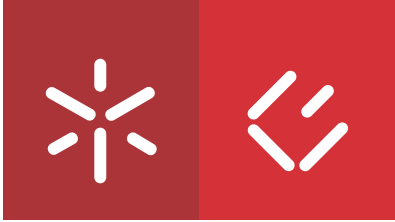


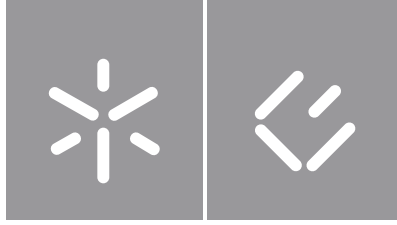


Monique Oliveira Moreira de Souza

Forecasting Crude Oil Prices Volatility with
GARCH Models

Universidade do Minho
Escola de Economia e Gestão





Universidade do Minho
Escola de Economia e Gestão

Monique Oliveira Moreira de Souza

Forecasting Crude Oil Prices Volatility with GARCH Models

A thesis presented for the degree of
Master in Monetary, Banking and Financial Economics

Research under the supervision of
Professor Cristina Alexandra Oliveira Amado

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença [abaixo](#) indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição

CC BY

<https://creativecommons.org/licenses/by/4.0/>

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

ACKNOWLEDMENTS

This dissertation represents the final stage of a long journey that became especially challenging due to the pandemic crisis that still persists, fortunately I found all the support I needed.

First, I would like to acknowledge my advisor, Professor Cristina Amado whose expertise and guidance were crucial not only for the results of this research work, but for my personal growth. Although we faced an atypical year, I learned a lot from all the honest and constructive criticisms during our online meetings.

Finally, I would like to acknowledge my beloved family and friends, especially my husband Daniel, for all the words of encouragement and kind gestures during all the challenging moments I faced in this thesis. You are always there for me.

RESUMO

As variações do preço do petróleo têm um grande impacto a nível financeiro em todo o mundo. Assim, o objetivo desta tese é realizar um estudo empírico de previsão da volatilidade dos preços do petróleo através da estimação de vários modelos univariados da classe GARCH para o período de Janeiro de 1986 até Dezembro de 2019.

As previsões conduzidas fora da amostra para horizontes de um, cinco e vinte dias foram obtidas através do *model confidence set* (MCS) proposto por Hansen et al. (2011) que consiste numa sequência de testes que retornam um ranking com o desempenho de cada modelo e cuja grande vantagem é a possibilidade de seleção de mais do que um melhor modelo, o que proporciona uma solução mais realista.

Os resultados empíricos indicam que os modelos que conseguem captar a persistência ou a longa memória na volatilidade e a assimetria na volatilidade do preço do petróleo, têm uma performance superior em todos os horizontes de previsão.

Palavras-Chave: Mercado do Petróleo, Previsões de Volatilidade, GARCH, Model Confidence Set

ABSTRACT

Changes in crude oil prices have a major impact on finances worldwide. Thus, the proposal of this thesis is to conduct an empirical study to forecast the volatility of crude oil prices by estimating several univariate GARCH-class models for the period from January 1986 to December 2019.

The out-of-sample forecasts horizons for one, five and twenty days were obtained through the model confidence set (MCS) procedure proposed by Hansen et al. (2011) which consists of a sequence of tests that return a ranking with the performance of each model and whose great advantage is the possibility of selecting more than one best model which provides a more realistic solution.

The empirical results indicate that models that captures the long-memory or volatility persistence and asymmetry in crude oil prices volatility perform better in all forecasting horizons.

Key Words: Crude Oil Market, Volatility forecasting, GARCH, Model Confidence Set

"Prediction is very difficult, especially if it's about the future."

Niels Bohr, Nobel laureate in Physics 1922

CONTENTS

1. Introduction	1
2. Brief Historical Overview of Crude Oil Price Returns	4
3. Literature Review.....	8
4. Methodology.....	14
4.1. The Class of GARCH Models	14
4.2. Forecasting Model Evaluation	18
5. Empirical Application.....	20
5.1. Preliminary Data Analysis	20
5.2. Estimation Results	24
5.3. Forecasting Crude Oil Volatility.....	29
6. Conclusions.....	33
7. References.....	35

LIST OF FIGURES

Figure 1. Crude Oil production by region between 1965 – 2018.....	5
Figure 2. Evolution chart of WTI prices between 1986 - 2019.....	8
Figure 3. Full-sample results of daily prices, 01/02/1986-12/31/2019.	20
Figure 4. Full-sample results of daily log-returns, 01/01/1986-12/31/2019.....	21
Figure 5. Autocorrelation function for log-returns.....	22
Figure 6. Autocorrelation function for absolute log-returns.	23
Figure 7. Autocorrelation function for squared log-returns.....	23

1. INTRODUCTION

The behavior of crude oil prices raises many concerns among regular consumers, industries, governments, and financial markets, as many uncertainties and changes related to crude oil prices have an immediate impact on the daily lives of millions of people worldwide. Therefore, much research has been done in order to answer a challenging question about the future: “Is it possible to make accurate forecasts about crude oil prices?”

It is well known that after several trends to measure risk became popular, volatility emerged as an important link for risk assessment and investment planning. As a result, Andersen and Bollerslev (1998) suggested that volatility forecasts may play the role of a barometer for financial markets vulnerability.

Consequently, many financial researchers adopted different volatility models in forecasting studies about performances and benchmark pricing which made volatility a tool often used to measure the intensity of unpredictable changes in asset returns.

Among these different volatility models, there are two time-varying models that have gained greater acceptance among researchers: the autoregressive conditional heteroskedasticity (ARCH) model introduced by Engle’s (1982) seminal paper and the generalized ARCH (GARCH) model proposed by Bollerslev (1986) which led to the development of several models to forecast volatility.

Recently, Zhang et al. (2019) estimated and forecasted crude oil prices volatility using several GARCH-type models and the MCS procedure of Hansen et al. (2011) to evaluate out-of-sample forecasting performance of single-regime and regime-switching models. They concluded that the regime switching models do not perform significantly better than the single-regime models using weekly and daily returns data.

In addition, Hasanov et al. (2020) also evaluated the forecasting performance of several univariate and multivariate volatility models of crude oil prices with the MCS procedure and concluded that when accounting for the model optimality and

structural breaks, the incorporation of asymmetric heavy-tailed distributions functions in the out-of-sample estimations leads to significant forecasting gains.

My contribution to the literature consists of forecasting the volatility of crude oil prices using several linear and nonlinear GARCH-type models that are widely accepted as statistical models in the vast literature regarding forecasting energy markets volatility to evaluate whether the incorporation of extended daily data and some statistical complexities, such as long-range dependence or long memory, volatility asymmetry and alternative distribution functions lead to significant gains in forecasting accuracy.

This dissertation is organized as follows. Chapter 2 presents a brief historical overview of crude oil prices which goes back more than 150 years. Chapter 3 presents the literature review where several contributions using GARCH volatility models designed to forecast future events were shown to be a good fit for modeling and forecasting crude oil prices.

Chapter 4 introduces the econometric methodology, such as the modelling estimation procedure, the optimal modeling approach and the forecasting evaluation method.

Chapter 5 describes and analyses the in-sample data, introduces the empirical study, carries out some robustness checks and discusses the empirical results for 1, 5, and 20-day-ahead forecasting horizons. For a better comprehension, this chapter is divided into three sections.

The first section approaches the in-sample data analysis concerning the daily prices and returns of West Texas Intermediate (WTI) from January 2nd, 1986 to December 31st, 2013 and applies tests of autocorrelation and heteroskedasticity effects.

The second section proceeds with the optimal modelling estimation and analysis of different ARMA-GARCH-type models considering two different distributions: the standard normal and the Student's t , since the correct specification of the conditional distribution of returns is an important step in modeling and forecasting crude oil volatility.

Although the standard normal it is a symmetric distribution since most observations are clustered around the mean and the most distant values on both tails are equally unlikely, that allows the fitting of many natural phenomena such as heights and blood pressure, and thus is considered the most important probability distribution, however, this study focuses in financial data that presents non-zero skewness and high kurtosis whereby, the Student's t-distribution can handle better data with fatter tails than the normal distribution.

Finally, the third section applies the MCS procedure to evaluate the out-of-sample forecasting performance computed by the different ARMA-GARCH-type models. Compared to conventional tests, the MCS procedure has two main advantages. Firstly, being the statements about significance. Since the MCS test does not need a pre-set benchmark or a pairwise comparison, it is especially helpful when there is a large number of models to be compared since the MCS test is performed only once and the output reports the p-value of each model. Secondly, the MCS procedure generates a superior set of models that allows the possibility of more than one best model which provides a more realistic solution since there is no agreement on the best model for forecasting volatility among researchers.

Chapter 6 summarizes and highlights the most important conclusions found in this study. Overall, the results indicate that none of the volatility models estimated outperforms the others in all forecasts horizons. However, models that capture asymmetry and volatility persistence or long-memory perform better in all forecasting horizons.

Tableau (version 2020.4) was used for numerical and graphical analysis shown in Chapter 3, while the code I developed for this study for Chapters 5 and 6 used the R software (version 4.0.3) combined with the rugarch package developed by Ghalanos (2016) and MCS package developed by Catania and Bernardi (2015).

2. BRIEF HISTORICAL OVERVIEW OF CRUDE OIL PRICE RETURNS

Before presenting the empirical study, a brief historical contextualization of how the era of crude oil began, since over the 150 years of its history, the oil industry has gone through clearly distinct phases that is important to understand and highlight, in order to obtain accurate forecasts of transition periods and periods of stability in the future.

The modern oil industry can trace its origins in 1859 when Edwin Drake became known as the "Father of the American Petroleum Industry" for inventing technology able to drill the first successful commercial oil well in Titusville, Pennsylvania. However, in those early days, oil was used only for domestic lighting as a substitute for whale oil.

During the mid-1950s, technological advances boosted logistics due to the high density and relative abundance of crude oil. Its production was finally revolutionized, and a large-scale industry was launched, marking the beginning of crude oil era that remains until nowadays, whereby modern society rely mostly on crude oil to sustain industry, heat buildings and fuel machines to transport goods and people across the globe, making crude oil the world's most important source of energy, despite being a fossil fuel¹.

