994), Harris and Sollis (2003)

(2008) also using WTI oil daily returns for a 14-year (1991-2005) period, concluded that GARCH type models tend to outclass implied volatility[8] models, while GARCH models with GED distribution were even more accurate than those with normal or Student-t distributions. Kang et al. (2008) collecting evidence from the oil market they evaluate a group of GARCH models consisting of the normal GARCH, IGARCH, CGARCH, and FIGARCH models.

Similarly to Sadorsky (2006) the models are judged with the help of both specific loss functions as well as the DM statistical test, with the results suggesting that risk analysts should always allow for the long memory volatility effect in their analyses concerning oil returns, as the CGARCH and FIGARCH models which can capture this phenomenon were the ones providing the best forecasts. Cheong (2009), investigating a 15-year (1993-2008) sample of daily returns for WTI oil and Brent oil, indicated that the long memory volatility effect is present at a relatively low degree at the Brent oil returns, as likewise Sadorsky (2006) it is found that the more simple short memory GARCH models, with either normal or Student-t distribution, perform better than those with higher complexity.

In contrast, the long memory effect is found to be significantly present in the WTI returns, with the relative tests highlighting the FIAPARCH model with Student-t distribution as the most appropriate model for volatility forecasting, as it dominated for both short and long forecasting horizons. Mohammadi and Su (2010) is the first and perhaps one of the few academic papers that examine the behavior of weekly oil returns. Specifically, the authors use an 11-year (1997-2009) sample of weekly oil returns containing data from several major oil markets, upon which they compare the out of sample forecasting performance of the basic GARCH, the EGARCH, the APARCH and FIGARCH models all assuming that the innovations distribution is approximated by the

[8] See Ledoit and Santa-Clara (1998), Schönbucher (1999) and Brace et al. (1997, 2001)

skewed Student-t distribution. The results based on the examined loss function and the DM accuracy test revealed that the APARCH model tends to outperform the other three models, a finding which is in contrast with that of Kang et al. (2008) who argued that volatility shocks regarding oil returns dissipate at a slow hyperbolic rate as indicated by the FIGARCH model.

Wei et al. (2010), further extend the research of Kang et al. (2008) by comparing the forecasting performance of a larger group of GARCH models using 18-years (1992-2009) of daily spot returns for WTI oil and Brent oil. Moreover, the basic GARCH, IGARCH, Risk Metrics, GJR-GARCH, EGARCH, APARCH, FIGARCH, FIAPARCH and HYGARCH models included in the study were tested based on six different loss functions instead of three as in the paper of Kang et al. (2008) as well as on both the DM and SPA performance tests. The outcomes of Wei et al. (2010) reveal that in contrast with Kang et al. (2008), no specific individual model was found to be superior to the other competitive models, with the authors pointing that a model which performs better under specific market conditions and loss functions there is good chance that it will not be as reliable for other markets and under different loss functions.

At the same time and contrary to other studies, the more complex non-linear GARCH models, which are able to account for both the long memory and the leverage effects, appear to be slightly better choices particularly for making volatility predictions over longer future time horizons. Hou and Suardi (2011), based on the methodology of Bühlmann and McNeil (2002) they develop a non-parametric GARCH model, with which they compare the same set of parametric GARCH models of Wei et al. (2010), however in this case the models rely on the Student-t distribution.

Next, in compliance with the findings of Patton (2011) that the use of many loss functions may lead to serious distortions when used with standard volatility proxies such as squared returns, the authors evaluate the models based on the two loss functions for which their optimal forecast varies the least and also the SPA test. The results using an 18-year (1992-2010) sample of daily spot prices for WTI oil and Brent oil show a clear superiority of the non-parametric GARCH model relative to all nine rival parametric models.

Wang et al. (2011), examine the ability of several GARCH class models, such as the basic GARCH, GJR-GARCH, EGARCH, APARCH and FIGARCH models, to efficiently capture the long memory volatility effect which is profoundly present in the energy commodity returns. In this study the non-parametric methods of de-trended fluctuation analysis (DFA)[9] and rescaled range analysis (R/S)[10] are used in order to compare the long memory properties of the conditional volatility, as captured by the specific GARCH models, with that of the actual volatility. Using a 20-year (1990-2010) sample of daily spot and futures prices for WTI oil, the two methods showed that the models performed quite satisfying in terms of capturing the long memory effect for time scales exceeding 12 months, while they did relatively poor for shorter time scales.

