Universitetet
i Stavanger

Faculty of Science and Technology

Master's Thesis

Study programme/specialisation: Industrial Economics/*Risk Management* and *Investment and Finance*	**Spring semester, 2020** Open
Author: Kristoffer Ree	.... (Signature of author)
Faculty supervisor: Sindre Lorentzen	
Title of master's thesis: Ex-Post (Pseudo) Out-of-Sample Forecast Accuracy of Proposed Oil Price Models	
Credits (ECTS): 30	
Keywords: Ex-Post Out-of-Sample Forecast Accuracy Crude Oil Price Models Time Series Loss Functions	**Number of pages:** 64 **+ Supplemental material/other:** 9 Stavanger, **June 28, 2020**

(This page is intentionally left blank)

Master Thesis
INDMAS

Ex-Post (Pseudo) Out-of-Sample Forecast Accuracy of Proposed Oil Price Models

Universitetet i Stavanger

Kristoffer Ree

University of Stavanger

June 28, 2020

Abstract

This thesis aims to test and compare some of the most frequently applied models in the forecasting literature, for their ability to produce accurate ex-post (pseudo) out-of-sample forecasts of the crude oil price. These models range vastly in complexity, ranging from the most parsimonious idea of price-today-is-price-tomorrow approach to more sophisticated and stochastic models. All models will be assessed with the commonly used proxy for the oil price, namely the West Texas Intermediate (WTI) benchmark price, sampled in both daily and monthly frequencies. A model's forecast accuracy will be evaluated employing a set of various loss functions that differ in their way of penalizing the forecast errors. Additionally, the models' forecasts will be tested for being directionally accurate in predicting the actual price changes. Finally, model selection and estimation will be analysed across different lengths of historical price data, to examine what effect the choice of sample period has on the forecast results.

The empirical results of this analysis show that neither the deterministic or stochastic models evaluated are able to forecast the price of crude oil with an adequately desired accuracy. It was also found that forecast results are highly sensitive to the choice of sample period for historical prices used as input for model estimation, and that certain models perform better when only recent market data is used as input.

Acknowledgement

This thesis marks the finalization of my Master of Science program in Industrial Economics at the University of Stavanger (UiS) with specialization within *Risk Management* and *Investment and Finance*. Throughout my five years at UiS, I have gained knowledge in various fields which will be of great value when I pursue my future career ambitions. I am grateful for the consistent service and teaching I have been given from the lecturers and other employees at the university.

I especially want to thank Sindre Lorentzen, my academic supervisor, for his guidance and motivational support during my work on this Master's thesis. He has also provided me with an interesting and challenging Master's thesis. His professional expertise within the field of forecasting and economics have been essential for me to learn the theoretical basis of this thesis.

I would like to express gratitude to my family and close friends during the time of my studies. They have been most helpful and encouraged me whenever needed. Finally, I want to thank my fellow students. You have inspired me and pushed me in the right directions regarding studying, as well as given me many good memories and laughs.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC	*Akaike Information Criteria*
AR	*AutoRegressive*
ARIMA	*AutoRegressive Integrated Moving Average*
DM	*Diebold-Mariano*
GBM	*Geometric Brownian Motion*
IEA	*International Energy Agency*
L	*Long*
M	*Medium*
MA	*Moving Average*
ME	*Mean Error*
MR	*Mean Reverting*
MSE	*Mean Square Error*
NYMEX	*New York Mercantile Exchange*
OPEC	*Organization of the Petroleum Exporting Countries*
PT	*Pesaran-Timmermann*
R	*Recent*
RMSE	*Root Mean Square Error*
S	*Short*
SDE	*Stochastic Differential Equation*
SR	*Success Rate*
WTI	*West Texas Intermediate*

1 Introduction

Crude oil is arguably one of the most important commodities in the world, accounting for one-third of global energy consumption. In addition to being a starting material for most of the products we use in everyday life, it has also emerged as being a highly important strategic commodity in terms of a nation's economical- and political strength. Abrupt movements in crude oil prices have been proven to affect the level of economic activity and consumer sentiment (Hamilton, 2009). Despite the paramount importance of fluctuations in crude oil prices for economic activity, forecasting and developing a better understanding of the price of oil continues to be a daunting task. Reliable forecasts of crude oil prices are of special interest for a wide range of applications. Central banks and international organizations view the price of oil as one of the critical variables in generating macroeconomic projections and for assessing and managing macroeconomic risks as well as to improve policy responses.

The price of oil is basically determined by its supply and demand. Although there is a common consensus of it being even more influenced by exogenous and irregular past, present and future events like weather, gross domestic product growth, stock levels, political aspects, quota decisions set by the largest producing capacity cartel: the Organization of the Petroleum Exporting Countries (OPEC), turmoil in oil-exporting and -importing countries and so on. In the research, there has been numerous attempts and suggested model specifications for forecasting oil prices. Regardless of these studies, a better understanding of crude oil prices and their formations are still highly sought after knowledge, which at the current state is not adequately sufficient in terms of accurately forecasting the price fluctuations. Accurate predictions of oil prices are highly required, for example, to guide natural resource development and investments in infrastructure. They also play an integral part in generating projections of energy use, predicting carbon emissions and climate change, and in particular to companies dependent on energy prices for optimal investment decisions, allocation of capital and risk management. Hence, identifying the stochastic processes governing the price of oil is essential for both policymakers and private energy actors.

This thesis aims to evaluate the ex-post (pseudo) out-of-sample forecast accuracy of some of the proposed models in the literature in terms of the crude oil price. As noted by Hansen and Timmermann (2012) the choice of sample period and sample split point considering forecasting analysis is not irrelevant. For this reason, different lengths of historical oil prices were applied to the models considered, with the logical supporting idea being that using more recent information of the oil market dynamics may improve upon a model's ability to forecast the future. Whereas by including information about past dynamics not representative for the current and future state of the market, may jeopardize a model's forecasting ability. This logic is supported by academics and practitioners who argue that crude oil markets have undergone structural transformations which have changed the impact of underlying factors and alternated the path of oil prices. Hence, the hypothesis that by using more recent oil price data performs better than including past and not relevant data when conducting a forecast will be evaluated.

For any forecast, there is a requirement for some loss or cost function to evaluate its performance when compared to the true observed values. In the forecasting literature, this has usually been done in terms of calculating the forecast's Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) or similar widely adopted loss functions. As different loss functions penalize under- and overprediction errors in different ways, a model's out-of-sample forecasting accuracy is thus dependent upon the choice of loss function. Hence, in this thesis a model's forecasting accuracy will be evaluated against various loss functions, to see what effects the choice of loss function has on the model's ranking and performance. Although a model may provide worse forecast results compared to another model in terms of any loss function metric, it might be able to capture other important aspects of the price path. This may, for example, be the direction of price changes, which is pivotal to firms and investors in the financial oil market. Thus a model's ability to predict the direction of future price changes will also be evaluated by computing the model's Success Rate (SR). To check whether the SR can be interpreted as evidence of directional accuracy or pure luck, a suitable significance test of the score will be reported.

The ultimate objective for economist and science in general is to make reliable and consistent forecasts in order to grant better decision-making e.g. for firms and stakeholders. Despite the frequent prevalence and academic interest in forecasting the crude oil price, the literature has proven itself to be lacking a common consensus regarding what model produces the most accurate forecasts. There have been numerous proposed models in the literature, with their complexity varying greatly. Ranging from the most parsimonious idea of price-today-is-price-tomorrow approach to extremely sophisticated deep learning machine models. This raises the question of whether more complex models perform better than the less computational expensive models? And does the benefits of more complex models outweigh the extra cost of applying them? Therefore in this thesis, some of the most widely applied models in the forecasting research, with varying complexity, will be evaluated and tested with various sample lengths and forecast horizons. Finally, analysts in the literature have usually aimed attention to one price sample frequency when conducting the forecasts. As daily crude oil prices have significantly more noise in the data compared to monthly, quarterly and annual prices, the models will be evaluated using both daily and monthly prices. This is done to be able to spot any differences in forecasting performance when the amount of noise in the data is reduced, and examine whether models performing well with one sample frequency also is consistent when another sampling frequency is utilized.