The chart from Figure 1 shows that global oil production more than doubled over the last 50 years. The contribution of each region to the global total remained generally consistent, being the Middle East the world's largest crude oil producer in 2018. The region represented by Commonwealth of Independent States (CIS) unites Azerbaijan, Armenia, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, Turkmenistan, Uzbekistan, and Ukraine.

¹ Source: BP's 2015 Statistical Review of World Energy pp. 1–48

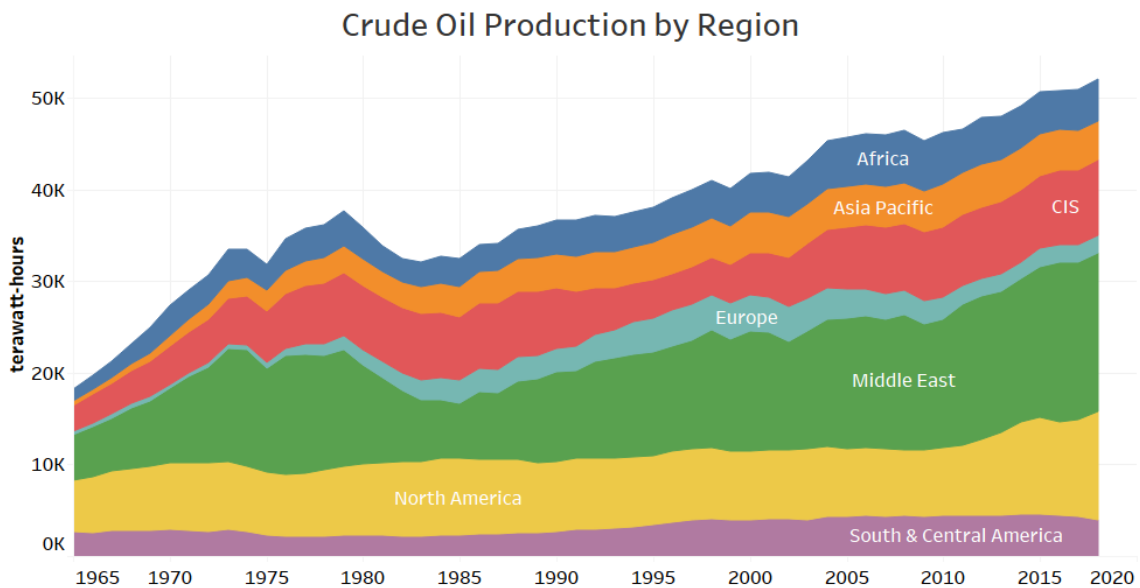


Figure 1. Crude Oil production by region between 1965 – 2018

In 1960, the Organization of the Petroleum Exporting Countries (OPEC) emerged through an agreement signed between Iran, Iraq, Kuwait, Saudi Arabia, and Venezuela whose goal was to increase the crude oil price produced by all five oil-exporting countries and simultaneously play a leading role in the crude oil competitive market.

During the 1970s there was a significant increase in crude oil global price, mainly due to the geopolitical conflicts that constitute the most important events of the 20th century, affecting the energy economy and causing major impacts on the economies of oil-importing countries which reflected the effectiveness of OPEC and temporarily acquired control of the global crude oil market.

In 1973 occurred the OPEC's Arab oil embargo, also known as the first oil shock that cut oil production, in response to the United States decision to support Israel in the Arab-Israeli war, causing prices to quadruple in 1974.

The second oil shock occurred in 1979-80 with the slowdown in crude oil production due to the Iranian Revolution which nearly stopped the oil production during the war, the global supply of crude oil was limited and its price more than doubled between 1978 and 1981 as the Iran-Iraq war prevented oil production.

In the early 1980s, the opposite happened as a result of the energy crisis of the 1970s, a lower consumer demand followed which caused an 'oil glut' and a consequent decrease in crude oil prices.

In 1990, the first Gulf War occurred when Iraq invaded Kuwait, causing the crude oil prices to skyrocket from \$34 per barrel to nearly \$77. After a U.S.-led military coalition managed to remove Saddam Hussein's Iraqi forces from Kuwait in early 1991, the prices dropped again.

In 2007, after the subprime mortgages depreciation quickly turned into an international banking crisis in 2008, the world emerged in a severe financial crisis, consequently oil prices had a sharp decline, as the general drop in asset prices increased the demand for energy around the world decreased as well as the demand for crude oil.

Through the 2010s there was again an 'oil glut' due a combination of many factors. During 2014–15 OPEC members consistently exceeded their production ceiling and China, the world's biggest importer of crude oil, experienced a sharp slowdown in economic growth. At the same time, in response to record oil prices, shale oil production in the United States almost doubled from 2008 levels, due to significant technology improvements.

In 2016, all these factors caused a collapse in oil prices because, although the supply glut increased it failed to meet the expected global growth. Still, according to BP's 2019 Statistical Review of World Energy, in 2018 OPEC accounted for nearly 44% of global oil production and 81.5% of the world's proven oil reserves.

At the present OPEC counts with eight more nations, in addition to the five founding countries: Libya, United Arab Emirates, Algeria, Nigeria, Gabon, Angola, Equatorial Guinea, and Congo, totalizing 13 members.

Recently, on April 20th, 2020, the crude market made history for the worst reasons, as WTI ended the day at minus \$37 per barrel, entering on negative territory for the first time. Although the time series analysis for this study is between 1986 and 2019 is inevitable to mention that the effect of COVID-19 pandemic on crude oil prices represent the biggest contraction ever on its price's history.

Another peculiarity about crude oil, besides its high influence in energy markets and great acceptance among financial assets, is its tangible dimension, that contribute to extremely high levels of price volatility, such as future supplies and reserves, political upheavals, adverse weather conditions, issues with the refineries or pipelines and world crises.

There are three factors that define crude oil grades and directly affect barrel prices: the American Petroleum Institute (API) gravity which represents an index that measures crude oil density and normally ranges from 15 to 45 degrees, where a lower density means a higher API which indicates a light crude and the opposite a heavy crude; the sulfur content which is characterized as sweet when has fewer impurities, but is harder to find, or sour when the sulfur level is higher than 0.5%, however is less expensive due to lower quality; and finally, the proximity to tidewater and refineries.

Thus, it is easy to understand why WTI is also known as “Texas Light Sweet”, since it has an API gravity around 39.6 degrees per barrel rating it as a light crude, contains about 0.24% sulfur, rating it as a sweet crude which is not only used as a benchmark in oil pricing but also as the underlying commodity of NYMEX oil future contracts.

As WTI is a light crude oil, its properties and production site make it ideal for being refined in the United States, mostly in the Midwest and Gulf Coast regions; for those reasons, the town of Cushing, in Oklahoma has been the price settlement point for WTI on the NYMEX for over three decades and became the major oil supply hub for crude oil in North America.

WTI historical spot prices FOB² in U.S. dollars per barrel since January 2nd, 1986 can be downloaded free of charges from U.S. Energy Information Administration (EIA) website. The graph from Figure 2 represents the annual average price for WTI.

² Free on Board is a transportation term that indicates that the price for goods includes delivery at the seller's expense to a specified point and no further.

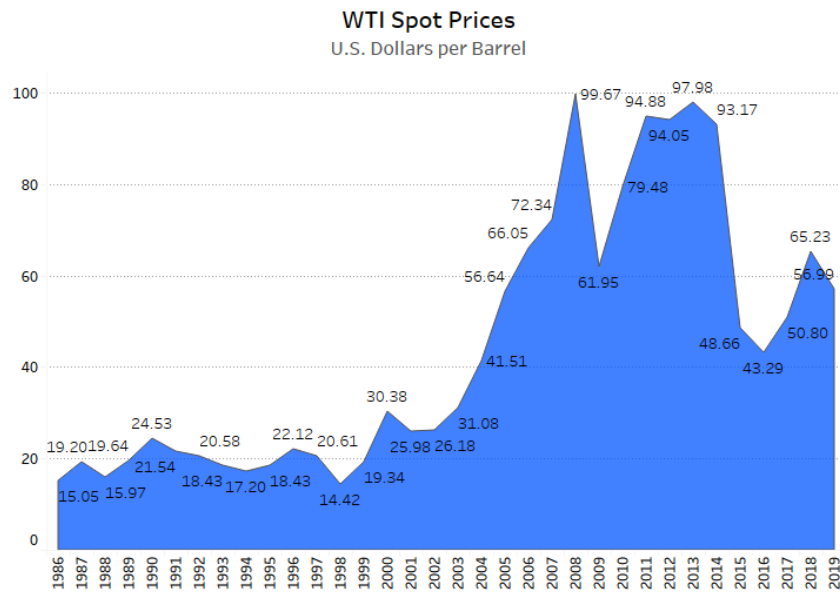


Figure 2. Evolution chart of WTI prices between 1986 - 2019

WTI prices have increased dramatically, rising from just 15.05 dollars per barrel in 1986, to a peak of 99.67 dollars per barrel in 2008.

3. LITERATURE REVIEW

The revolution in modeling and forecasting volatility began in 1982, and had as starting point the seminal paper of Engle whom instead of assuming a constant one-period forecast variance, as in traditional econometric models, presented a new class of stochastic processes to model the effects of serial correlation and conditional heteroskedasticity that became known as the autoregressive conditional heteroskedasticity (ARCH) model (Engle, 1982). The model describes important stylized facts, very common in financial returns, such as fat-tailed distributions which have a much higher probability of extreme profits or losses than predicted by the normal distribution; volatility clustering characterized by large price changes being followed by large price changes and small price changes being followed by small price changes; long-range dependence, also called long memory or long-range persistence which is observed when the autocorrelation function of absolute or squared returns decays slowly across lags; and finally, the asymmetry of gains and losses which denotes that the distribution of losses has a heavier tail and, therefore, a greater influence on future volatilities than the distribution of gains.

A few years later, the generalized autoregressive conditional heteroskedasticity (GARCH) models was proposed by Bollerslev (1986) with a more flexible lag structure, capable of representing current variances in terms of past variances, allowed GARCH models to quickly gain greater acceptance due to its high performance in modelling and forecasting time-dependent volatility.