Arouri et al. (2012), using a large sample consisting of 25-years (1986-2011) of data concerning daily spot returns as well as one and two month futures returns for WTI oil, gasoline and heating oil, they test the potential ability of a series of GARCH type models to capture the long memory volatility effect. In harmony with previous similar studies the authors determined that the FIGARCH model is better able to account for this phenomenon and generally to describe the time varying energy commodity return volatility. Furthermore, it is argued that a FIGARCH model suitably modified to account for both structural brakes and long memory effect would be better than any other related model under any criteria.

A rather interesting study is also that of Wang and Wu (2012), in which they compare the accuracy of the forecasted volatilities of a set of univariate models, including

[9] See Peng et al. (1994)
[10] See Hurst (1951)

the basic GARCH, the GJR-GARCH, the EGARCH and the FIGARCH models, with that of a set of multivariate models, including the Full BEKK, the Diagonal BEKK and Scalar BEKK models, under six different loss functions and the SPA test. Under these criteria and based on a 19-year (1992-2011) sample of weekly spot returns for WTI oil, gasoline, heating oil and jet oil, it is revealed that more complex multivariate models outclass the relatively simple univariate models as far it concerns the accurate volatility forecasting of the examined energy commodities. This can be justified up to an extent by the fact that the estimation of the constant conditional correlations in the CCC model showed a statistically significant and strong implication of correlation between the oil market and the markets of the other three refined energy commodities. The authors argue that this robust relationship between oil and its refined products can provide vital information for all participants involved in the specific market, as the market dynamics of the refined products can influence the oil price.

Finally, Kang and Yoon (2012) combine the ARIMA methodology and its generalized version the ARFIMA methodology with several GARCH type models, such as the basic GARCH, the IGARCH and the FIGARCH models, in an attempt to create mix models that will account for both long memory effect and volatility at the same time. The models were developed and tested based on a 17-years (1995-2012) sample of daily futures returns for WTI oil, heating oil and gasoline, with the evidence from the two loss functions and the DM test included in the study showing that none of the examined models managed to outperform the other competitive models in all three markets. Nevertheless, it was found that the less complex ARIMA-GARCH models was superior in terms of volatility forecasting regarding crude oil futures, while the more sophisticated ARFIMA-FIGARCH model dominated in forecasting volatility for heating oil and gasoline.

3.2 *Energy commodity risk management via Markov-Switching GARCH models*

A rather common assumption for the vast majority of researches, which try to model and eventually attempt to make forecasts relative to price volatility of energy products using GARCH family models, is that the unconditional variance of these commodities is constant. However, the fact that the energy market is by far more volatile than any other and is reacting in a more extreme way to the various unexpected economic and geopolitical events, such as the global recession of 2007 and the political instability and the military conflicts in the middle east, constitutes the hypothesis for constant unconditional variance reasonably unrealistic.

During periods of high economic instability like the aforementioned, the energy market is exposed to persistent shocks of largely steep price changes, which can practically cause structural brakes to the unconditional variance and according to Bauwens et al. (2014) can seriously affect the precision and predicting ability of single regime GARCH models. This may happen because the extremely high persistence of the shocks that is being captured by these models possibly indicate also a regime switch in the estimated parameters of the variance process, hence revealing a structural brake in the model's volatility dynamics. Lamoureux and Lastrapes (1990), Hamilton and Susmel (1994), Gray (1996), Klaassen (2002), Mikosch and Stărică (2004), and Hillebrand (2004) strongly support that in case there are no switching parameters included in the model, a substantial upward bias is most likely to appear in the approximations regarding the persistence of the conditional volatility that is extracted through a GARCH type model.

In order to overcome the above shortcoming, researchers incorporated the Markov-Switching regimes methodology in their GARCH based volatility analysis. Cai (1994) and Hamilton and Susmel (1994) were the first to combine the two techniques building models that connected regime switching with the ARCH innovations. Gray

(1996) and Klaassen (2002) further developed the original model by modifying it properly to allow for GARCH dynamics. Finally, Haas et al. (2004) diversification over Grey's (1996) model, is by far the most widely academia as it offers better analytical tractability, providing stationarity conditions and a more clear–cut interpretation of the variance dynamics.

In the relative literature concerning the issue of volatility modeling and forecasting for energy commodities, more and more academic researchers use Markov-Switching GARCH models to address the problems of the asymmetric and volatility clustering effects that are intensively present in their price returns. Specifically, the fact that an overwhelming number relative papers, such as that of Arouri et al. (2012), report that most data samples containing energy product returns reveal a strong indication of the presence of structural brakes in their volatility, making a necessity to compare, in terms of suitability and predicting ability, all proposed GARCH family models with a Markov-Switching regime GARCH model.