The remainder of this thesis is structured as follows: Chapter 2 gives a brief introduction to grasp the concept of the oil price and its history. Chapter 3 gives an overview of the existing forecasting techniques in the literature, as well as presents the selected models to be employed throughout the thesis. Chapter 4 includes a description of the price data sets to be studied, and discusses essential aspects of the forecasting procedure and its evaluation. Chapter 5 presents and examines the empirical results of the selected models from Chapter 3. Finally, Chapter 6 concludes on the obtained results.

2 The Oil Price

As this thesis scrutinizes modelling the dynamics of the oil price, this chapter will give a brief introduction to what we mean by the oil price, its history, and its determining factors. When referring to the oil price one generally means the spot price of one barrel[1] of crude oil. Crude oil is an unrefined naturally occurring petroleum product accumulated and buried in reservoirs underneath the surface. This product is composed of hydrocarbons and other organic material, which serves as the world's most dominant source of primary energy. Oil is the most consumed primary energy source worldwide, and in 2018 oil accounted for approximately 31%[2] of the world's total primary energy demand.

This chapter provides a brief theoretical framework relevant to grasp the oil price. In Section 2.1, a historical view of some of the major events that have led to oil price fluctuations, are given attention. Whereas Section 2.2 describes some of the most commonly utilized benchmarks to price one barrel of crude oil in the international market.

2.1 Oil Price History

Fluctuations in the price of a barrel crude oil are influenced by the three primary factors: demand, supply and oil inventories. However, over the last two decades, the behaviour of oil prices have become increasingly more complex with many driving factors, such as politics, government regulations, interest rates, technological advances, environmental concerns, natural disasters, population- and economic growth, etc., influencing its behaviour. Oil prices are highly volatile compared to other commodities as a result of the fact that oil demand and supply have low price elasticities. Supply is almost inelastic in the short-run and can only be increased up to its full capacity. Equivalently, oil demand is also rather inelastic as it has limited substitution potential in the short-run in response to an oil price increase. Depending on the most important underlying driving factors, oil prices can behave very differently over time. A historical view of the spot prices for the two most common benchmarks West Texas Intermediate (WTI) and Brent Crude can be seen in Figure 2.1. From which the evolution of the oil price over time has been varying between being stable, collapsing abruptly and trending upwards.

[1] 1 barrel ≈ 159 litres.
[2] According to IEA's World Energy Outlook 2019 (IEA, 2019).

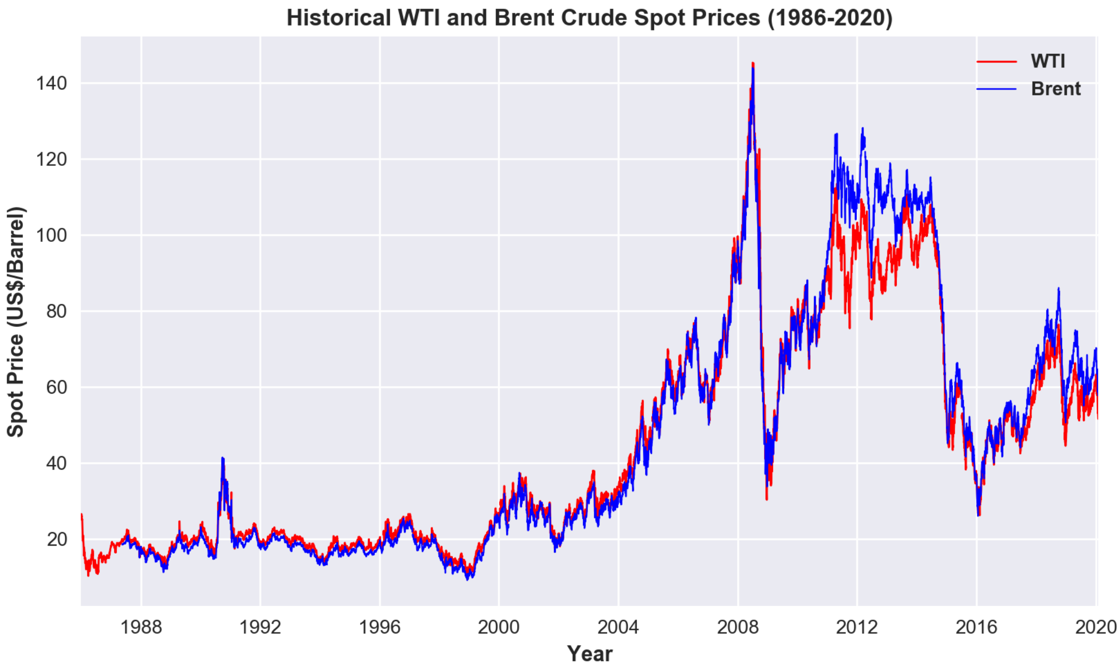


Figure 2.1: *Historical daily WTI and Brent Crude spot prices (1986-2020). Data retrieved from the U.S. Energy Information Administration's website.*

Major global events are strongly related to the observed fluctuations in the oil price. Where some are more predictable and understood, whereas others are utterly random and unpredictable, these latter events are often referred to as "black swans"[3]. One great example of such an event was the terrorist attacks by the Islamic group Al-Qaeda against the United States on the morning of September 11, 2001. After the hit, oil prices shot upward due to fears that oil imports from the Middle East would be curtailed. This event underlined the connection between the oil market and political instability. In August 1990, during the Iraqi invasion of Kuwait, the United Nations put a total embargo on oil-exporting from Iraq and Kuwait. In addition, with the outbreak of the Persian Gulf War, it resulted in a daily oil supply reduction of 4.7 million barrels in the international market, which accounted for 7% of the global aggregate demand. Referring to Figure 2.1, the oil price rose from 14 dollars to 40 dollars per barrel in as little time as three months. However, the oil price was quickly pulled back, due to an increased production in OPEC countries supplementing the supply shortage. After

[3]The term, black swan, was first introduced by Nassib Talib through his book, *The Black Swan*, in 2007 (Taleb, 2007). In this book, a black swan is understood as a surprising extreme event relative to one's knowledge or beliefs following three attributes. Firstly, it is an outlier, as it lies outside the realm of regular expectations because nothing in the past can convincingly point to its possibility. Secondly, it carries an extreme impact. Thirdly, in spite of its outlier status, human nature makes us concoct explanations for its occurrence after the fact, making it explainable and predictable.

1996, with gradually increasing oil production and exporting in Iraq plus the impact of the Asian financial crisis on the world economy and oil demand, the oil price continuously declined to a historically low level of 10 dollars per barrel in 1998. Once again OPEC adjusted its production levels, and after a period of underproduction oil price rebounded back up at the start of the 21th century. At the same time, the black swan event of 9/11 occurred which triggered a new round of increasingly oil prices.

From 2002 onward, strong growth in the global economic activity driven by developing market economies, particularly in China, resulted in a prolonged period of increasing oil prices. This economic resurgence led to an accelerated increase in crude oil demand, which put upward pressure on the oil price, and in the middle of 2008, the price had risen to a record high price of nearly 150 dollars per barrel. However, this price level did not last long, by the end of 2008 due to the outbreak of a global financial crisis, the oil price plummeted to near 30 dollars per barrel, an over 100 dollars decline in just five months. At the beginning of 2009, after the initial shock of the financial crisis and the financial market had retained some stability again, the oil price picked up and rose to about 70 dollars per barrel in a short time. The oil price became broadly stable for the next four years owing to the rise in shale oil production in North America and a diminishing oil demand growth, only counteracted by supply-side concerns related to geopolitical tensions in the Middle East and Russia. Simultaneously, gains in energy efficiency and improved development in other substitutional energy resources contributed to restricting oil demand growth. The oil price fluctuated greatly around the 100 dollars mark these years until the steep fall in the middle of 2014. This was a result of oversupply of oil in the market due to booming shale oil production levels in North America together with stagnant oil demand growth, especially in China.