Palm's (1996) extensive literature about the developments of the first 15 years in modeling time-varying volatility using ARCH and GARCH processes on financial applications demonstrated that empirical applications achieved significant improvements, especially on statistical properties regarding financial time series as they were not accounted for any existing models until then.

Even nowadays, the importance of ARCH and GARCH models as tools to model time series data is indisputable. However, since it is intrinsically symmetric, when forecasting skewed time series, the asymmetric GARCH model often leads to biased results. Consequently, several extensions and alternative specifications for the standard GARCH model have been presented in an attempt to appropriately take into account the volatility persistence and assist in modeling conditional volatility, but so far a complete model covering a variety of stylized facts is yet to be proposed.

Bollerslev and Ole Mikkelsen (1996) and Granger and Ding (1996) research works about non-stationary behavior of volatility using long-memory persistence became very popular as they focused on the volatility of high-frequency data, especially in financial time series and crude oil prices is no exception, as persistence shows that shocks to conditional variance die at a hyperbolic rate which is slower than the exponential rate of the decay of shocks in GARCH models as accounted by Baillie (1996), whereby in this dissertation I shall use a sample period longer than 30 years, as it makes sense to apply a modelling procedure for long-term volatility movements over long observation periods.

Unfortunately, when volatility persistence is high forecasts are harmed because when fitting a single-regime GARCH models on crude oil price changes, potential structural changes may be ignored and the model persistence is not reduced causing often inaccurate predictions. To solve this issue, regime-switching GARCH models were introduced into the GARCH-type models, in order to model nonlinear behavior

and address potential structural changes contributing to the improvement of asset prices volatility forecasts.

Hamilton (1988, 1989) proposed the Markov Regime-Switching GARCH (MRS-GARCH) model to allow the regimes in the Markov chain to have different volatility structures, capturing effectively potential state transition and nonlinearity which were relevant to Zhang and Wang (2015) study about the reasons behind WTI strong volatility from 2003 to 2012.

The first surprising issue was the difference between the market-trading price, that is influenced by the speculation in crude oil markets and the fundamental crude oil price, that is determined by crude oil supply and demand which fluctuated less than the market-trading price.

Thus, Zhang and Wang (2015) defined the deviation between crude oil market-trading price and its fundamental price as crude oil price bubbles and measured them with multivariate regression model and then employed the MRS model of Hamilton (1989) to measure the evolution of crude oil price bubbles over time.

The empirical results indicated that during these price bubbles process, prices evolved between two regime-switching states: the stable state and the upheaval state. The price bubbles regime-switching process also followed a time homogeneous first-order Markov chain that made possible to statistically differentiate these two states during the price bubbles processes concluding that, in most cases, the stable states were often observed after upheaval states, being the stable state more persistent than the upheaval state, because during most of the time there were not huge surprising events during the sample period which highlight the importance of future policy actions in order to prevent instabilities in crude oil markets.

Li et al. (2013) also proposed a regime-switching model named Mixture Memory GARCH (MM-GARCH) which is a mixture of the standard GARCH and the Fractionally Integrated GARCH (FIGARCH) model proposed by Baillie et al. (1996) and became often used to describe the long-range dependence on conditional variances of financial time series which was also supported by the empirical study by Zhang et

al. (2019) for in-sample data, where they concluded that regime-switching models perform better than the single-regime models.

More recently, Zhang et al. (2019) employed a great number of linear and nonlinear GARCH-type models to forecast the volatility of daily and weekly WTI and Brent crude oil prices, from 1986 until June 2017 the single-regime GARCH, GJR-GARCH and EGARCH models were applied.

The empirical results showed that, although the MRS-GARCH models provided superior results than the single-regime GARCH and MM-GARCH models in in-sample results, when confronted with out-of-sample experiments the single-regime GARCH models are not inferior than the regime-switching GARCH models, concluding that the regime-switching models do not provide an indisputable improvement in the accuracy of forecasts for crude oil price volatility. As a matter of fact, Zhang et al. (2019) recommended that volatility forecasting models should not simply elect multiple regimes for crude oil prices because the simplest models as the single-regime models in most cases outperforms regime-switching models.

Kang et al. (2009) investigated the efficacy of volatility long-memory models and its performance to forecast Brent, Dubai and WTI crude oil daily prices, from 1992 until 2006 combining the Component GARCH (CGARCH) model proposed by Ding and Granger (1996) and Engle and Lee (1999) which has one component to capture short-term and another component to capture the long-term effects of volatility, together with FIGARCH model proposed by Baillie (1996). The out-of-sample forecasts achieved better results and proved that persistence or long-memory models are more useful than the linear GARCH and the Integrated GARCH (IGARCH) models proposed by Engle and Bollerslev (1986). However, it is important to highlight that the long-memory volatility CGARCH and FIGARCH models cannot capture the asymmetric leverage effect.

Nevertheless, when Kang et. al (2009) paper was extended a further analysis conducted by Wei et. al (2010) employed a large number of linear and nonlinear GARCH-type models and six loss functions as a forecasting criteria and the Superior Predictive Ability (SPA) test of Hansen (2005) to ensure more robust results, but unlike Kang et. al (2009), none of the nine GARCH-type models used (including the FIGARCH) demonstrated to be superior to the others, across any loss functions for

both Brent and WTI prices. However, the nonlinear GARCH-type models could capture long-memory and asymmetric volatility with greater accurate forecasts than the linear ones, especially over long-time horizons, warning that when a volatility model is chosen it should not only take into account existing research, but also the correlation between the sample data, the loss functions and ultimately, the forecast aim.

Recently, Hasanov et al. (2020) paper examined a greater number of volatility models, their forecasting performances and the relevance of some statistical complexities regarding four types of petroleum futures contracts traded on NYMEX.

In this work nine univariate models were used, such as the random walk, historical mean, moving average, exponentially smoothing, linear regression, autoregressive, GARCH, threshold GARCH, GARCH-in-mean, the vector autoregression and bivariate GARCH to forecast WTI daily volatility of heating oil, unleaded gasoline and natural gas futures contracts, from February 1988 to December 2018³. They also polled all single-regime models, with two-state Markov-Switching models proposed by Haas et al., (2004) under three different distributions (Gaussian, Student's t, and skew Student's t) and used the MCS test of Hansen et al. (2011) to analyze the forecasting accuracy and account for structural breaks.

The results showed that the Markov-Switching models are particularly useful for heating oil and unleaded gas markets but for crude oil and natural gas the results remained unchanged across many forecasting horizons.

In addition, the authors included statistical models on the out-of-sample volatility forecast analysis that widely confirmed the results obtained on Sadorsky (2006), as GARCH-type models appeared to be superior in one-period forecasts results on crude oil futures case.

Unlike previous forecasting studies, related with the energy market volatility using GARCH-type models which ignored the optimality of the mean equation in the

³ The only exception is the unleaded gasoline data which was discontinued after December of 2006

context of a single moving window approach Hasanov et al. (2020) considered the relevance of the mean equation optimality in addition to the significance of conditional distributions, using a large number of the moving windows to find out whether the introduction of statistical complexities such as regime switches and alternative distribution functions lead to any forecasting accuracy gains. In most cases, results revealed that models with optimal mean were greater than models that ignore optimality in the mean which highlights the importance of optimality in the mean of GARCH-type models.

In sum, until the early 2010s there were few articles analyzing the key features of crude oil prices volatility because once the standard GARCH model cannot capture persistence or long memory, there is not an agreement about the best way of modelling and forecasting crude oil prices volatility and neither about a specific GARCH-type model, that consistently proved to be superior, as discussed in studies by Kang et al. (2009), Wei et al. (2010), Zhang et al. (2019) and Hasanov et al. (2020). In addition, since the combination of GARCH models with loss functions during forecasts estimation may lead to heterogeneous results, authors also recommend caution when combining volatility models with loss functions to evaluate forecasting performances.

Bhowmik and Wang (2020) found that in the last 10 years there have been significant improvements in the research work and, although GARCH-type models have increased and presented excellent models, each model has its own strengths and weaknesses. Therefore, in order to achieve better results, other statistical techniques must be combined. However, the authors observed that few researchers used multivariate GARCH statistical techniques to analyze market volatility and returns and asymmetric GARCH models such as EGARCH, GJR-GARCH and TGARCH were often used to capture the effect of negative news on volatility of stock returns.

Finally, the authors also concluded that when accounting for the model optimality, structural breaks and the use of asymmetric heavy-tailed distribution functions in the estimations, lead to significant forecasting accuracy gains.

4. METHODOLOGY

4.1. THE CLASS OF GARCH MODELS

ARCH-type models were the pioneers in modeling conditional variance processes, since they were the first to allow the incorporation of volatility clustering and mean reversion which became instantly a very flexible and powerful tool to describe the time-varying nature of conditional volatility that characterizes financial assets.

Therefore, nowadays the most popular volatility model is the generalized ARCH of Bollerslev (1986), since it is more parsimonious than the ARCH model and its simplicity of implementation contributed to its great acceptance and since GARCH also avoids overfitting thanks to the ability to account for the main stylized facts common in financial returns, GARCH models became widely accepted in quantitative finance to measure and forecast volatility.

However, despite its popularity, GARCH models treat shocks symmetrically which led to the proposal of several asymmetric extensions in the financial econometrics literature and thus the GARCH-family models specifications emerged to jointly estimate the conditional mean and conditional variance equations.

This thesis will focus on the following seven univariate models to model and forecast crude oil returns volatility: the standard GARCH, EGARCH, GJR-GARCH, TGARCH, CGARCH, APARCH, NGARCH models.

The GARCH model of Bollerslev (1986) uses both an autoregressive and moving average structure to the conditional variance. Since the return series may be linearly dependent, an Autoregressive Moving Average (ARMA) model may be needed for the conditional mean and the GARCH model to fit the conditional variance.

In this dissertation the continuously compounded return series r_t shall have the following ARMA(m,n)-GARCH(p,q) specification:

$$\begin{aligned} r_t &= E(r_t|\Omega_{t-1}) + \mu_t \\ \mu_t &= \varepsilon_t \sqrt{h_t} \end{aligned} \tag{1}$$

where Ω_{t-1} is the information set up to time $t - 1$, ε_t is the independently and identically distributed (i.i.d.) standardized innovations conditionally with a standard normal distribution or the Students't t-distribution.