One of the first papers in that specific field is that of Fong and See (2002) in which a data sample is used containing observations for an 8 year period about the daily returns of the WTI crude oil futures. They conclude that their regime switching GARCH model constantly outperformed the basic GARCH(1,1) model in capturing the unobserved volatility effects, due to the existence of regime switches in the volatility. Next, Vo (2009) shows that based on a sample for weekly crude oil spot prices from 1986 to 2008, there is evident regime switching volatility giving the wrong impression of simple highly persistent volatility and thus increased predictability when examined with a single regime GARCH model. Furthermore, it is supported that MS-GARCH models are more appropriate to model and forecast crude oil volatility, as well as they are capable to incorporate major events that are able to seriously affect the oil price.

Nomikos and Pouliasis (2010), comparing the forecasting ability of MS-GARCH models to the basic GARCH on a large data set containing information about daily futures prices from 1991 to 2008 for WTI crude and heating oil and ICE Brent crude and gas oil, it is supported that the Markov-Switching models clearly dominate accounting for the special volatility characteristics of the specific commodities and making future price forecasts. Additionally, Luo et al. (2010) underline that the two regime MS-GARCH models that were used, accounting for normal and turmoil periods respectively, exhibit the essential advantage of being able to the stochastic volatility of crude oil returns into several different characteristic groups.

Another interesting study is that of Cifter (2013), which explores the Nordic electricity market using daily price returns for a highly volatile period between 2008 and 2011 and evaluates the performance of MS-GARCH models relative to the GARCH(1,1) and the GJR-GARCH model, testing for three different return distribution, including the normal, the Student-t and the Skewed Student-t. Again, the results favor the MS-GARCH models, with the Skewed Student-t distribution enhancing the performance of both types of models. Similarly, Gunay (2015), Runfang et al. (2017) and Zhang et al. (2018) based on crude oil data sets of daily returns compare the MS-GARCH model with the basic GARCH and two non-linear GARCH models, the GJR-GARCH and the EGARCH, with the MS-GARCH model dominating repeatedly.

Finally, Di Sanzo (2018) using an extensive sample of both Brent and WTI crude oil daily returns from 1992 until 2015, evaluates both the in-sample as well as the out-of-sample performance of the MS-GARCH model with a series of GARCH type models, including the GARCH(1,1) model, the IGARCH, the EGARCH, the GJR-GARCH, the FIGARCH and the FIAPARCH. The models were tested based on the Mean Square Error (MSE) and the Quasi-Likelihood (QLIKE) loss functions, while the AIC and BIC criteria are penalizing all compared models. Clearly the MS-GARCH model was found to be superior above all other GARCH type models, proving that it is the most appropriate and robust model, meeting all the necessary requirements to perform a thorough examination of the price volatility of energy commodities.

Nevertheless, at this point it is considered essential to warn the reader that the vast majority in the abovementioned academic researches, the authors use data samples that include periods of intensive economic crises and generally uncertain geopolitical climate and as such the volatility models that were developed based on Markov-Switching GARCH methodology are most likely under such conditions to be found more appropriate to model energy product volatility.

4. Conclusion

The energy market is characterized by excessive uncertainty and price volatility as a result of the high impact of geopolitical and environmental factors as well as the global demand for competition rise and market deregulation. These highly unstable market conditions make a necessity for both energy economists and all the businesses actively involved in that particular market, to examine and determine the most appropriate methodologies and tools to manage the energy commodities' extreme price risk. Several academic researches analyzing data samples containing a significant number of price observations for multiple time periods regarding energy products with a high trading volume, revealed that energy commodity price volatility exhibit strong signs of both long memory and asymmetric effects.

Hence, independent of which volatility model is applied to measure and forecast the price volatility, it is essential first to test for the most appropriate return distribution and then for this return distribution to be incorporated into the model. As it is an exceptionally rare phenomenon not to observe extreme levels of skewness and kurtosis in the return distribution of energy products, making the unrealistic assumption of a normal return distribution it would almost certainly lead to a dangerous underestimation of the total price risk and poor forecasting ability.