Regardless of the oversupply and a low oil price, OPEC decided to not reduce its production levels at its meeting in November 2014. The found member of OPEC, Saudi Arabia, has historically operated as a "swing producer" in the oil market. Where it has stabilised the oil price by reducing or increasing its production levels to respectively, either rise the oil price in case of oversupply or lower the oil price in case of shortages. By February 2016, the oil price had dropped to below 30 dollars per barrel as oil-producing countries had since the middle of 2014 produced 1-2 million barrels of crude oil daily exceeding demand, as well as China's economy hit its lowest growth in a generation. Finally, after a long period of hibernation, OPEC and other non-OPEC members, including Russia, tried to agree on a coordinated production freeze to rebalance the oil market. By the end of November 2016, an upswing came as OPEC members agreed to cut their production levels by 1.2 million barrels per day. Following OPEC's restriction on the supply, the global oil inventories declined along with robust demand, the market tightened significantly and the oil price reached levels ranging between 50 to 75 dollars. Most recently, in March 2020 (not included in Figure 2.1), the price went through a massive plunge of 34% in just one day. Induced by the potential fears of a three-way oil price war between OPEC, Russia and the U.S., with the additional effect from the outbreak of the global coronavirus disease (COVID-19), reducing the global demand. This latter proves just how prone the oil price movements are to the current state or even expected future state of the world, which more often than not, are consequences of unforeseen events or news. For any academics or practitioners attempting to predict future oil price movements, this serves as an unfavourable feature making it an undeniable very difficult task to do.

2.2 Crude Oil Characteristics

Crude oil is indisputably one of the world's most important commodities. Although the resource often is referred to as "black gold", it has ranging viscosity and can vary in colour from black to yellow depending on its hydrocarbon composition. Many types of crude oil are produced globally and the market value per barrel heavily depends upon the quality and grade of the crude oil. Because some crude oil types are more preferable than others for refiners to distillate and convert into petroleum products, there exist price differentials among different qualities of crude oil. The two most important characteristics of crude oil are its density and sulfur content. Density ranges from light to heavy, whereas sulfur content is characterised as sweet or sour. Crude oils with low sulfur content and low density are defined as light sweet crude oils and are usually priced higher than heavy sour crude oils. Sulfur is considered an impurity and for sour crude oils containing more than 0,5% sulfur, this impurity needs to be removed before the crude oil can be refined into petrol, and thus the cost of processing is increased compared to processing a sweet crude oil.

For convenient trade and use of crude oil around the world different benchmarks are defined to clarify a crude oil's quality and grade. The three most important and actively traded are the WTI, Brent Crude and Dubai Crude. WTI refers to light sweet crude oil extracted from wells in the U.S. and thereafter sent via pipelines to Cushing, Oklahoma. It is the most actively traded futures contract and serves as the main benchmark in North America. The quality of Brent Crude is very similar to that of the

WTI, however, it is not as light or as sweet as the WTI. Roughly two-thirds of all crude oils around the world are priced using Brent Crude as the benchmark price. Brent Crude refers to crude oil extracted from fields in the North Sea between the Shetlands Islands and Norway. An advantageous feature of the Brent Crude is its waterborne supply, making it easier to transport to distant locations, while the supply of WTI is land-locked and thus transportation costs are generally more onerous. Dubai Crude refers to Middle Eastern heavy and medium sour crude oil with lower quality than WTI and Brent. The benchmark is typically used to price Persian Gulf crude oil exports to Asia. In addition to these primary crude benchmarks, there exist more than 100 crude oil benchmarks. The prices of other crudes are set at a differential to the most utilized benchmarks, where the differentials are adjusted according to changes in supply and demand, transportation costs, as well as quality premiums or discounts. Some of the others most commons are; OPEC Reference Basket (ORB) used by OPEC to standardize crude oil prices among its member countries, Tapis Crude which is traded in Singapore, Bonny Light used in Nigeria, and Urals oil used in Russia.

3 Forecasting Techniques and Selected Models

Over the past recent years, there has been an increasing interest from academics and practitioners on how to understand and accurately forecast the development of oil prices. This interest is subject to the increasing fluctuations observed in the oil prices lately, which makes accurate forecasting more difficult. A vast amount of methodologies exist in the literature on how to forecast oil prices, with no widely accepted consensus on which performs best. These methodology's complexities differ greatly, ranging from the most basic idea stating that current prices are the best predictor of tomorrow's prices, to more extensive models requiring powerful data analytical tools. In addition, some models perform well in the short-term horizon but not in the long-term horizon, and others vice versa. A common approach for companies has been to use models relying on spot and futures prices to make short- and medium-term predictions on the oil market. With the underlying concept that the relationship between futures price fluctuations and spot price fluctuations will point towards future period's oil prices.

This part of the thesis aims to give an overview of the existing crude oil price forecasting techniques and introduce the most widely adopted models in the literature. In Section 3.1 the various techniques applied when forecasting crude oil prices are presented. Selected deterministic and stochastic models used to forecast crude oil prices are described in Sections 3.2 and 3.3, respectively.

3.1 Crude Oil Price Forecasting Techniques

Crude oil is one of the strategic commodities and its use and availability have an impact on the world's macroeconomic factors such as interest rates, inflation, gross domestic product, exchange rates, industrial production, etc. Due to this importance, the determinants of oil price fluctuations have received substantial attention from energy researchers and economists in the literature. As a result, numerous techniques and models have been proposed in the pursuit of a reliable and consistent approach to forecast the fluctuations, volatility and movements of crude oil prices. In the existing literature, we distinguish between two main categories of forecasting techniques: qualitative and quantitative techniques. Quantitative techniques are divided into two categories: econometric models and non-standard methods. These techniques deal with numerical and quantitative factors influencing oil prices. The econometric models are further subdivided into three types: time series models, financial models and structural models. The characteristics of the three types of econometric models to forecast oil prices can be described as follows:

a) time series models exploiting the statistical properties of the data, namely autocorrelation and non-stationarity;

b) financial models based on the relationship between spot and futures prices; and

c) structural models describing how specific economic factors and the behaviour of economic agents affect the future values of oil prices (Frey et al., 2009).

For the non-standard or computational methods, which have recently received a lot of attention regarding forecasting, the most frequently utilised methods are artificial neural networks (ANN) and support vector machines (SVM). On the other hand, qualitative techniques deal with erratic factors' influence on oil prices such as political events, speculations and wars. This is a knowledge-based approach to model the oil price and incorporate infrequent and erratic events which might influence the future oil market. As proposed by Bashiri Behmiri and Pires Manso (2013), all forecasting strategies adopted in the literature can be classified according to the following list of techniques:

1. **Quantitative techniques:**
   - 1.1. Econometric models:
     - 1.1.1. Time series models
     - 1.1.2. Financial models
     - 1.1.3. Structural models
   - 1.2. Non-standard methods: Artificial Neural Networks, Support Vector Machines.
2. **Qualitative techniques:** Fuzzy Logic and Expert Systems, Delphi method, Web Text Mining method, Belief Networks.

Selection of forecasting technique depends greatly on which aspect of the crude oil market one sets out to forecast. In the literature, there is a clear distinction between models used to forecast the volatility of crude oil prices and those used to forecast the actual crude oil prices or associated change in crude oil prices (i.e. logarithmic- or arithmetic returns). For the remaining part of this thesis, the main focus will be on deterministic and stochastic quantitative techniques that forecast the actual crude oil prices or associated returns.

3.2 Selected Deterministic Forecasting Models

This section aims to give a brief theoretical introduction to some of the deterministic models widely applied when forecasting crude oil prices, or even commodities in general and stock prices. A deterministic model is one where the value of the dependent variable of interest is completely determined by the parameters of the model, hence for the same parameter values, the outcome will be the same every time calculated. Quantitative techniques apply mathematical models in an attempt to exploit relevant characteristics of historical data to make short- and medium-term predictions. Among the existing literature on forecasting crude oil prices, econometric models serve as the most frequently used technique and will be the main focus area of this study. Due to the increasing complexity of the oil price dynamics described in the last chapter, there have been numerous attempts in the literature to fit different models and incorporate these factors, but few of these have provided adequately strong forecast results.