The conditional mean can be written as follows:

$$r_t = C + \sum_{i=1}^m \phi_i r_{t-i} + \sum_{j=1}^n \theta_j \varepsilon_{t-j}, \quad (2)$$

where r_t is the daily returns, C is a constant term, ϕ_i are the parameters of the autoregressive component of order m , θ_j are the parameters of the moving average component of order n , and both orders m and n are non-negative integers.

$$h_t = \omega + \sum_{i=1}^p \beta_i h_{t-i} + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2, \quad (3)$$

where ω is the long-run volatility with condition $\omega > 0, \beta_i \geq 0; i = 1, \dots, p$ and $\alpha_j \geq 0; j = 1, \dots, q$ to ensure the non-negativity of the variance process. The sum of $\beta_i + \alpha_j$ quantifies the persistence of shocks to volatility, if equal to the unity then every shock has an infinite persistence. When $\beta_i + \alpha_j < 1$, then GARCH(p,q) model is covariance stationary.

A more specific model is the ARMA(1,1)-GARCH(1,1) model that is defined as:

$$r_t = C + \phi r_{t-1} + \theta \varepsilon_{t-1} \quad (4)$$

$$h_t = \omega + \beta h_{t-1} + \alpha \varepsilon_{t-1}^2 \quad (5)$$

A more flexible GARCH extension is the exponential GARCH (EGARCH) model proposed by Nelson (1991) who argued that the non-negative restrictions on parameters α and β in the standard GARCH are very restrictive, thus the EGARCH does not have any parameters restrictions. EGARCH also accommodates asymmetry where negative shocks can have a larger impact on volatility than positive shocks and by modeling the logarithm of the variance, will always be positive regardless of the sign of the parameters. EGARCH is defined as:

$$\log(h_t) = \omega + \alpha z_{t-1} + \gamma (|z_{t-1}| - E|z_{t-1}|) + \beta \log(h_{t-1}), \quad (6)$$

where γ is the asymmetric leverage coefficient that describes the volatility leverage effect. EGARCH models allows good and bad news to have different impact on volatility since a negative parameter γ indicates that negative returns have a larger impact on volatility than positive returns.

As already addressed in section 4.1 volatility clustering is a phenomenon that comes and goes, since major changes in asset prices are often followed by other major changes, and minor changes are often followed by minor changes as well, this asymmetry can be modeled using the threshold GARCH (TGARCH) models of Zakoian (1994). The variance equation is given by:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma \varepsilon_{t-1}^2 I_{(\varepsilon_{t-1} < 0)}, \quad (7)$$

where $I(\cdot)$ is a dummy variable that assumes unity if $\varepsilon_{t-1} < 0$ and zero otherwise. TGARCH model allows asymmetric effects by including a dichotomous dummy variable into the standard GARCH model.

The Glosten-Jagannathan-Runkle-GARCH (GJR-GARCH) model developed by Glosten et al. (1993) was designed to capture the potential larger impact of negative shocks on return volatility, so is commonly named as asymmetric leverage volatility effect. GJR-GARCH is similar to TGARCH, the only difference is the variance instead of the standard deviation in the specification, as follows:

$$h_t = \omega + [\alpha + \gamma I_{(\varepsilon_{t-1} < 0)}] \varepsilon_{t-1}^2 + \beta h_{t-1}, \quad (8)$$

where $I(\cdot)$ is an indicator function which means that, when the condition in $(\cdot)$ is met, $I(\cdot) = 1$ and 0 otherwise; γ is the asymmetric leverage coefficient which describes the volatility leverage effect.

Another interesting specification is the CGARCH model of Engle and Lee (1993) that investigated the long and short-run memory of volatility. CGARCH models decomposes the conditional variance into a permanent (ξ_t) and transitory component, where one component captures the short-term and another the long-term effects of shocks. The specification for the conditional variance is given by:

$$h_t = \xi_t + \alpha(\varepsilon_{t-1}^2 - \xi_{t-1}) + \beta(h_{t-1} - \xi_{t-1}), \quad (9)$$

$$\xi_t = \omega + \rho\xi_{t-1} + \eta(\varepsilon_{t-1}^2 - h_{t-1}), \quad (10)$$

where ξ_t represents the long-run volatility, the past forecast error ($\varepsilon_{t-1}^2 - h_{t-1}$) provides the driving force for the time-dependent movement of ξ_t , and the difference between the conditional variance and its long-run volatility ($h_{t-1} - \xi_{t-1}$) defines the transitory component of volatility. The transitory component then converges to zero with powers of $(\alpha + \beta)$, whereas the permanent component converges on $\omega/(1 - \rho)$ with powers of ρ . It is assumed that $0 < \alpha + \beta < \rho < 1$ for the long-run volatility component to be more persistent than the short-run volatility component.

The Asymmetric-Power-ARCH (APARCH) models of Ding et al. (1993) represents a general class of models that combine both ARCH and GARCH models. The specification for its conditional variance is given by:

$$h_t^\delta = \omega + \alpha(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^\delta + \beta h_{t-1}^\delta, \quad (11)$$

where $\omega \geq 0$; $\alpha, \beta, \delta > 0$ and $\gamma \in [-1, 1]$. The parameter δ performs a Box-Cox transformation and γ reflects the leverage effect, so a positive γ indicates that negative returns have a bigger impact on the conditional variance than positive returns.

The APARCH model nests several ARCH extensions as special cases which means that some models can be derived or estimated from this model. For example, the standard GARCH when δ and $\gamma = 0$, the GJR-GARCH when $\delta=2$, the TGARCH when $\delta=1$ and the NGARCH when $\gamma=0$ and $\beta=0$.

The Nonlinear GARCH (NGARCH) model proposed by Higgins and Bera (1992) introduced a smooth transition from the standard GARCH models through a correlation between financial markets and volatility processes where a lagged value of conditional variance (h_{t-1}^δ) performs as a transitional variable in order to find useful models for highly persistent volatility. Since the high persistence in conditional variance is related to relatively unusual changes in regime, this can be captured by an adequate specification of the new model, as follows:

$$h_t^\delta = \omega + \alpha(|\varepsilon_{t-1}|)^\delta + \beta h_{t-1}^\delta, \quad (12)$$

where parameter ω determines the level of the conditional variance h_t^δ . Thus, as demonstrated previously, NGARCH can be represented as the APARCH model without the leverage effect.

In sum, these models were chosen due to their heterogeneity since each of them focuses on a different kind of stylized fact. Although some specifications seem similar, the way each model account for the stylized fact is different.

4.2. FORECASTING MODEL EVALUATION

Many authors have argued that in practice when alternative models' forecasts are combined the accuracy is better than forecasts based on a single "best" model. Since observations are subject to structural breaks and models are subject to different levels of misspecification, a strategy that groups information from different models usually performs better than methods that only try to select the best forecast model.

To fill this gap, emerged the MCS test that compares volatility forecasts computing several alternative model specifications with test statistics and returns a selection from a smaller set of "superior" models named "Superior Set Models" (SSM), where the null hypothesis of equal predictive ability (EPA) is not rejected at a pre-specified confidence level α . MCS procedure has a predefined loss function matrix of dimension $r \times m$ where m is the competing models with a series of r 1-day-ahead forecast values.

The initial set of models is denoted by:

$$M^0 = \{1, \dots, m\} \quad (13)$$

Where MCS ranks models according to their expected loss combining the following widely used statistical loss functions to test and compare the prediction ability of the models on the out-of-sample data:

$$AE_{1,t+1} = |\tilde{v}_{t+1} - \hat{v}_{t+1}|, \quad (14)$$

$$AE_{2,t+1} = |\tilde{v}_{t+1}^2 - \hat{v}_{t+1}^2|, \quad (15)$$

$$QLIKE_{t+1} = \log(\hat{V}_{t+1}^2) + \tilde{V}_{t+1}^2 \hat{V}_{t+1}^2, \quad (16)$$

$$SE_{1,t+1} = (\tilde{V}_{t+1} - \hat{V}_{t+1})^2, \quad (17)$$

where the loss functions AE_1 and AE_2 denotes the absolute error, $QLIKE$ denotes the Quasi-Likelihood and SE_1 denotes the squared error, $\tilde{V}_{t+1}$ is a proxy for actual volatility, $\hat{V}_{t+1}$ is the forecasts of volatility generated by the considered ARMA-GARCH-type models at time t .

The loss differential between models i and j is given by:

$$d_{ij,k} = \ell_{i,k} - \ell_{j,k} \text{ for } i, j = 1, \dots, m \text{ and } k = 1, \dots, R \quad (18)$$

where $d_{ij,k}$ indicate the difference in the losses between models i and j . This iterative procedure continues until the null hypothesis of EPA is not rejected at a fixed level of confidence $(1 - \alpha)$ for each step in the testing procedure, where MCS procedure ($\hat{M}_{1-\alpha}^*$) sequentially eliminates the models with worst forecasting performance from the first set of models M^0 .

Finally, when the null hypothesis is accepted the procedure ends and the SSM is created with all the models that remain after the consecutive eliminations and usually a smaller SSM or ideally a single model, M^* , of dimension $m^* \leq m$ is selected.

According to Hansen et al. (2011) the computation of the range statistic (T_R) is given by:

$$t_{ij} = |\bar{d}_{ij}| / [\widehat{var}(\bar{d}_{ij})]^{1/2} \text{ for } i, j = 1, \dots, m \quad (19)$$

where $\bar{d}_{ij} = R^{-1} \sum_{k=1}^R d_{ij,k}$ denotes the relative sample loss between the i -th and j -th models, while $\widehat{var}(\bar{d}_{ij})$ is the bootstrapped estimate of $var(\bar{d}_{ij})$. The statistic t_{ij} is used to decide on the following hypotheses:

$$H_{0,M}: c_{ij} = 0 \text{ for all } i, j = 1, \dots, m \quad (20)$$

$$H_{1,M}: c_{ij} \neq 0 \text{ for some } i, j = 1, \dots, m$$

where $c_{ij} = E(d_{ij})$ is assumed to be finite and time invariant.