However, because accuracy comprises a crucial part in such analyses, testing the examined models under different loss functions may reveal which specific model measures price risk with the greatest precision. Additionally, in case the data set of the returns of the examined energy products contain periods of severe economic crisis like the debt crisis of 2007, or other significant geopolitical events, then researchers turn to the Markov-Switching GARCH methodology to test and account for potential regime shifts in the volatility process

Finally, as it is indicated by the current analysis as well as it is highlighted by the vast majority of the more recent studies, there is no particular individual model or methodology that can outperform all the others in modeling and forecasting the price changes of all the main energy commodities. Most researchers in their concluding remarks express the belief that the suitability and therefore the performance of a single model or methodology strictly relies on the exact sample that is examined along with the any special characteristic features that affect the trade of the specific energy product.

References

Agnolucci P. (2008). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, 31(2), pp. 316-321.

Andersen T.G., Bollerslev T., Diebold F. X. and Labys P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), pp. 529–626.

Ardia, D., Bluteau, K., Boudt, K., & Catania, L. (2018). 'Forecasting risk with Markov-switching GARCH models:A large-scale performance study'. *International Journal of Forecasting*, 34(4), pp. 733-747.

Arouri, M., Lahiani, A., Levy, A. and Nguyen, D. (2012). Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. *Energy Economics,* 34, pp. 283-293.

Artzner P., Delbaen F., Eber J. and Heath D. (1999). Coherent measures of risk, *Mathematical Finance*, 9(3), pp. 203–228.

Baba Y., Engle R.F., Kraft D.F. and Kroner K.F. (1991). Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, 11(1), pp. 122-150.

Badescu A.M. and Kulperger R.J. (2008). GARCH option pricing: A semi-parametric approach. *Insurance: Mathematics and Economics*, 43(1), pp. 69–84.

Baillie R.T., Bollerslev T. and Mikkelsen H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1), pp. 3-30.

Barone-Adesi G., Engle R.F. and Mancini L. (2008). A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies*, 21(3), pp.1223-1258.

Barone-Adesi G., Giannopoulos K. and Vosper L. (2002). Backtesting Derivative Portfolios with FHS. *European Financial Management*, 8(1), pp. 31-58.

Basel Committee on Banking Supervision (1996). Supervisory framework for the use of "backtesting" in conjunction with the internal model-based approach to market risk capital requirements. *Bank for International Settlements*, Basel, Switzerland. Retrieved from: https://www.bis.org/publ/bcbsc223.pdf

Basel Committee on Banking Supervision, (2004). International convergence of capital measurement and capital standards. *Bank for International Settlements,* Basel, Switzerland. Retrieved from: https://www.bis.org/publ/bcbs128.pdf

Basel III. (2010). A global regulatory framework for more resilient banks and banking systems. *Bank for international settlements,* Basel, Switzerland. Retrieved from: https://www.bis.org/publ/bcbs189_dec2010.pdf

Bauwens L., De Backer B. and Dufays A. (2014). A Bayesian method of change-point estimation with recurrent regimes: Application to GARCH models. *Journal of Empirical Finance*, 29, pp. 207-229.

Bauwens, L., Preminger, A., & Rombouts, J. V. K. (2010). Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, 13(2), pp. 218–244.

Bertsimas D., Lauprete G.J. and Samarov A. (2004). Shortfall as a Risk Measure: Properties, Optimization and Applications. *Journal of Economic Dynamics and Control*, 28(7), pp. 1353–1381.

Bierbrauer M., Menn C, Rachev S.T. and Truck S. (2007). Spot and derivative pricing in the EEX power market. *Journal of Banking and Finance*, 31(11), pp. 3462-3485.

Black F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1-2), pp. 167-179.

Bollerslev T, RY Chou and KF Kroner (1992). ARCH modeling in finance. *Journal of Econometrics*, 52(1-2), pp. 5–59.

Bollerslev T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3), pp. 542-547.

Bollerslev T. (1990). Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach. *Review of Economics and Statistics,* 72, pp. 498-505.

Bollerslev T, Engle R.F. and Nelson D.B. (1994). ARCH Model. *Handbook of Econometrics,* Volume IV, (pp.2959–3038). Elsevier Science B.V. Retrieved from: http://public.econ.duke.edu/~boller/PublishedPapers/benhand94.pdf

Bollerslev T. and Mikkelsen H. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1), pp. 151–184.

Bollerslev T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics,* 31(3), pp. 307-327.

Bollerslev T. and Wooldridge J.M. (1992). Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances. *Econometric Reviews*, 11(2), pp. 143-172.

Bollerslev T., Engle R.F. and Wooldridge J.M. (1988). A Capital Asset Pricing Model with Time Varying Covariances. *Journal of Political Economy*, 96(1), pp. 116-131.