3.2.1 Benchmark Model: Random Walk

A random walk without drift forecast, also referred to as the No-Change forecast, has been utilized as the conventional benchmark in the literature on forecasting asset prices (Alquist and Kilian, 2010). Complex and computationally expensive models have typically been compared to this benchmark, to examine whether the introduction of more complexity indeed is justified by an increased forecasting performance relative to the benchmark's performance. According to findings in Alquist and Kilian (2010), the No-Change forecast outperformed other econometric models and was much more accurate than professional survey forecasts of the crude oil price across horizons from 1 month to 12 months. The model forecasts the expected spot price h-period in the future, S_{t+h}, conditional on all available information at present time t, as equal to the last observed spot price, S_t. Hence, the concept is that changes in the spot price are unpredictable, and thus the best available h-period forecast of the oil price is the currently observed oil price, which can be written as:

$$\hat{S}_{t+h|t} = S_t \tag{3.1}$$

Equation 3.1 simply states that the change in the actual oil price at time t and the predicted oil price h-period ahead is zero, and if extrapolated into the distant future it follows a horizontal line equal to the last observed value at present time t. When this concept is fitted to a known time series of oil prices, the model predicts the next oil prices as equal to the previous step's actual price, hence it "shadows" the actual oil price, lagging one time-step behind. Thus, a very obvious problem with this naïve model is when the oil price constantly moves up and down between each period, then the model will predict the opposite movements for each respective period.

An extension of this model is the random walk with drift model, which can be applied if the time series either has an observed increasing or decreasing trend. Throughout the thesis, this model will be referred to as the RW with Drift model. The model expresses the expected h-period ahead oil price by:

$$\hat{S}_{t+h|t} = S_t + h\alpha \tag{3.2}$$

In equation 3.2, α is the estimated drift or the average change from one period to the next of the historical data, which is multiplied by the number of periods ahead, h, one is set out to forecast. Hence, an h-period forecast is modelled as a trend line with slope α anchored at the last observed price, S_t. The random walk application in finance is related to the "efficient market hypothesis", often credited to Fama (1965), stating that in an efficient market with complete information, actual prices reflect all available relevant information and expectations about the future.

3.2.2 Futures Prices as Predictors of Future Spot Prices

A widely applied approach when forecasting the crude oil price is to relate expected future spot prices to futures contract's prices of crude oil. A futures contract is a standardized legally binding agreement to buy or sell a particular commodity, asset, or security at a predetermined price at a specified time in the future. They are traded at futures exchanges, such as the New York Mercantile Exchange (NYMEX) for WTI contracts and the Intercontinental Exchange (ICE) for Brent Crude contracts, which allow traders to transact anonymously. The NYMEX exchange *WTI Light Sweet Crude Oil Futures* is the world's most liquid and actively traded crude oil contract. Each day approximately 1.2 million contracts are traded (as of 2020), where each contract comprises 1,000 barrels, although only a small fraction of the traded volumes are physically settled. Central banks and international institutions, such as the International Monetary Fund (IMF), commonly use futures as a proxy for the market's expectation about the expected future spot price of crude oil. In addition, futures-based forecasts of the oil price play an integral part in policy discussions at the Federal Reserve Board. Despite the widely adopted approach of using futures-based forecasts, this is not to say that forecasters do not recognize the potential limitations of such an approach. However, the consensus among policymakers, macroeconomists and financial analysts is that oil futures, as imperfect as they might be, are still the best available forecasts of the expected future spot price of oil. This perception has persisted regardless of recent empirical evidence to the contrary and the advancement of theoretical models designed at explaining the inadequacy of the predictive power of oil futures prices.

Utilizing futures as a proxy for future spot prices have the great advantage that it is relatively easy to generate and communicate to the public. An h-period forecast of the nominal price of crude oil can be generated by using the price of a futures contract with maturity h, F_t^h. Under the assumption that futures oil prices represent an unbiased and efficient predictor of expected future spot oil prices, implies the forecasting model, and will be referred to as the Futures model:

$$\hat{S}_{t+h|t} = F_t^h, \qquad h = 1,2,3,\dots n \tag{3.3}$$

However, there are several different approaches suggested in the literature on how to utilize futures prices to portray expected future spot prices.[4] Following the suggested approach of Alquist et al. (2011), the forecasted oil price can be based on the spread between the spot price and futures price which act as an indicator of whether the oil price is likely to move up or down. If the futures price equals the expected future spot price, the spread should indicate the expected changes in spot prices. The suggested model to explore the forecasting accuracy can be expressed as:

$$\hat{S}_{t+h|t} = S_t\left[1+\ln\left(\frac{F_t^h}{S_t}\right)\right], \qquad h = 1,2,3,\dots n \tag{3.4}$$

Where S_t is the current spot price of oil, F_t^h is the current oil futures price with maturity h. The logarithmic term of $(\frac{F_t^h}{S_t})$ represents the spread and is commonly assumed by practitioners to equal the expected change in the nominal price of oil over the next h period. This model will be referred to as the Spread model.

The common view that futures prices contain information about future spot prices implicitly relies on the hypothesis that oil futures contracts are actively traded at the

[4]See for example Baumeister et al. (2013) and Alquist and Kilian (2010) who outlines different futures models modifications.

given horizon. This is an important condition as one would not expect F_t^h to have sufficiently strong predictive power for future spot prices if the market is lacking liquidity at the relevant horizon. Typically the liquidity and trading activity for futures contracts declines as their maturities increases. In 1989, the NYMEX introduced futures contracts extending beyond 12 months, and later on in 1999, contracts with maturities of 7-years were first introduced. Even though these long-maturities contracts are available, the futures market for contracts extending 12 months is not particularly liquid. This observation is essential in the way that one should not expect futures with maturities over one year to provide consistently accurate forecasts, due to the low amount of such contracts being traded. In the empirical literature there exist evidence that futures prices, in fact, does not serve as any particularly accurate predictor to forecast the future spot oil price. For example, Alquist and Kilian (2010), Alquist et al. (2011) and Baumeister and Kilian (2012, 2014) concludes that futures-based forecasts does not significantly improve upon the accuracy achieved by a monthly No-Change forecast up to a 12 month horizon. However, some improvements were observed at certain horizons in the statistical loss metric utilized (Mean Square Percentage Error), but generally, none of these reductions turned out to be statistically significant. Whereas, in a study by Reeve and Vigfusson (2011), futures-based oil price forecasts seemed to outperform the No-Change benchmark prior to the early 2000s with respect to the statistical loss metric Mean Square Error (MSE), although in the latter first decade of the 2000s the No-Change forecast seemed to be the superior forecasting method. Regardless of these empirical research results exposing the weaknesses of the futures price forecasting approach, central banks and international organizations persist to employ this method as their baseline forecast.

3.2.3 ARIMA Model

The AutoRegressive Integrated Moving Average (ARIMA) model is a class of statistical models which are widely applied in statistics, econometrics, and in particular for univariate time series analysis and forecasts. This method is often referred to as the Box-Jenkins approach, credited to Box and Jenkins (1976). ARIMA models are simple linear time series models that have extensively been utilized to predict crude oil prices (Alquist et al. (2011); Baumeister and Kilian (2012); Akpanta and Okorie (2014); Dbouk and Jamali (2018)), interest rates changes (Gospodinov and Jamali (2011); Dbouk et al. (2016)) and other financial asset returns. An ARIMA model applies a forecasting algorithm based on the conception that information from previous values in a time series can alone be used to predict future values. This information is retrieved from the time series' own lagged values and the lagged forecast errors. Any specific ARIMA(p,d,q)

model is characterised by its order of the parameters p, d, and q, which respectively describes the AutoRegressive (AR), Integrated (I), and Moving Average (MA) parts of the specific model. Hence, one of the great features of the ARIMA model is the ability to transform it into various other models by carefully setting the parameters to specific orders. The AR part indicates that it is a linear regression model where the evolving variable of interest is regressed on its own lagged values. Hence, a linear relationship to forecast the variable is made from using its past values as predictors. For a linear regression model to have any statistical power in its estimates, the regressors should neither be correlated nor dependent on each other.