The initial set of models M^0 for this study contains seven ARMA-GARCH-Type models specifications and considers two different distributions: the standard

normal and the standardized Student-t, combining a total of 14 models' specifications for WTI prices volatility.

5. EMPIRICAL APPLICATION

5.1. PRELIMINARY DATA ANALYSIS

The data for this study consists of 8569 observations obtained from the U.S. Energy Information Administration on the daily spot prices (in US dollars per barrel) of West Texas Intermediate (WTI) traded on NYMEX. The in-sample data is plotted in blue and the out-of-sample data is plotted in red in Figure 3 which consists in exactly 34 years long, from 2nd January 1986 to 31st December 2019.



Figure 3. Full-sample results of daily prices, 01/02/1986-12/31/2019.

However, this dataset was divided into an in-sample set with the first twenty-eight years from 1986 to 2013 with 7064 observations for the in-sample analysis and modeling, then the out-of-sample set includes the last six years from 2014 to 2019, with the 1505 final observations which shall be used for the out-of-sample analysis that will be conducted in the next section.

The plot in Figure 4 illustrates the daily returns, where periods of low returns are often followed by periods of high returns which highly suggests the presence of volatility clustering.

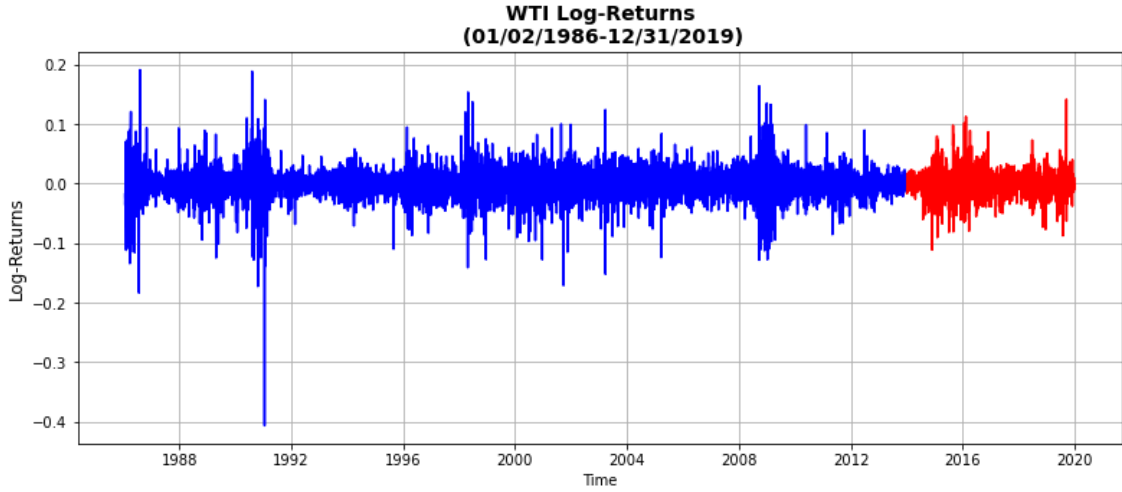


Figure 4. Full-sample results of daily log-returns, 01/01/1986-12/31/2019.

Most financial studies involve returns series rather than prices series, first because investors are more interested in assets returns for investment opportunities and secondly, returns series have more attractive statistical properties than prices, since a return is a percentage defined as the change of price expressed as a fraction of the initial price. Therefore, makes more statistical sense to analyze returns data rather than price series.

Thus, the transformation of prices to returns in this study are based on the logarithmic returns function, given by:

$$r_t = \ln p_t - \ln p_{t-1} \text{ for } t = 1, 2, \dots, T \quad (21)$$

where p_t represents WTI daily spot price at time t and r_t represents WTI daily returns at time t . Following Sadorsky (2006) and Kang et al. (2009) the daily volatility or variance will be estimated by daily squared returns.

The descriptive statistics in Table 1 for the in-sample period show that the mean of the return series is close to zero, the skewness is negative, kurtosis assumes a value greater than the normal value of 3 which indicates the presence of excess kurtosis in daily returns which means that the probability of obtaining an extreme outcome either extremely large or extremely small is higher and thus riskier. However, the low standard deviation indicates that returns are clustered around the mean.

Table 1. Descriptive Statistics – 01/02/1986-12/31/2013

	Prices	Returns
Observations	7064	7063
Mean	40.8489	0.0002
Median	25.45	0.0007
Minimum	10.25	-0.4064
Maximum	145.31	0.1915
Standard Deviation	30.0394	0.0254
Skewness	1.1029	-0.7627
Kurtosis	-0.0753	14.9717

The plot of the autocorrelation of a time series by lag is called the autocorrelation function (ACF) and is a useful qualitative tool to assess the presence of linear dependence in the returns at individual lags. The ACF for log-returns plotted in Figure 5 up to 250 lags, shows that several both lower and higher order autocorrelation coefficients are significantly different from zero.

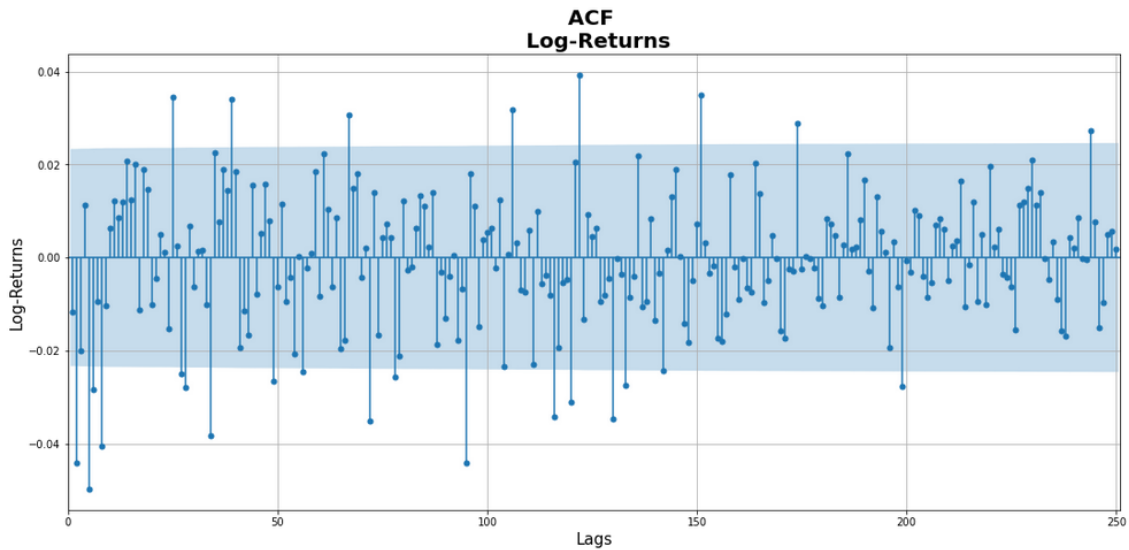


Figure 5. Autocorrelation function for log-returns.

The ACF for absolute returns and squared returns plotted in Figures 6 and 7, respectively, shows that the autocorrelation coefficients are statistically significant within 5% up to the first half, that is approximately until the 125th lag, are significantly different from zero which indicates strong dependence in the higher order movements in the returns, providing evidence for the presence of time-varying volatility in the oil returns.

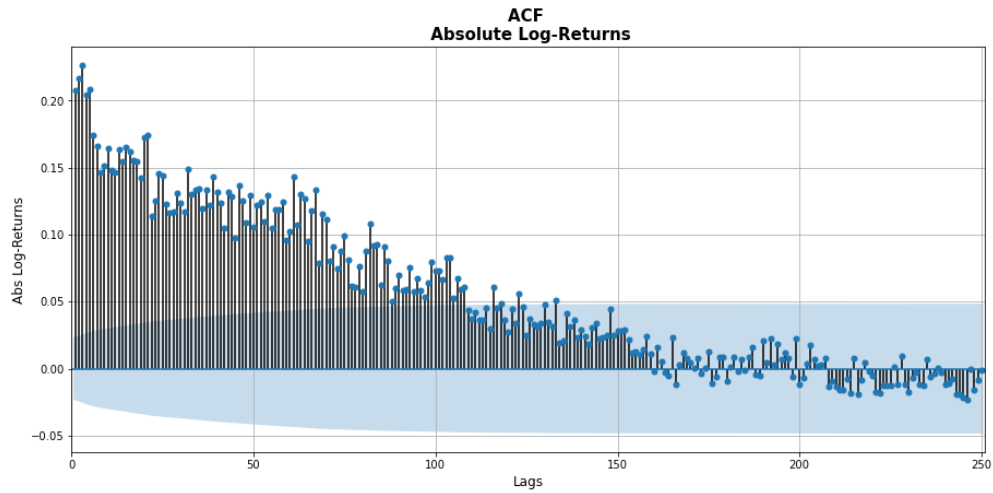


Figure 6. Autocorrelation function for absolute log-returns.

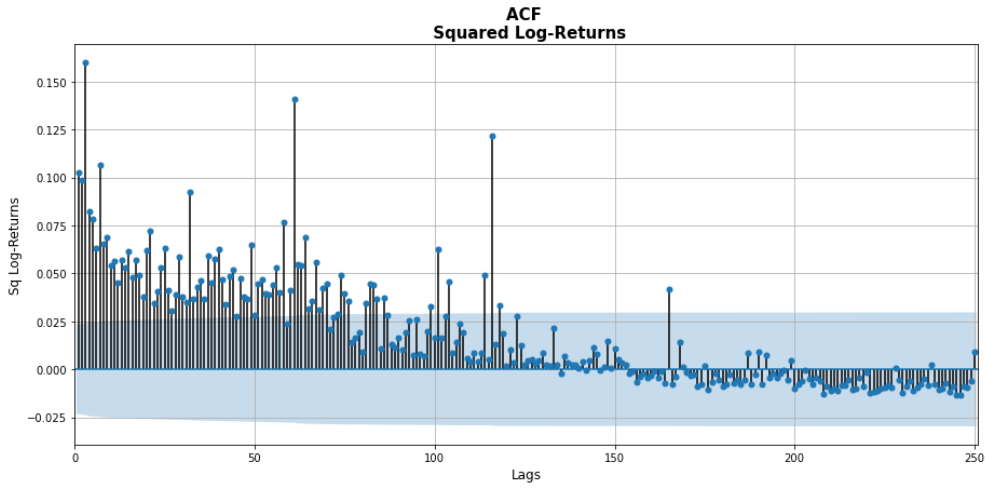


Figure 7. Autocorrelation function for squared log-returns.