Brace A., Gatarek D. and Musiela M. (1997). The Market Model of Interest Rate Dynamics. *Mathematical Finance*, 7(2), pp. 127-155.

Brace A., Goldys B., Klebaner F. and Womersley R. (2001). Market Model of Stochastic Implied Volatility with Application to the BGM Model. *Working Paper,* Department of Statistics, University of New South Wales.

Bühlmann, P & McNeil, AJ (2002). An algorithm for nonparametric GARCH modelling *Computational Statistics & Data Analysis,* 40(4), pp. 665-683.

Cabedo J.D. and Moya I. (2003). Estimating oil price value at risk using the historical simulation approach. *Energy Economics*, 25(3), pp. 239–253.

Cai J. (1994). A Markov Model of Switching-Regime ARCH. *Journal of Business & Economic Statistics*, 12(3), pp. 309-315.

Ceczy C.C., Minton B.A. and Schrand C.M. (2000). Choices among alternative risk management strategies: Evidence from the natural gas industry. Working Paper, University of Pennsylvania.

Chan K.F., Gray P. and Van Campen B. (2008). A new approach to characterizing and forecasting electricity price volatility. *International Journal of Forecasting,* 24(4), pp. 728-743.

Chang C., McAleer M. and Tansuchat R. (2009). Modelling conditional correlations for risk diversification in crude oil markets. *The Journal of Energy Markets*, 2(4), pp. 29-51.

Chang C., McAleer M. and Tansuchat R. (2010). Analyzing and forecasting volatility spillovers, asymmetries and hedging in major oil markets, *Energy Economics*, 32(6), pp. 1445-1455.

Chang C., McAleer M. and Tansuchat R. (2011). Crude oil hedging strategies using dynamic multivariate GARCH. *Energy Economics*, 33(5), pp. 912-923.

Chen C.W., So M.K. and Lin E.M. (2009). Volatility forecasting with double Markov switching GARCH models', *Journal of Forecasting*, 28(8), pp. 681–697.

Cheong C.W. (2009). Modeling and forecasting crude oil markets using ARCH-type models. *Energy Policy*, 37(6), pp. 2346-2355.

Chkili W., Hammoudeh S. and Nguyen K.D. (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics*, 41, pp. 1-18.

Choi K. and Hammoudeh S. (2009). Long memory in oil and refined products markets. *The Energy Journal,* 30(2), pp.48-55.

Christoffersen P.F. (1998). Evaluating interval forecasts, *International Economic Review*, 39(4), pp. 841-862.

Christoffersen P.F. and Diebold F.X. (2006). Financial asset returns, direction-of-change fore-casting and volatility dynamics. *Management Science*, 52(8), pp. 1273–1287.

Christoffersen, P., 1998. Evaluating Intervals Forecasts, *International Economic Review*, 39, pp. 841-862.

Cifter A. (2013). Forecasting electricity price volatility with the Markov-switching GARCH model: Evidence from the Nordic electric power market. *Electric Power Systems Research*, 102, pp. 61-67.

Clewlow, L., & Strickland, C. (2000). Energy derivatives: Pricing and risk management. London: Lacima Publications.

Davidson J. (2004). Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business & Economic Statistics*, 22(1), pp. 16–29.

Degiannakis S. (2004). Volatility forecasting: evidence from a fractional Integrated asymmetric Power ARCH skewed-t model. *Applied Financial Economics,* 14(18), pp. 1333–1342.

Denton M., Palmer A., Masiello R. and Skantze P. (2003) Managing Market Risk in Energy. IEEE *Transactions on Power Systems*, *18*(2), pp. 494-502.

Devlin J. (2004). Managing oil price risk in developing countries. *The World Bank Research Observer*, 19(1), pp. 119-139.

Di Sanzo S. (2018). A Markov switching long memory model of crude oil price return volatility. *Energy Economics*, 74, pp. 351-359.

Dickey D.A. and Fuller W.A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366), pp. 427-431.

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

Dueker M.J. (1997). Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility. *Journal of Business & Economic Statistics*, 15(1), pp. 26-34.

Elder J. and Serletis A. (2008). Long memory in energy futures prices. *Review of Financial Economics*, 17(2), pp. 146-155.

Energy Information Administration (EIA) (2002). Derivatives and Risk Management in the Petroleum, Natural Gas and Electricity Industries. Washington, DC: Department of Energy.

Engle R.F. and Bollerslev T. (1986). Modeling the persistence of conditional variances. *Econometric Reviews*, 5(1), pp. 1–50.

Engle R.F. and Mustafa C. (1992). Implied ARCH models from option price. *Journal of Econometric*s, 52(1-2), pp. 289–311.