When fitting an ARIMA model to any time series, the first step is to determine how to make the time series stationary. A stationary time series is one whose properties do not depend on the time at which they were observed, e.g. a white noise process is stationary. Hence, for a stationary time series we would expect statistical attributes such as the mean, variance, autocorrelation, etc., to remain approximately constant over time (i.e. low heteroscedasticity). There are several techniques to transform a non-stationary time series into a stationary one. Logarithmic transformations aid to stabilise the variance of a time series, whereas differencing is an effective way of stabilising the mean by removing changes in the level of a time series and thereby reducing any trend or seasonality observed. The integrated part of the ARIMA model represents the order of differencing required to make a non-stationary time series, stationary. Usually, a first-order differencing is enough to induce stationarity, this would be the case of setting d equal to 1, while higher orders of differencing are rarely necessary. Considering the spot price of crude oil, a first-order differencing would imply the following relationship:

$$S'_t = S_t - S_{t-1}$$

As the crude oil prices previously observed in Figure 2.1 clearly exhibit periodically trends and a strongly varying variance over time, a first-order difference seems suitable to induce stationarity of the time series. When treating the first-order differenced crude oil prices as the independent variable of interest, a pure autoregressive model of order p (AR(p)) can be modelled as:

$$S'_t = \alpha + \phi_1 S'_{t-1} + \phi_2 S'_{t-2} + ... + \phi_p S'_{t-p} + \varepsilon_t$$

Or equivalently:

$$S_t' = \alpha + \sum_{i=1}^{p} \phi_i S_{t-i}' + \varepsilon_t \tag{3.5}$$

The parameter p implies that the variable of interest (S_t'), is estimated from its previous p-lagged values acting as explanatory variables. The coefficients ϕ_i represents the weights associated with each lagged variable S_{t-i}', α is an optional intercept term, and ε_t is a white noise process with zero mean and zero correlation across time.

Finally, a pure moving average model of order q (MA(q)) can be modelled as:

$$S_t' = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \varepsilon_t \tag{3.6}$$

In an MA(q) model, instead of regressing the variable of interest on its past values, it uses past forecast errors, ϵ_{t-i}, as explanatory variables. It follows the intuition that each value of S_t' can be modelled as a weighted average of the past q forecast errors. In equation 3.6, the coefficients θ_i represents the weights of each lagged forecast error, μ is the expectation for S_t', and finally, ε_t is again white noise.

An ARIMA model incorporates all the features of an AR(p) and a MA(q) model by the appropriate order of differencing of the time series in question. By introducing a new constant term c, being the sum of α and μ, and using the expected value for the white noise term; $E[\varepsilon_t] = 0$, the final outcome is an ARIMA(p, d, q) model, which in our case becomes:

$$S_t' = c + \sum_{i=1}^{p} \phi_i S_{t-i}' + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{3.7}$$

This model is defined as a non-seasonal ARIMA model where the predictors on the right-hand side include both lagged values and lagged errors of the variable of interest, and the intercept term c can either be included or not. As mentioned earlier, by carefully setting the order of the model, various other models are constructed. For example, an ARIMA(0,1,0) without the intercept term is equivalent of a random walk or as previously denoted the no-change model, whereas including the intercept term yields the random walk with drift model.

3.3 Selected Stochastic Forecasting Models

This section lays the fundamentals behind the selected stochastic forecasting models used to obtain the results in the forthcoming chapters. From a financial point of view, such models try to describe the price behaviour and uncertainty of the underlying commodity. A stochastic model is used to forecast the probability of various outcomes under different scenarios, by using random variables. The random variables are built into the model, which produces many outcomes to see their different effect on the solution, then this is repeated numerous times under different scenarios. These models differ from the previous section's selected deterministic models which gives the exact same outcomes for a set of inputs, no matter how many times the model is recalculated. The models described in this section includes the Geometric Brownian Motion model and the Mean Reverting Vašíček model.

3.3.1 Brownian Motion

The phenomenon Brownian motion is such a fundamental characteristic observed in financial modelling that it needs some further explanation. It was the Scottish botanist Robert Brown, who first discovered the phenomenon in 1827. He observed the random motion of a pollen particle immersed in water on a microscopic scale. The motion was caused due to the collision between the pollen particle and the fast-moving water molecules. From a mathematical perspective, Brownian motion can be described as a limiting case of some random walk as its time increments approach zero, i.e. the number of steps becomes infinitely large. In the last decades, there have been various models that incorporate this phenomenon, especially in reference to stock and commodity price modelling.

A *standard* Brownian motion[5] represents a stochastic process (also referred to as a Wiener process), W_t, that can be defined in mathematical terms by three main properties:

1. W_t has a continuous path in t, and $W_0 = 0$.
2. The increment of the Brownian motion in the time interval of dt between the two points t and $t + dt$, where $dt \geq 0$, is $W_{t+dt} - W_t$. This increment follows a normal distribution with mean zero and variance equal to the time interval dt.

[5] A *standard* Brownian motion is the case when $\sigma^2 = 1$, as the increments of a Brownian motion follows a normal distribution $\mathcal{N}(0, \sigma^2 dt)$. (Throughout the thesis, when we say Brownian motion we are considering a *standard* Brownian motion.)

Mathematically expressed as; $W_{t+dt} - W_t \sim \mathcal{N}(0, dt)$.

3. The Markov property: W_t has independent increments, for every $t \geq 0$, the future increments $W_{t+dt} - W_t$, where $dt \geq 0$, are independent of the past values W_s, for $s \leq t$.

3.3.2 Geometric Brownian Motion Model

Ever since discovering the process of Brownian motion it has been extensively applied in multiple fields, including finance, to model behaviour of stock prices, commodity prices, macroeconomic factors, etc. The Geometric Brownian Motion (GBM) model, as the name suggests, incorporates the principle of Brownian motion to describe the behaviour of a continuous-time stochastic process. The model implies that returns of a variable of interest follow a lognormal distribution, and thus meaning that the logarithmic returns, which are continuously compounded returns, follow a normal distribution. Consistent with reality, the model restricts prices from falling below zero due to the nature of a lognormal distribution (i.e. maximum negative return is 100%). It models the asset price as a sum of a positive deterministic function of time and a stochastic Brownian motion term. A stochastic process, such as the asset's spot price S_t, is said to follow the GBM model if it satisfies the stochastic differential equation (SDE) given by equation (3.8), or modelled as the instantaneous rate of return on S_t given by equation (3.9):

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{3.8}$$

or

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \tag{3.9}$$

In the same way as Brownian motion was observed for the pollen particle, an asset's price deviates from a steady-state as a result of being altered by trades in financial markets. Considering an asset with spot price S_t at time t, a mean percentage drift of μ and a mean expected volatility of σ. Then the relative change in its price during the next period of time dt can be decomposed into a deterministic and a stochastic part. The deterministic and predictable part is the expected price change during a time period of dt. Hence, the change in price is equal to $\mu S_t dt$. Whereas the stochastic and unpredictable part mirrors the random changes in the asset's price during the time interval of dt. These random changes may have various underlying reasons, such as changed demand and supply for the asset, speculations and unexpected news about the asset, etc. The asset's mean volatility is represented by σ, and W_t is a standard Brownian motion process as described in the previous subsection. This Brownian motion process (W_t), is the sum of all preceding Brownian increments, (W_i). Each Brownian increment is computed by drawing a standard random variable, (z_i), from a normal distribution with mean zero and standard deviation one, i.e. $z_i \sim \mathcal{N}(0,1)$, and multiplied with the square root of the incremental time period dt.

$$W_i = z_i \sqrt{dt}$$

The Brownian discretized path now becomes the cumulative sum of each single Brownian increments, which defines W_t as:

$$W_t = \sum_{i=1}^{n} W_i$$

Generally, the deterministic part is called the drift term, while the stochastic part is called the diffusion term and gives the model features simulating a random walk process. A visual presentation of these two terms' effect on the model can be seen in Figure 3.1. The GBM model also possesses the Markov property, meaning that "the future, given the present state, is independent of the past" (Sigman, 2006). In this case, this transforms to; given the present value of S_t, the values of S_T, where $T > t$, are independent of the values of S_u, where $u < t$.