In order to formally detect if the returns have serial autocorrelation, the Ljung–Box test proposed by Ljung and Box (1978) was applied to test whether the returns are jointly autocorrelated at multiple lags.

The results of Table 2 show that: the null hypothesis of no serial autocorrelation until the 10th order is rejected as the p-value indicates that there is significant autocorrelation in the returns at the 5% level of significance. Therefore, the hypothesis that the return series is white noise is rejected and conclude there is strong evidence of autocorrelation in the data.

Table 2. Ljung-Box test for autocorrelation

Lags	χ^2	p-value	Reject H_0
1	14.179	8.336 e-04	True
5	36.501	7.541e-07	True
10	52.602	8.826e-08	True

Finally, the ARCH-Lagrange Multiplier (LM) test introduced by Engle (1982) is applied to test for conditional heteroskedasticity in the returns.

Table 3. Engle's ARCH-LM test for conditional heteroskedasticity

Lags	χ^2	p-value	Reject H_0
1	76.753	< 2.2e-16	True
5	307.19	< 2.2e-16	True
10	363.02	< 2.2e-16	True

The results of Engle's ARCH-LM test shown in Table 3 assess the significance of ARCH effects for WTI returns. The p-value indicate that the null hypothesis of no ARCH effects is rejected at the 5% level of significance which shows the presence of ARCH effects in WTI returns, therefore, the GARCH model is necessary to model in order to capture the presence of conditional heteroskedasticity in the returns.

5.2. ESTIMATION RESULTS

In this study, volatility forecasts are computed using an estimation window of twenty-eight years which consists of 7,063 daily returns observations to cover both economic expansions and contractions. The length of the forecast periods will be 1, 5 and 20-day-ahead.

As already discussed in Chapter 4, the modelling of the conditional variance shall be the focus of this work. However, before conducting the empirical results of variance forecasting, the in-sample data was examined to determine which types of models perform best in-sample. If the full sample were applied it would result in very rigid models with flat variance term structures.

The approach for modeling the conditional mean and the conditional variance was to search over several alternatives of $ARMA(m, n) - GARCH(p, q)$ models where m and n ranged between 0 to 3 while p and q parameters were either 1 or 2, since in most cases that is sufficient to capture the conditional dependencies.

When comparing among different specifications of ARMA-GARCH-type models in order to select the optimal model, I used the statistical measures of model information criteria, such as the Akaike Information Criterion (AIC) proposed by Akaike (1974), the Bayesian Information Criterion (BIC) proposed by Schwarz (1997), the Hannan-Quinn Information Criterion (HQIC) proposed by Hannan and Quinn (1979). The AIC, BIC and HQIC can be computed respectively as:

$$AIC = -2 \ln(L) + 2k, \quad (22)$$

$$BIC = -2 \ln(L) + \ln(n)k, \quad (23)$$

$$HQIC = 2 \ln[\ln[n]]k - 2 \ln(L), \quad (24)$$

where n is the number of observations, k is the number of estimated parameters and shows the degree to which the number of model parameters is penalized and L is the value of the likelihood function evaluated at the parameter estimates. When comparing among ARMA-GARCH-type models, the aim is to select the model with the smallest value information criteria for each specification however, BIC and HQIC are stricter in penalizing loss of degree of freedom than AIC since the higher the number of parameters k the more the model is penalized.

Tables 4 and 5 present the in-sample estimation results for the different volatility models assuming for the innovations the standard normal distribution and the Student's t-distribution, respectively.

First, the results reported in Tables 5 and 6 show that for the ARMA-GARCH, ARMA-EGARCH, ARMA-GJR-GARCH, ARMA-TGARCH, ARMA-NGARCH and ARMA-APARCH models, β is close to 1 which indicates a high degree of volatility persistence ($\alpha + \beta$) which is also an greater indication of the presence of long memory in volatility.

Second, the asymmetric leverage coefficients γ reported in Table 4 and 5 are all significantly different from 0 in the ARMA-EGARCH, ARMA-TGARCH and ARMA-APARCH models while GJR-GARCH leverage coefficient γ is only different from 0 in Table 6 which indicates that WTI market also exhibits a leverage effect.

Third, the values of the power coefficient δ shown in Tables 4 and 5 for the ARMA-NGARCH and ARMA-APARCH models feature a relatively high and statistically significant leverage as both models present positive coefficients which emphasizes that negative returns have a bigger impact on the conditional variance than positive returns.

Fourth, Tables 4 and 5 report that the estimated sum $\alpha + \beta$ assumes the lowest value for the ARMA-CGARCH model which indicates that the short-run volatility component is weaker while the long-run volatility component (ρ) is estimated as 0.9986–0.9935, indicating that the permanent component of the conditional variance evidences a high degree of persistence.

Finally, the ARMA(3,3)-CGARCH(1,1) model performs best for the standard normal distribution and the ARMA(3,3)-EGARCH(1,1) model performs best for the Student's t-distribution regarding the LL, AIC, BIC and HQIC criteria as shown in Tables 4 and 5.

Table 4: Estimation results of volatility models for standard normal distribution

	ARMA (3,3) GARCH (1,1)	ARMA (3,3) EGARCH (1,1)	ARMA (3,3) GJR-GARCH (1,1)	ARMA (3,3) CGARCH (1,1)	ARMA (3,3) TGARCH (1,1)	ARMA (3,3) NGARCH (1,1)	ARMA (3,3) APARCH (1,1)
μ	0.0304 (0.0216)	0.0285 (0.0232)	0.0306 (0.0219)	0.0396 (0.0217)	0.0259 (0.0191)	0.02794 (0.0338)	-0.0504 (0.007)
ϕ_1	-0.1207 (0.0192)	0.052 (0.0187)	-0.1207 (0.0194)	0.0458 (0.0108)	-0.1145 (0.0048)	-0.1141 (0.0268)	0.1269 (0.0000)
ϕ_2	-0.1526 (0.0265)	-0.3549 (0.0066)	-0.1525 (0.0262)	-0.3557 (0.013)	-0.2129 (0.0082)	-0.1746 (0.0895)	-0.0204 (0.0012)
ϕ_3	0.7667 (0.0097)	0.7995 (0.0139)	0.7668 (0.0097)	0.793 (0.0085)	0.7128 (0.0157)	0.7339 (0.0911)	0.8929 (0.0004)
θ_1	0.1025 (0.0164)	-0.0754 (0.009)	0.1025 (0.0165)	-0.0647 (0.0085)	0.0948 (0.0043)	0.0964 (0.0267)	-0.1096 (0.0002)
θ_2	0.1276 (0.0219)	0.3266 (0.0092)	0.1275 (0.0219)	0.3297 (0.0101)	0.1772 (0.0051)	0.1457 (0.0826)	0.0129 (0.0001)
θ_3	-0.8062 (0.0047)	-0.8277 (0.0000)	-0.8063 (0.0047)	-0.8184 (0.0000)	-0.7574 (0.0151)	-0.7755 (0.0872)	-0.9057 (0.0057)
ω	0.0569 (0.019)	0.0259 (0.0151)	0.057 (0.0191)	0.0319 (0.0068)	0.0228 (0.0119)	0.0335 (0.0116)	0.0305 (0.06)
α	0.0925 (0.0164)	-0.0074 (0.0119)	0.0927 (0.0222)	0.0859 (0.0203)	0.1019 (0.0213)	0.1039 (0.0159)	0.1009 (0.0554)
β	0.9035 (0.0158)	0.9889 (0.0075)	0.9035 (0.0161)	0.6560 (0.0669)	0.9144 (0.0193)	0.9088 (0.0137)	0.9123 (0.004)
δ						1.3998 (0.1389)	1.3333 (0.4353)
η				0.0674 (0.0075)			
ρ				0.9986 (0.0009)			
γ		0.192 (0.1327)	-0.0003 (0.0188)		0.0621 (0.0706)		0.0183 (0.0853)
$\alpha + \beta$	0.996 (0.0322)	0.9815 (0.0194)	0.9962 (0.0383)	0.7419 (0.0872)	1.0163 (0.0406)	1.0127 (0.0296)	1.0129 (0.0874)
LL	-15476	-15465	-15476	-15450	-15470	-15464	-15466
AIC	4.3850	4.3822	4.3853	4.3784	4.3837	4.3821	4.3830
BIC	4.3947	4.3929	4.3960	4.3901	4.3943	4.3938	4.3956
HQIC	4.3883	4.3859	4.3890	4.3824	4.3873	4.3862	4.3873

Notes: Robust standard errors are in parenthesis.