Engle R.F. and Ng V.K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5), pp. 1749-1778.

Engle R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, 50(4), pp. 987-1007.

Engle R. F. (2001). Garch 101: The use of ARCH and GARCH models in Applied Econometrics. *Journal of Economic Perspectives*, 15(4), pp. 157-168.

Engle R.F. (2002). Dynamic Conditional Correlation. *Journal of Business and Economic Statistics*, 20(3), pp. 339-350.

Engle R.F. and Kroner, K.F. (1995). 'Multivariate simultaneous generalized ARCH'. *Econometric Theory*, 11(1), pp. 122-150.

Elie, L. and Jeantheau, T. (1995). Consistency in heteroskedastic models. *Comptes Rendus de l 'Acad´emie des Sciences*, 320, pp. 1255–1258.

Eydeland A. and Wolyniec K. (2003). *Energy and Power Risk management: new developments in modeling, pricing and hedging*, Chicago: Wiley.

Fama E.F. (1965). The behavior of stock market prices. *Journal of Business*, 38(1), pp 34-105.

Fleming J., Kirby C. and Ostdiek B. (2001). The economic value of volatility timing. *Journal of Finance*, 56(1), pp. 329–352.

Fong W.M. and See K.H. (2002). A Markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics*, 24(1), pp. 71–95.

Gao Y., Zhang C. and Zhang L. (2012). Comparison of GARCH Models based on Different Distributions'. *Journal of Computers*, 7(8), pp. 1967-1973.

Gray S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1), pp. 27-62.

Gunay S. (2015). Markov Regime Switching Generalized Autoregressive Conditional Heteroskedastic Model and Volatility Modeling for Oil Returns. *International Journal of Energy Economics and Policy*, 5(4), pp. 979-985.

Haas M. (2004). A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, 2(4), pp. 493-530.

Halkos G.(2006). *Econometrics: Theory and practice*. Giourdas Publications.

Halkos G.(2011). *Econometrics: Theory and practice: Instructions in using Eviews, Minitab, SPSS and excel*. Gutenberg: Athens.

Halkos G. and Papadamou S. (2006). An investigation of bond term premia in international government bond indices. *Research in International Business and Finance*, 20(1), pp. 45-61.

Halkos G. and Papadamou S. (2007). Significance of risk modelling in the term structure of interest rates. *Applied Financial Economics*, 17(3), pp. 237-247.

Halkos G. and Sepetis A. (2007). Can capital markets respond to environmental policy of firms? Evidence from Greece, *Ecological Economics,* 63(2-3), pp. 578-587,

Halkos G. and Zisiadou A. (2018a). Reporting the natural environmental hazards occurrences and fatalities over the last century, MPRA Paper 87936, University Library of Munich, Germany.

Halkos G. and Zisiadou A. (2018b). Analysing last century's occurrence and impacts of technological and complex environmental hazards, MPRA Paper 90503, University Library of Munich, Germany.

Hamilton J.D. (1989). A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle. *Econometrica*, 57(2), pp. 357-384.

Hamilton J.D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2), pp. 39-70.

Hamilton J.D. and Susmel R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2), pp. 307-333.

Hamilton J.D. (1994). *Time series Analysis*, New Jersey: Princeton University Press.

Hammoudeh S.E. and Aleisa E. (2004). Dynamic relationships among GCC stock markets and NYMEX oil futures, *Contemporary Economic Policy*, 22(2), pp. 250–269.

Harris R. and Sollis R. (2003). *Applied time series Modeling and Forecasting,* West Sussex: Wiley and Sons Limited.

Harvey D.I., Leybourne S.J. and Newbold P. (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, 13(2), pp. 281-291.

Hibon M. and Evgeniou T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, 21(1), pp. 15–24.

Hillebrand E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129(1-2), pp. 121-138.

Hou A. and Suardi S. (2012). A nonparametric GARCH model of crude oil price return volatility. *Energy Economics*, 34(2), pp. 618-626.

Hudson Doran S.J. and Ronn E.I., (2008). Computing the Market Price of Volatility Riskin the Energy Commodity Markets. *Journal of Banking and Finance*, *32*(12), pp. 2541-2552.

Huisman R. and Mahieu R.J., (2001). Regime jumps in electricity prices. *SSRN Electronic Journal*, Volume 25, Issue 5, pp. 425-434.