Figure 3.1: *Characteristics of a generalized Geometric Brownian Motion model, where the blue line displays the combined effect from a positive drift and a random walk process.*

The solution to the GBM model is found by applying the itô's lemma identity. For a function $f(S_t)$, where S_t satisfies the SDE of (3.8), then by use of Itô's lemma identity, the equation can be written as:

$$df(S_t) = \frac{\partial f}{\partial S_t} dS_t + \frac{1}{2}\frac{\partial^2 f}{\partial S_t^2} dS_t^2$$

With $dS_t^2 = S_t^2 \sigma^2 dt$, it becomes[6]:

$$df(S_t) = \frac{\partial f}{\partial S_t} dS_t + \frac{1}{2}\frac{\partial^2 f}{\partial S_t^2} S_t^2 \sigma^2 dt$$

By substituting $f(S_t)$ to $\ln(S_t)$, where:

$$\frac{\partial \ln(S_t)}{\partial S_t} = \frac{1}{S_t}, \qquad \text{and} \qquad \frac{\partial^2 \ln(S_t)}{\partial S_t^2} = -\frac{1}{S_t^2}.$$

Then by inserting equation (3.8) for dS_t this yields:

$$\begin{aligned} d\ln(S_t) &= \frac{1}{S_t}\left(\mu S_t dt + \sigma S_t dW_t\right) + \frac{1}{2}\left(-\frac{1}{S_t^2}\right) {S_t}^2 \sigma^2 dt \\ &= \mu dt + \sigma dW_t - \frac{\sigma^2}{2} dt \\ &= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t \end{aligned}$$

Considering the time interval dt, starting from $t = 0$ to a future point $t = t$. Then $d\ln(S_t)$ can be written as $\ln\big(\hat{S}_t\big) - \ln(S_0) = \ln\left(\frac{\hat{S}_t}{S_0}\right)$, which yields: (Remark from last subsection $W_0 = 0$.)

$$\ln\left(\frac{\hat{S}_t}{S_0}\right) = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t$$

Which after applying the exponential rule to both sides gives the final solution to equations (3.8) and (3.9) as:

$$\hat{S}_t = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right] \tag{3.10}$$

[6] As $dS_t^2 = S_t^2(\mu^2 dt^2 + 2\mu\sigma dt dW_t + \sigma^2 dW_t^2)$, for the limit $dt \to 0$, the terms dt^2 and $dt dW_t$ tend to zero quicker than dW_t^2. Hence, substituting dt for dW_t^2 (due to the second property of quadratic variance for a Wiener process in Section 3.3.1), and setting dt^2 and $dt dW_t$ to zero, yields this relationship.

A simpler form of equation (3.10) can be written as:

$$\hat{S}_t = S_0 e^{X_t} \qquad \text{with,} \qquad X_t = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t \tag{3.11}$$

Where X_t simply is a Brownian motion with a drift term of $\mu - \frac{\sigma^2}{2}$, and a diffusion term of σW_t, where σ is interpreted as a scaling parameter for the random walk process W_t. It is worth mentioning that if we neglect the diffusion term in our model (since μ and σ are constants based on historical data) the future price will move smoothly up or down depending on whether the drift is positive or negative, respectively. By including the diffusion term or random shock component, it allows for different future price scenarios, and thus gives the possibility for simulation. This diffusion term is also what makes the model a stochastic process, due to the incorporation of Brownian motion. The discrete-time GBM process $\hat{S}_t$, makes predictions for the asset price, t periods ahead, using only the current price, expected percentage drift and the expected volatility. The expected volatility can either be estimated from historical data (historical volatility) or by the volatility implied by the option market prices (implied volatility). However, the ideal volatility to use for forecasting purposes would be the future volatility, which would not be possible to know without knowing the future state of the market. Hence, the volatility used as input parameter should be the best estimate of future volatility and reflect future market expectations and belief for the time period of consideration. Furthermore, the estimated volatility is assumed to be constant, which as any trader knows does not reflect the reality observed in a market due to features such as volatility clustering[7]. Other more complex models include time-varying volatility and Risk Management models, which assumes that volatility fluctuates in a similar way as the actual prices do. The great features of the GBM model concerning financial asset or commodity price modelling, is the fact that the exponential function always yields positive integers, and only depends on the input parameters μ and σ to be estimated.

3.3.3 Mean Reverting Model

For commodities and interest rates modelling a Mean Reverting (MR) model has more economic logic than the GBM model presented before. Economic theory states that when the price of a good goes up, the demand will shrink and more supply will enter the market, such that there will be an oversupplied market. Conversely, as the price declines, the demand increases and there will be a tight market regarding supply. In

[7]Volatility clustering refers to the phenomenon observed in financial data where large changes tend to be followed by large changes in either direction, and similarly for small changes.

commodity markets, this works as a balancing mechanism. On the one hand, if the price increases to an abnormally high price, the market is expected to eventually revert to a lower price due to the resulting oversupplied market. On the other hand, if the price drops too low, the demand for the commodity will be high whereas the supply will shrink, and eventually the market is expected to revert to a higher price due to shortage in the supply. These price-reverting cycles are often observed in historical prices for commodities, however, they generally have varying magnitudes and does not occur in fixed time intervals. An MR model aims to capture this market balancing mechanism and various model specifications have been proposed in the literature for different mean reverting processes.

In financial mathematics, a model for describing the evolution of interest rates was proposed by Vašíček (1977), which has been labelled the Vašíček model. This is a mean reverting Ornstein-Uhlenbeck stochastic process where an additional drift parameter is included to represent the long-term equilibrium price level, θ. The Vašíček model is defined by the following continuous-time SDE:

$$dX_t = \kappa(\theta - X_t)dt + \sigma dW_t \tag{3.12}$$

A generalized mean reverting process with paths for various initial values of X_0 is illustrated in Figure 3.2. For initial values above or below the long term mean, θ, the process shows either a decreasing or increasing trend, respectively. Whereas for an initial value equal to the long-term mean, the process fluctuates around this value as a result of the stochastic behaviour from the diffusion term.

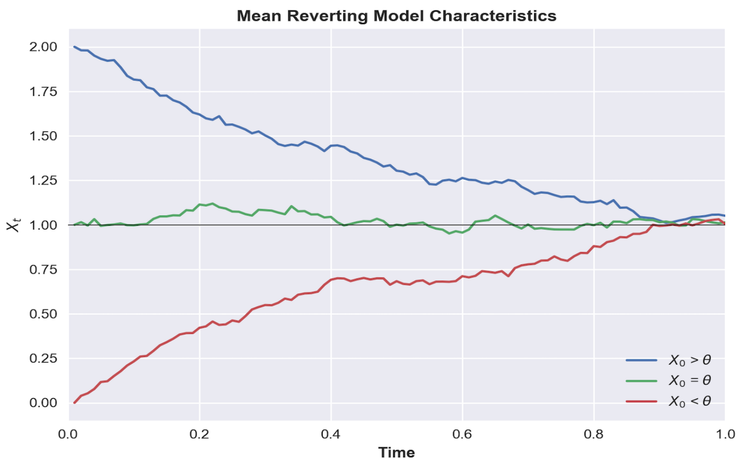


Figure 3.2: *Characteristics of a generalized Mean Reverting process, where* $\kappa = 2$, $\theta = 1$, $\sigma = 0.15$ *and* $\Delta t = 0.01$. *Initial values of* 2, 1 *and* 0 *for the blue, green and red lines, respectively.*

For the purpose of modelling crude oil prices, Schwartz (1997) applied the Vašíček model to model logarithmic crude oil prices. This corresponds to interpreting X_t as $\ln(S_t)$ in equation (3.12), where S_t represents the crude oil price at time t, and the time increment dt is infinitely approaching zero. The magnitude of the speed of adjustment κ (> 0) measures the rate of mean reversion to the long-term mean logarithmic crude oil price, θ. In the second term, σ is the volatility of logarithmic oil prices, and $dW_t \sim \mathcal{N}(0, dt)$ being an increment to standard Brownian motion. The first term corresponds to the drift term, while the second term is the diffusion term which induces stochastic behaviour. The parameters; κ, θ and σ, are considered to be constants in this specification, although other specifications of this model have incorporated a time-varying volatility (e.g. by modelling the volatility process with one of the GARCH[8]-class models proposed in the literature).