Table 5: Estimation results of volatility models for Student's t-distribution

	ARMA (3,3) GARCH (1,1)	ARMA (3,3) EGARCH (1,1)	ARMA (3,3) GJR-GARCH (1,1)	ARMA (3,3) CGARCH (1,1)	ARMA (3,3) TGARCH (1,1)	ARMA (3,3) NGARCH (1,1)	ARMA (3,3) APARCH (1,1)
μ	0.0569 (0.0214)	0.0463 (0.0222)	0.0528 (0.0220)	0.0602 (0.0217)	0.0478 (0.0429)	0.0274 (0.0352)	0.0195 (0.0389)
ϕ_1	-0.0062* (0.0076)*	1.5319 (0.0006)	-0.0168 (0.0077)	-0.2379* (0.1723)*	-0.4038 (0.0355)	-0.2659 (0.0823)	-0.4811 (0.0100)
ϕ_2	-0.3976 (0.0127)	-0.2813 (0.0022)	-0.4056 (0.0126)	-0.2811* (0.1734)*	-0.4492 (0.0088)	-0.3181 (0.0408)	-0.5265 (0.0081)
ϕ_3	0.7377 (0.0076)	-0.3679 (0.0019)	0.7275 (0.0077)	0.6314 (0.1966)	0.4515 (0.0117)	0.5838 (0.0196)	0.3731 (0.0384)
θ_1	-0.0025* (0.0035)*	-1.547 (0.0008)	0.0082 (0.0031)	0.2297* (0.1665)*	0.3922 (0.0312)	0.2562 (0.0785)	0.4696 (0.0096)
θ_2	0.3783 (0.0113)	0.3077 (0.0000)	0.3865 (0.0111)	0.2675* (0.1672)*	0.4310 (0.0076)	0.3011 (0.0404)	0.5086 (0.0079)
θ_3	-0.7545 (0.0000)	0.3523 (0.0000)	-0.7441 (0.0000)	-0.6609 (0.1914)	-0.4853 (0.0115)	-0.6163 (0.0187)	-0.4059 (0.0357)
ω	0.0593 (0.0202)	0.0152 (0.0014)	0.0588 (0.0206)	0.047 (0.0072)	0.0233 (0.1616)	0.0288 (0.0069)	0.0269 (0.0072)
α	0.0679 (0.0151)	-0.0229 (0.0087)	0.0573 (0.0159)	0.0493 (0.0181)	0.0763 (0.2879)	0.0799 (0.0087)	0.0761 (0.0094)
β	0.9228 (0.0169)	0.9896 (0.0001)	0.9238 (0.0171)	0.6586 (0.0737)	0.9322 (0.2874)	0.9280 (0.0078)	0.9311 (0.0089)
δ						1.2042 (0.1163)	1.1348 (0.1149)
η				0.0603 (0.0122)			
ρ				0.9935 (0.0026)			
γ		0.1424 (0.0054)	0.0183 (0.0136)		0.1756 (0.6189)		0.1574 (0.0651)
$\alpha + \beta$	0.9907 (0.032)	0.9667 (0.0089)	0.9811 (0.033)	0.7079 (0.0918)	1.0085 (0.5753)	1.0079 (0.0165)	1.0072 (0.0183)
LL	-15246	-15222	-15245	-15243	-15229	-15232	-15227
AIC	4.3204	4.3137	4.3202	4.3200	4.3156	4.3169	4.3158
BIC	4.3311	4.3253	4.3318	4.3326	4.3273	4.3295	4.3294
HQIC	4.3241	4.3177	4.3242	4.3244	4.3197	4.3212	4.3205

Notes: Robust standard errors are in parenthesis. * denote non-significance at the 5% level.

5.3. FORECASTING CRUDE OIL VOLATILITY

Although the ARMA-CGARCH model and the ARMA-EGARCH model are a good fit for in-sample data, this does not necessarily imply that these models can accurately forecast volatility for the out-of-sample data.

Thus, this section employs the MCS procedure proposed by Hansen et al. (2011) to evaluate the performance of volatility forecasts computed by the different ARMA-GARCH-type models estimated in section 5.2 with the standard normal (N) and standardized Student's-t (STD) distributions for the out-of-sample period from January 2nd, 2014 to December 31st, 2019 which consists of 1505 observations.

The MCS sequence of tests is firstly performed to reduce the initial number of models while at the same time generates a model confidence set which includes the superior set of models (SSM) with Equal Predictive Ability (EPA) with the pre-specified confidence level (α) of 10% for this study which set a confidence level of 90% for MCS ($\hat{M}_{0.9}^*$) procedure.

This flexibility helps to discriminate models with respect to desired characteristics which for this study, are their forecasting performances for 1, 5, and 20-day-ahead which corresponds to 1 day, 1 week, and 1 month trading periods in business days. The higher the number of eliminated models, the higher the heterogeneity of the competing forecasts.

Tables 6, 7 and 8 presents the MCS superior set of models (SSM) ranked according to their p-values for 1, 5, and 20-day forecasts horizons in terms of four different loss criteria that consequently serve different practical uses: the AE_1 and AE_2 denotes the absolute error and are known for devaluing the outliers, SE_1 denotes the squared error and measures the difference between the estimator and what is estimated according to the expected value of the squared error loss, therefore, is characterized by heavily weighting outliers and finally, the QLIKE denotes the Quasi-Likelihood and corresponds to the percentage absolute errors or the loss implied by a Gaussian likelihood, hence is characterized by allowing overdispersion.

Table 6: MCS superior models for 1-day-ahead volatility forecasting

AE ₁		AE ₂		QLIKE		SE ₁	
EGARCH(STD)	1.000	TGARCH(N)	1.000	NGARCH(STD)	1.000	APARCH(N)	1.000
APARCH(N)	1.000	GJR-GARCH(N)	0.999	GJR-GARCH(STD)	0.999	APARCH(STD)	0.997
NGARCH(STD)	1.000	APARCH(N)	0.793	APARCH(STD)	0.998	NGARCH(STD)	0.991
GJR-GARCH(STD)	1.000	GARCH(N)	0.625	GARCH(STD)	0.998	GJR-GARCH(STD)	0.991
APARCH(STD)	0.996	NGARCH(N)	0.309	APARCH(N)	0.996	TGARCH(STD)	0.991
TGARCH(STD)	0.994			EGARCH(N)	0.990	GJR-GARCH(N)	0.972
CGARCH(STD)	0.994			TGARCH(N)	0.989	GARCH(STD)	0.972
GARCH(STD)	0.957			CGARCH(N)	0.989	CGARCH(STD)	0.969
GJR-GARCH(N)	0.747			GARCH(N)	0.912	TGARCH(N)	0.917
CGARCH(N)	0.747			NGARCH(N)	0.887	EGARCH(STD)	0.916
GARCH(N)	0.663			EGARCH(STD)	0.803	EGARCH(N)	0.869
NGARCH(N)	0.641			GJR-GARCH(N)	0.719	NGARCH(N)	0.803
EGARCH(N)	0.444			CGARCH(STD)	0.529	CGARCH(N)	0.803
TGARCH(N)	0.246			TGARCH(STD)	0.509	GARCH(N)	0.744

Notes: The table shows the corresponding p-values computed according to TR statistics under four loss functions. The confidence level for MCS is 90%.

Table 7: MCS superior models for 5-day-ahead volatility forecasting

AE ₁		AE ₂		QLIKE		SE ₁	
EGARCH(STD)	1.000	EGARCH(N)	1.000	EGARCH(N)	1.000	APARCH(N)	1.000
EGARCH(N)	0.999			NGARCH(N)	1.000	EGARCH(STD)	1.000
NGARCH(STD)	0.999			GJR-GARCH(N)	0.999	EGARCH(N)	1.000
GJR-GARCH(STD)	0.999			GJR-GARCH(STD)	0.999	APARCH(STD)	0.991
APARCH(STD)	0.997			APARCH(STD)	0.997	NGARCH(STD)	0.981
GARCH(STD)	0.997			APARCH(N)	0.997	GJR-GARCH(STD)	0.979
CGARCH(STD)	0.996			NGARCH(STD)	0.994	TGARCH(STD)	0.974
TGARCH(STD)	0.994			TGARCH(N)	0.992	TGARCH(N)	0.970
APARCH(N)	0.993			TGARCH(STD)	0.984	GARCH(STD)	0.969
CGARCH(N)	0.982			GARCH(N)	0.917	GJR-GARCH(N)	0.955
GARCH(N)	0.175			EGARCH(STD)	0.899	CGARCH(STD)	0.942
GJR-GARCH(N)	0.165			GARCH(STD)	0.893	NGARCH(N)	0.907
NGARCH(N)	0.146			CGARCH(N)	0.854	GARCH(N)	0.876
				CGARCH(STD)	0.807	CGARCH(N)	0.759

Notes: The table shows the corresponding p-values computed according to TR statistics under four loss functions. The confidence level for MCS is 90%.

Table 8: MCS superior models for 20-day-ahead volatility forecasting

AE ₁		AE ₂		QLIKE		SE ₁	
GJR-GARCH(STD)	1.000	EGARCH(N)	1.000	CGARCH(STD)	1.000	APARCH(N)	1.000
GARCH(STD)	1.000	GARCH(N)	0.864	EGARCH(N)	1.000	NGARCH(N)	1.000
NGARCH(STD)	1.000	CGARCH(N)	0.188	NGARCH(N)	0.999	GARCH(N)	1.000
APARCH(STD)	1.000			GJR-GARCH(STD)	0.998	EGARCH(N)	0.999
EGARCH(STD)	1.000			GJR-GARCH(N)	0.995	GARCH(STD)	0.999
TGARCH(STD)	1.000			NGARCH(STD)	0.992	GJR-GARCH(STD)	0.999
EGARCH(N)	0.999			APARCH(N)	0.992	EGARCH(STD)	0.995
CGARCH(STD)	0.999			CGARCH(N)	0.992	APARCH(STD)	0.987
APARCH(N)	0.999			APARCH(STD)	0.992	CGARCH(N)	0.987
GARCH(N)	0.985			GARCH(N)	0.986	NGARCH(STD)	0.973
CGARCH(N)	0.963			EGARCH(STD)	0.956	TGARCH(STD)	0.965
NGARCH(N)	0.896			TGARCH(STD)	0.935	GJR-GARCH(N)	0.951
GJR-GARCH(N)	0.695			TGARCH(N)	0.718	CGARCH(STD)	0.912
TGARCH(N)	0.145			GARCH(STD)	0.633	TGARCH(N)	0.611

Notes: The table shows the corresponding p-values computed according to TR statistics under four loss functions. The confidence level for MCS is 90%.

The estimated SSM for all forecast's horizons considered the 14 estimated models which suggests that WTI may not be affected by some stylized facts such as the leverage effect or complex nonlinear conditional volatility dynamics. The homogeneous composition of the final SSM, report an empirical evidence that this series appear to be accurately described by the standard normal and the Student's t distributions of the single-regime ARMA-GARCH-type models.

The four loss functions also treat outliers very differently which result in some dissimilarities, however, they agree on which models should be preferred except for AE₂.

In general, the results of the ARMA-EGARCH, ARMA-GJR-GARCH, ARMA-NGARCH and ARMA-APARCH models perform better in smaller forecasting horizons such as the one and five-day-ahead, while the ARMA-EGARCH, ARMA-GJR-GARCH, ARMA-NGARCH and ARMA-GARCH models perform better in forecasts for the 20-day-ahead.