Huisman R., Dahlen K.E. and Westgaard S. (2015). Risk Modelling of Energy Futures: A comparison of RiskMetrics, Historical Simulation, Filtered Historical Simulation, and Quantile Regression. In *Stochastic Models, Statistics and Their Applications*, pp 283-291. Wraclow, Switzerland: Springer.

Hurst, H.E., 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116, pp. 770-799.

Jackson J. (2010). Promoting energy efficiency investments with risk management decision tools. *Energy Policy*, 38(8), pp. 3865-3873.

Jarque C.M. and Bera A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), pp. 255-259.

Kang S.H. and Yoon S. (2012). Modeling and forecasting the volatility of petroleum futures prices'. *Energy Economics*, 36, pp. 354-362.

Kang S.H., Kang S. and Yoon S. (2008). Forecasting volatility of crude oil markets'.*Energy Economics*, 31(1), pp. 119-125.

Khindarova I. and Atakhanova Z. (2002). Stable modeling in energy risk management. *Mathematical Methods of Operations Research (ZOR)*, 55(2), pp. 225-245.

Klaassen F.F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27(2), pp. 363-394.

Klar B., Lindner F. and Meintanis S. (2012). Specification tests for the error distribution in GARCH models. *Computational Statistics and Data Analysis*, 56(11), pp. 3587–3598.

Kupiec P.H. (1995). Techniques for verifying the accuracy of risk measurement Models. *The Journal of Derivatives*, 3(2), pp. 73–84.

Lamoureux C.G., and Lastrapes W.D. (1990). Persistence in Variance, Structural Change and the GARCH Model. *Journal of Business & Economic Statistics*, 8(2), pp. 225-234.

Lanza A., Mateo M. and McAleer M., (2004). Modelling Dynamic Conditional Correlations in WTI oil forwards and futures returns, *Finance Research Letters*, Elsevier, 3(2), pp. 114-132.

Ledoit O. and Santa-Clara P. (1998). Relative Pricing of Options with Stochastic Volatility. Los Angeles: Anderson Graduate School of Management, UCLA.

Ljung G.M., and Box G.E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), pp. 297-303.

Lopez J.A. (2001). Evaluating the predictive accuracy of volatility models. *Journal of Forecasting*, 20(2), pp. 87-109.

Luo C., Seco L.A., Wang H. and Dash Wu D. (2010). Risk modeling in crude oil market: a comparison of Markov switching and GARCH models. *Kybernetes*, 39(5), pp. 750-769.

Marcucci J. (2005). Forecasting Stock Market Volatility with Regime-Switching GARCH Models. *Studies in Nonlinear Dynamics & Econometrics*, 9(4), pp. 1-55.

McNeil A.J. and Frey R. (2000). Estimation of tail-related risk measures for Heteroskedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, *7*(3-4), pp. 271-300.

Menn C. and Rachev S.T. (2005). A GARCH option pricing model with α-stable innovations. *Europrean Journal of Operational Research*, 163(1), pp. 201–209.

Mikosch T. and Stărică C. (2004). Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. *Review of Economics and Statistics*, 86(1), pp. 378-390.

Mincer, J. and V. Zarnowitz (1969). The evaluation of economic forecasts. In J. Mincer (Ed.), *Economic Forecasts and Expectations*. New York: Columbia University Press.

Mohammadi H. and Su L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics*, 32(5), pp. 1001-1008.

Morana C. (2001). A semi-parametric approach to short-term oil price forecasting. *Energy Economics*, 23(3), pp. 325–338.

Morgan J.P. (1996). RiskMetrics Technical Document 4th ed', New York.

Mount T.D., Ning Y. and Cai X. (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, 28(1), pp. 62-80.

Narayan P.K. and Narayan S. (2007). Modeling oil price volatility. *Energy Policy*, 35(12), pp. 6549-6553.

Nelson D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica,* 59(2), pp. 347-370.

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

Panas E. (2001). Estimating fractal dimension using stable distributions and exploring long memory through ARFIMA models in Athens Stock Exchange. *Applied Financial Economics*, 11(4), pp. 395-402.

Papapetrou E. (2001). Oil price shocks, stock markets, economic activity and employment in Greece. *Energy Economics,* 23(5), pp. 511–532.

Patton, Andrew J., (2011). Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics*, 160(1), pp. 246-256.

Peng C., Buldyrev S.V., Simons H.M., Stanley H.E. and Goldberger A.L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2), pp. 1685–1689.

Pilipovic D. (1998). *Energy risk: Valuing and managing energy derivatives*, New York: McGraw-Hill.