Transformation of a theoretical continuous-time model into a relevant discrete-time interval, will allow the usage of historical crude oil price data for parameter estimation. For sufficiently small enough time intervals, the process can be considered as the continuous-time approximation of a discrete-time AR(1) process. Which by expanding and rearranging terms in equation (3.12) yields:

$$X_{t+\Delta t} = \kappa\theta\Delta t + X_t(1 - k\Delta t) + \sigma(W_{t+\Delta t} - W_t) \tag{3.13}$$

Considering equation (3.13), the following empirical representation can be estimated by an ordinary least squares method (OLS):

$$X_{t+\Delta t} = a + X_t b + \varepsilon_{reg} \tag{3.14}$$

Where $X_{t+\Delta t}$ represents the logarithmic oil price Δt time forward, with e.g. monthly frequency of discrete historical data, $\Delta t = \frac{1}{12}$. The direct linkage between equations (3.13) and (3.14) suggest the following relationships; an intercept of $a = \kappa\theta\Delta t$, a coefficient of $b = (1 - k\Delta t)$ and distribution of the regression residual as $\varepsilon_{reg} \sim \mathcal{N}(0, \sigma^2\Delta t)$. As the residual incorporates a Brownian motion increment and a constant volatility parameter, by knowing the theoretical distribution properties of the residual where $Var[\varepsilon_{reg}] = \sigma^2\Delta t$, it becomes possible to estimate the volatility. From these relationships, the constant parameters of the MR model can be estimated as follows:

[8]Short for Generalized Autoregressive Conditional Heteroskedasticity

$$k = \frac{1-b}{\Delta t}, \qquad \theta = \frac{a}{\kappa \Delta t}, \qquad \text{and} \qquad \sigma = \sqrt{\frac{Var[\varepsilon_{reg}]}{\Delta t}}. \tag{3.15}$$

By setting the variable $X_t = \ln(S_t)$ and applying Itô's lemma, the stochastic differentiated process of equation (3.12) can be discretized and approximated by[9]:

$$\ln\left(\hat{S}_{t+\Delta t}\right) = \theta(1-e^{-\kappa\Delta t}) + e^{-\kappa\Delta t}\ln(S_t) + \sigma\sqrt{\frac{1-e^{-2\kappa\Delta t}}{2\kappa}}N_{[0,1]}, \tag{3.16}$$

where $\ln(S_t)$ denotes the natural logarithm of the crude oil price at time t, and $N_{[0,1]}$ is an independent identically distributed standard normal variable (or similar to the increments of a Brownian motion process) with zero mean and unit variance. Having estimated the parameters' values; κ, θ and σ by OLS, it becomes possible to simulate logarithmic oil prices having an initial price. This closed-form solution prevents the values from going negative, as a logarithmic oil price of below zero still would correspond to an oil price above zero after exponentiation. In contrary to the GBM model, the MR model assumes that price changes are not completely independent of one another, but rather are related and the evolution of price and its volatility is bounded.

[9]See for example; Pereboichuk (2014) p.44 and Bahar et al. (2017) p.423.

4 Data Description and Forecasting Procedure

A model's forecasting performance is normally conducted by splitting the relevant data set into an in-sample set used for initial parameter estimation and model selection, and an out-of-sample set used to evaluate and analyse the forecast performance. Commonly the in-sample set is denoted as the training set, whereas the out-of-sample set is denoted as the test or evaluation set. Empirical evidence of out-of-sample forecast performance is considered more reliable than evidence based on in-sample forecast performance which can be more sensitive to outliers and data mining. Out-of-sample forecasts also reflect a more realistic scenario of the information available to the forecaster in "real-time" when attempting to forecast a variable of interest. Hence, this has led the majority of researchers to regard out-of-sample forecast performance as the "ultimate test" for a proposed model (Hansen and Timmermann, 2012). For an out-of-sample forecast to be unbiased and valid there should be no information leakage from data in the out-of-sample period considered. Therefore, only information from the training set is used for model specification, while the test set is used for model validation only. A true out-of-sample forecast, also referred to as ex-ante forecast, implies the use of data all the way up to the present time at when the forecast is to be done. This approach involves a cumbersome procedure to accumulate sufficient amounts of data points in order to examine models' forecasting abilities. In response to this, the pseudo out-of-sample, or also referred to as ex-post, forecasting procedure was developed. A pseudo out-of-sample procedure implies the use of a splitting point, at which the data on the one side represents the in-sample data, whereas the data on the other side represents the "unknown", out-of-sample data to be forecasted and evaluated against.

This chapter is structured as follows: Section 4.1 introduces the monthly and daily sampled crude oil price data sets utilised for forecasting together with the previous chapter's selected models. Different sample sets defined as training and test sets are presented in Section 4.2. Finally, in Section 4.3, various evaluation criteria used to measure forecast accuracy are defined and described.

4.1 Data

As discussed in Section 2.2, there exists no uniform crude oil price due to quality differentials. However, the most popular crude benchmarks serve as good proxies to model the crude oil price as a unique variable. For the remaining part of this study, the benchmark WTI crude oil will be used as this proxy. The price data for WTI crude oil was retrieved from EIA's website[10] in both daily and monthly frequencies between the start of January 1986 till the end of January 2020 (daily frequented prices can be seen in Figure 2.1). Monthly sampled prices were mainly collected to be able to compare the futures-based models, as these models are based around the contract's monthly maturities. In Table 4.1 descriptive statistics are printed for the data set used for the different frequencies. The daily and monthly price histories show observations of 8609 and 409, respectively. Crude oil prices during the sample period have a mean of $44 and ranges from as low as $10 to over $145 per barrel.

Table 4.1: *Descriptive statistics for daily and monthly crude oil prices (January 1986-January 2020)*

Crude Oil Prices

	Observations	**Mean**	**Std.Dev**	**Min**	**Max**
Daily	8609	$44.2	$29.2	$10.3	$145.3
Monthly	409	$44.2	$29.1	$11.4	$133.9

When modelling crude oil prices the daily or monthly price fluctuations are an important aspect. The daily and monthly logarithmic price returns give an indication of how prone the prices are to fluctuations. Large relative changes often observed in such data is what causes the main problem for any forecasting model as the underlying effects of the price change often are unexpected and unforeseen. In Table 4.2 descriptive statistics of crude oil price fluctuations, calculated as the logarithmic returns, are presented. The absolute mean is the mean of all absolute price returns, which shows that on average the crude oil changes by 1.71% and 6.46% in either direction for the daily and monthly sampled data sets, respectively. Large fluctuations in oil prices are confirmed by the minimum and maximum price returns seen. Daily and monthly standard deviation in price returns characterise volatility of oil prices, and for the sample period, these are 2.49% and 8.65%, respectively. Inspection of the skewness statistic reveals that the price fluctuations are negatively skewed compared to a symmetrical bell curve, or normal distribution. Meaning that the left tail of the density plot centred at the mean is longer or fatter relative to the right tail. For computation of the kurtosis, the Fisher's kurtosis formula was used, which defines a normal distribution to have a kurtosis of zero. Thus, the kurtosis observed is equivalent of the excess kurtosis, which indicates a heavy-tailed distribution compared to a normal distribution. The excess kurtosis for the daily frequented data indicates as expected

[10]Data sourced from: `https://www.eia.gov/dnav/pet/hist/RWTCD.htm`

more extreme outliers compared to the monthly averaged data. As both the daily and monthly price fluctuations have positive excess kurtosis, their distributions are said to be leptokurtic, which from a financial perspective indicates large risk for extreme price fluctuations on either side. These descriptive statistics of the return's distributions are consistent with stylized features of financial returns as noted in Dbouk and Jamali (2018).