Specifically, for the 1-day-ahead forecasting results, the ARMA-CGARCH and ARMA-EGARCH models are excluded for AE_2 loss function and ARMA-TGARCH models performs the worst under loss functions AE_1 and QLIKE while the ARMA-GJR-GARCH and ARMA-APARCH models performs the best for 3 loss functions, and ARMA-NGARCH performs the best for 2 loss functions indicating a higher forecasting accuracy.

For the 5-day-ahead forecasting results, the ARMA-EGARCH model performs the best since it survives in the MCS under all the loss functions, while ARMA-TGARCH(N) is excluded from the MCS under two of four loss functions.

Regarding the 20-day-ahead forecasting results, the ARMA-TGARCH model is included in the MCS under three of four loss functions, but its performance is worse than ARMA-GARCH and ARMA-CGARCH models that were included in all loss functions. Meanwhile, the ARMA-EGARCH, ARMA-GJR-GARCH, ARMA-NGARCH and for the first time ARMA-GARCH models performs better in three loss functions.

In short, based on the analysis above is possible to generally conclude that none of the ARMA-GARCH-type models outperforms the others in all forecasting horizons. However, the nonlinear models do perform better than the linear one since they can capture long-memory or volatility persistence and asymmetry in WTI return series.

In addition, on the forecasts for 20-day-ahead the linear ARMA-GARCH models performs the best for 2 loss functions, indicating that for longest forecast horizons the incorporation of nonlinear models fails to provide an evident improvement in forecasting accuracy for crude oil price volatility.

6. CONCLUSIONS

The purpose of this dissertation is to forecast WTI volatility using extended daily data from 2nd January 1986 to 31st December 2019 which consists of 8569 observations and estimate seven single-regime ARMA-GARCH models in an attempt to find the best volatility model with the ability to forecast and identify volatility stylized facts, such as volatility persistence or long-memory and the asymmetric leverage effect in WTI return series.

In the in-sample estimation, the different ARMA-GARCH models were compared based on the AIC, BIC and HQIC criteria to select the best model where a smaller criterion indicates a better goodness of fit. The ARMA-CGARCH and ARMA-EGARCH models displayed superiority for the standard normal and Student's *t* distributions, respectively. However, when confronted with the challenge of volatility forecasting in an out-of-sample experiments, only the ARMA-EGARCH models performed significantly better.

The criteria for forecast evaluation are performed using the MCS procedure proposed by Hansen et al. (2011) under four loss functions: AE₁, AE₂, QLIKE and SE₁. Unfortunately, since these loss functions cannot distinguish whether the loss differences are statistically significant. The MCS procedure evaluates the out-of-sample forecasting performance and provides a statistical comparison for the different estimated volatility models and allows the possibility of more than one best model.

The MCS procedure considered the 14 estimated models for all forecast's horizons which is a strong evidence that ARMA-GARCH-type models are able to identify structural changes in volatility.

Overall, in line with recent literature, the empirical results reveal that models that capture asymmetry such as ARMA-EGARCH, ARMA-GJR-GARCH and ARMA-APARCH, as well as models that capture volatility persistence or long-memory like ARMA-NGARCH perform better when forecasting WTI daily prices volatility, according to MCS procedure results with a confidence level of 90%.

However, for longer forecasts horizons such as 20-day-ahead, the MCS procedure indicates that the nonlinear models do not perform significantly better than the linear ARMA-GARCH model which suggests that energy economists, energy policymakers and energy practitioners should prefer the more parsimonious models for crude oil volatility forecasting.

Possible extensions of this study should be explored in the future and can lead to interesting results. First the use of regime-switching models to estimate time-varying variance results and thus identify the structural changes in volatility, such as the Mixture Memory GARCH (MM-GARCH). Second, considering more loss functions and combine with other conditional distributions that can capture other stylized facts observed in the financial return series.

Finally, despite all the research and models that have been proposed the empirical literature is far from an agreement on the appropriate model for forecasting crude oil price volatility. However, it should be noted that there is still some interesting research to be explored in the future, as the vital but complex role of crude oil prices, in the macro and microeconomics will continue to be present in the studies of the coming years, especially after the pandemic crisis of COVID-19 that started in 2020 and still persists.

7. REFERENCES

- Akaike, H. (1974) "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, 19(6), pp. 716–723. doi: 10.1109/TAC.1974.1100705.
- Andersen, T. G. and Bollerslev, T. (1998) "Answering the skeptics: yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39(4), p. 885-905. doi: 10.2307/2527343.
- Baillie, R. T. (1996) "Long memory processes and fractional integration in econometrics," *Journal of Econometrics*, 73(1), pp. 5–59. doi: 10.1016/0304-4076(95)01732-1.
- Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996) "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74(1), pp. 3–30. doi: 10.1016/S0304-4076(95)01749-6.
- Bhowmik, R. and Wang, S. (2020) "Stock market volatility and return analysis: a systematic literature review," *Entropy*, 22(5), p. 522. doi: <https://doi.org/10.3390/e22050522>.
- Bollerslev, T. (1986) "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31(3), pp. 307–327. doi: 10.1016/0304-4076(86)90063-1.
- Bollerslev, T. and Ole Mikkelsen, H. (1996) "Modeling and pricing long memory in stock market volatility," *Journal of Econometrics*, 73(1), pp. 151–184. doi: 10.1016/0304-4076(95)01736-4.
- Catania, L. and Bernardi, M. (2015) "MCS," R package version 0.1.3. <https://cran.r-project.org/web/packages/MCS/index.html>.
- Ding, Z. and Granger, C. W. J. (1996) "Modeling volatility persistence of speculative returns: A new approach," *Journal of Econometrics*, 73(1), pp.186-215 doi: 10.1016/0304-4076(95)01737-2.
- Engle, R. F. (1982) "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica*, 50(4), p. 987. doi: 10.2307/1912773.

Engle, R. F. and Bollerslev, T. (1986) "Modelling the persistence of conditional variances," *Econometric Reviews*, 5(1), pp. 1–50. doi: 10.1080/07474938608800095.

Engle, R. F. and Lee, G. G. J. (1993) "A permanent and transitory component model of stock return volatility," *University of California at San Diego, Economics Working Paper Series*, pp. 475–497.

Ghalanos, A. (2016) "Rugarch: univariate GARCH models," R package version 1.4-4. <https://cran.r-project.org/package=rugarch>.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993) "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The Journal of Finance*, 48(5), pp. 1779–1801. doi: 10.2307/2329067.

Granger, C. W. J. and Ding, Z. (1996) "Varieties of long memory models," *Journal of Econometrics*, 73(1), pp. 61–77. doi: 10.1016/0304-4076(95)01733-X.

Haas, M. (2004) "A new approach to markov-switching GARCH models," *Journal of Financial Econometrics*, 2(4), pp. 493–530. doi: 10.1093/jjfinec/nbh020.

Hamilton, J. D. (1988) "Rational-expectations econometric analysis of changes in regime," *Journal of Economic Dynamics and Control*, 12(2–3), pp. 385–423. doi: 10.1016/0165-1889(88)90047-4.

Hamilton, J. D. (1989) "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57(2), pp. 357–384. doi: 10.2307/1912559.

Hannan, E. J. and Quinn, B. G. (1979) "The determination of the order of an autoregression," *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2), pp. 190–195. doi: 10.1111/j.2517-6161.1979.tb01072.x.

Hansen, P. R. (2005) "A test for superior predictive ability," *Journal of Business & Economic Statistics*, 23(4), pp. 365–380. doi: 10.1198/073500105000000063.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011) "The model confidence set," *Econometrica*, 79(2), pp. 453–497. doi: 10.2139/ssrn.522382.

Hasanov, A. S., Shaiban, M. S. and Al-Freedi, A. (2020) "Forecasting volatility in the petroleum futures markets: A re-examination and extension," *Energy Economics*, 86, p. 104626. doi: 10.1016/j.eneco.2019.104626.

Higgins, M. L. and Bera, A. K. (1992) "A class of nonlinear ARCH models," *International Economic Review*, 33(1), pp. 137–158. doi: 10.2307/2526988.

Kang, S. H., Kang, S.-M. and Yoon, S.-M. (2009) "Forecasting volatility of crude oil markets," *Energy Economics*, 31(1), pp. 119–125. doi: 10.1016/j.eneco.2008.09.006.

Li, M., Li, W. K. and Li, G. (2013) "On mixture memory GARCH models," *Journal of Time Series Analysis*, 34(6), pp. 606–624. doi: 10.1111/jtsa.12037.

Ljung, G. M. and Box, G. E. P. (1978) "On a measure of lack of fit in time series models," *Biometrika*, 65(2), pp. 297–303. doi: 10.1093/biomet/65.2.297.

Nelson, D. B. (1991) "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica*, 59(2), p. 347–370. doi: 10.2307/2938260.

Palm, F. C. (1996) "7 GARCH models of volatility," in *Handbook of Statistics*, 14, pp. 209–240. doi: 10.1016/S0169-7161(96)14009-8.

Poon, S.-H. and Granger, C. W. J. (2003) "Forecasting volatility in financial markets: a review," *Journal of Economic Literature*, 41(2), pp. 478–539. doi: 10.1257/jel.41.2.478.

Sadorsky, P. (2006) "Modeling and forecasting petroleum futures volatility," *Energy Economics*, 28(4), pp. 467–488. doi: 10.1016/j.eneco.2006.04.005.

Schwarz, E. S. (1997) "The stochastic behavior of commodity prices: implications for valuation and hedging," *The Journal of Finance*, 52(3), pp. 923–973. doi: 10.1111/j.1540-6261.1997.tb02721.x.

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. doi: 10.1016/j.eneco.2010.07.009.

Zakoian, J. M. (1994) "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18(5), pp. 931–955. doi: 10.1016/0165-1889(94)90039-6.

Zhang, Y.-J. *et al.* (2019) "Volatility forecasting of crude oil market: can the regime switching GARCH model beat the single-regime GARCH models?," *International Review of Economics & Finance*, 59, pp. 302–317. doi: 10.1016/j.iref.2018.09.006.

Zhang, Y.-J. and Wang, J. (2015) "Exploring the WTI crude oil price bubble process using the markov regime switching model," *Physica A: Statistical Mechanics and its Applications*, 421, pp. 377–387. doi: 10.1016/j.physa.2014.11.051.