Pindyck R.S. (1999). The Long-Run Evolution of Energy Prices'. *The Energy Journal*, 20 (2), pp. 1-27.

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

Pritsker M. (2005). The hidden dangers of Historical Simulation. *Journal of Banking and Finance,* 30(2), pp. 561-582.

Runfang Y., Jiangze D. and Xiaotao L. (2017). Improved Forecast Ability of Oil Market Volatility Based on combined Markov Switching and GARCH-class Model. *Procedia Computer Science*, 122, pp. 415-422.

Sadorsky P., (2006). Modeling and forecasting petroleum futures volatility', *Energy Economics*, 28(4), pp. 467-488.

Schönbucher P. (1999). A Market Model for Stochastic Implied Volatility. *SSRN Electronic Journal*. 357(1758), pp. 2071-2092.

Smithson C.J. and Simkins B.J. (2005). Does Risk Management Add Value? A Survey of the Evidence. *Journal of Applied Corporate Finance*, 17(3), pp. 8-17.

Sun P. and Zhou C. (2014). Diagnosing the Distribution of GARCH Innovations. *Journal of Empirical Finance*, 29, pp. 287–303.

Vo M.T. (2009). Regime-switching stochastic volatility: Evidence from the crude oil market. *Energy Economics*, 31(5), pp. 779-788.

Wang Y. and Wu C. (2012). Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models?, *Energy Economics,* 34(6), pp. 2167-2181.

Wang Y., Wu C., Wei Y. (2011). Can GARCH-class models capture long memory in WTI crude oil markets?. *Economic Modelling*, 28(3), pp. 921-327.

Wei Y., Wang Y. and Huang D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32(6), pp. 1477-1484.

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

Yue-Jun Zhang, Ting Yao, Ling-Yun He, (2018). Forecasting crude oil market volatility: can the Regime Switching GARCH model beat the single-regime GARCH models?' *International Review of Economics & Finance*, 59, pp. 302-317.

Xiao L. and Aydemir A. (2007) Volatility modelling and forecasting in finance. **In**: Knight J. and Satchell S. (Eds). *Forecasting volatility in the financial markets*. Elsevier Finance. Elsevier.

Appendix

Table 1. Proposed models based on specific examined energy commodity data sets.

Year	Author	Examined Data Set	Examined Energy Commodity	Outperforming Model
2002	Fong and See	Daily Spot	WTI oil	MS-GARCH
2006	Sadorsky	Daily Futures	WTI oil, heating oil, gasoline, natural gas	TGARCH and GARCH(1,1)
2007	Narayan et al.	Daily Spot	WTI oil	EGARCH
2008	Agnolucci	Daily Spot	WTI oil	GED-GARCH
2008	Kang et al.	Daily Spot	WTI oil, Brent oil, Dubai crude	FIGARCH and CGARCH
2009	Vo	Weekly Spot	WTI oil	MS-GARCH
2009	Choi and Hammoudeh	Daily Futures	WTI oil	FIGARCH
2009	Cheong	Daily Spot	WTI oil, Brent oil	Student-t GARCH, Student-t FIAPARCH
2010	Nomikos and Pouliasis	Daily Futures	WTI oil, Brent oil, heating oil, gasoline	MS-GARCH
2010	Luo et al.	Daily Spot	WTI oil	MS-GARCH
2010	Mohammandi and Su	Weekly Spot	11 crude oil types	APARCH
2010	Wei et al.	Daily Spot	WTI oil, Brent oil	Non-linear GARCH models
2011	Hou and Suardi	Daily Spot	WTI oil, Brent oil	Non-parametric GARCH
2011	Wang et al.	Daily Spot, Daily Futures	WTI oil	Non-parametric GARCH models
2011	Arouri et al.	Daily Spot,1&2-month Futures	WTI oil, heating oil, gasoline	FIGARCH
2012	Wang and Wu	Weekly Spot	WTI oil, heating oil, gasoline, jet oil	Multivariate GARCH models
2012	Kang and Yoon	Daily Futures	WTI oil, heating oil, gasoline	ARIMA-GARCH and ARFIMA-FIGARCH
2013	Cifter	Daily Spot	Electricity	Skewed Student-t MS-GARCH
2015	Zhang et al.	Daily Spot, Weekly Spot	WTI oil, Brent oil	MS-GARCH
2015	Gunay	Daily Spot	Brent oil	Student-t MS-GARCH
2017	Runfang et al.	1-month Spot	WTI oil	MS-GARCH
2018	Di Sanzo	Daily Spot	WTI oil, Brent oil	MS-GARCH