Table 4.2: *Descriptive statistics for daily and monthly logarithmic crude oil price returns (January 1986-January 2020)*

Logarithmic Crude Oil Price Returns

	Observations	**Absolute Mean**	**Std.Dev**	**Min**	**Max**	**Skewness**	**Kurtosis**
Daily	8608	0.0171	0.0249	-0.406	0.192	-0.630	13.6
Monthly	408	0.0646	0.0865	-0.394	0.392	-0.447	2.67

4.2 Sample Horizons

Splitting the data set into a training and testing set is a choice variable which there is no broadly accepted guideline for how it should be done. Welch and Goyal (2008, p. 1464) state that "It is not clear how to choose the periods over which a regression model is estimated and subsequently evaluated", whereas Stock and Watson (2015, p. 613) recommend "Pick a date near the end of the sample, estimate your forecasting model using data up to that date, then use that estimated model to make a forecast". Researchers have adopted several different approaches to measure pseudo out-of-sample forecast accuracy. One common approach is to choose an initial training sample set for estimation and use the remaining data with various lengths as forecast evaluation samples. Another approach is based on the belief that more recent data better reflect the current state of the variable of interest. This approach suggests the choice of training data to be selected at points as near the present time as possible where noticeable structural breaks are observed in the data set, while still remaining a sufficiently large enough data set for model estimation. Consider the case, where a new improved technology enters the market for exploration and production of oil and impose a pivotal change in the market (e.g. how fracking improvements changed the U.S., from heavily depending on imported foreign oil to producing enough for its own domestic consumption and international exports in less than two decades). Such a development could potentially change the current market in a substantial way, and the inclusion of market data before this point in the training set could potentially jeopardize model selection and parameter estimations. This brings into question how important and significant the sample split point is to the model selection and forecasting performance. According to Hansen and Timmermann (2012), the split point is not irrelevant and generate a potential for data mining, implying that not even the aforementioned "ultimate test" for forecasting is immune to data mining.

I will now proceed to present how the train and test sample horizons were constructed. Consider the whole data set to be of length T, then, for a selected forecast horizon of length h, the train sample size will be from the start date specified up to and including date $T-h$. Whereas the out-of-sample or test sample will be from dates greater than $T-h$ till the end date of the data T. A visual representation of the estimation and forecast windows can be seen in Figure 4.1.

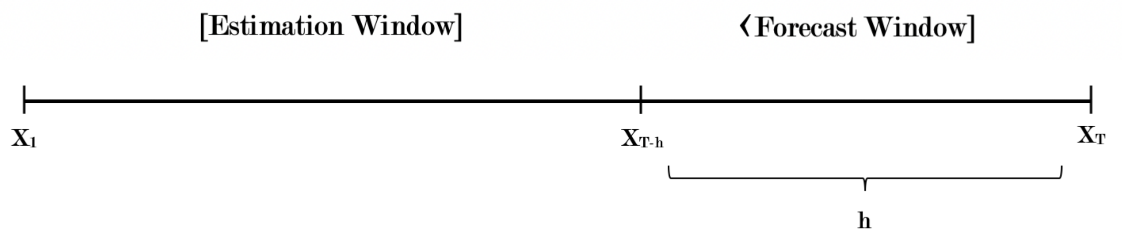


Figure 4.1: *Timeline of estimation window and forecast window.*

Different start dates used to construct the train samples for daily and monthly prices are presented in Table 4.3. As daily prices are not available for all weekdays during a year, the first available price at the start of the month will be used. Four different train sample lengths will be evaluated. With the *Long* sample length starting at the beginning of the time series. The *Medium* sample length start date was set to capture the strong growth in global economic activity driven by developing market economics in 2002, which caused a prolonged period of increasing crude oil prices. The idea behind the *Short* sample length was to exclude the global financial crisis of 2008, where the crude oil price dropped over $100 in just five months. Finally, the *Recent* sample length was constructed in a way to exclude the steep fall observed in crude oil prices as a result of great oversupply in 2014. The various train sample lengths will be denoted by L, M, S and R throughout the thesis.

Table 4.3: *Specification of train sample's start dates used for model fitting.*

Train Sample	
Sample Length	**Start**
Long (L)	01/01/1986
Medium (M)	01/01/2002
Short (S)	01/01/2009
Recent (R)	01/01/2016

In Table 4.4, different start dates used for the desired forecast horizons are specified. The table contains both dates for the monthly and daily frequented price series and represents the start date of the test sample. For the monthly sampled prices, the forecast horizon ranges from 1 to 24 months. Whereas for the daily sampled prices, the forecast horizon ranges from 1 to 256 days.

Table 4.4: *Specification of test sample's start dates used for forecasting.*

Test Sample			
Monthly		**Daily**	
Forecasted Months	**Start**	**Forecasted Days *(Months)***	**Start**
1	01/01/2020	1	31/01/2020
3	01/11/2019	5	27/01/2020
6	01/08/2019	22 *(1)*	01/01/2020
9	01/05/2019	64 *(3)*	01/11/2019
12	01/02/2019	130 *(6)*	01/08/2019
24	01/02/2018	256 *(12)*	01/02/2019

Referring to Figure 4.1, as an example of how an arbitrary train and test sample set were constructed, let us consider monthly frequented data with *Medium* train sample length and the desired forecast horizon of 6 months. This would suggest an estimation window lasting from 01/01/2002 up to and including 01/07/2019, and a forecast window lasting from 01/08/2019 up to and including 01/01/2020 (i.e. end of data set). The specified train and test start dates, gives a total of 48 different train and test samples to be evaluated for monthly and daily frequented prices.

4.3 Forecast Evaluation

The main focus of this thesis is to generate forecasts of the crude oil price with different models proposed in the literature and to evaluate their out-of-sample forecast accuracy. Hence, there is a need for some evaluation criteria to rank and quantify a model's forecast performance. This is typically accomplished by using loss (or cost) functions. In the literature, there exists a vast amount of proposed loss functions for evaluating a model's prediction performance. However, the choice of loss functions to be utilized introduces a level of arbitrariness to model ranking, as different loss functions penalize forecast errors differently. For this reason, several loss functions will be evaluated in order to establish to what extent the choice of loss functions impacts the ranking of models' forecast performance.

When a forecast $f_{t,h}$, of a true variable S_{t+h}, is made at time t for h periods ahead, the loss will emerge if the forecast turns out to be different from the actual value observed. The forecast error or bias is defined as $e_{t+h} = S_{t+h} - f_{t,h}$, and the loss function L is derived from the argument e, which is dependent upon S_{t+h} and $f_{t,h}$, i.e. $L(e) = L(S_{t+h},\ f_{t,h})$. For a true out-of-sample forecast (i.e. ex-ante forecast), the value of S_{t+h} will not be known until time $t+h$ has passed by. While the main advantage of a pseudo out-of-sample forecast (i.e. ex-post forecast) is that this value will be known to the analyst at time t, and thus allows for a derivation of an incurred loss function to be assessed. According to the discussion of Granger (1999), loss functions must possess several different properties. First, if the forecasted value is exactly equal to the true observed value, then $e = 0$ and there will be no loss (i.e. $L(0) = 0$). Second, $\min_e L(e) = 0$, so that $L(e) \geq 0$, meaning that the loss function is defined as a numeric value taking on any value greater than or equal to zero, i.e. $L \in [0, \rightarrow \infty\rangle$. Finally, $L(e)$ must be continuous and monotonically non-decreasing as e moves away from zero. This implies that $L(e_2) > L(e_1)$ if and only if $e_2 > e_1$, and indicates that the forecast producing e_2 is statistically worse than the forecast producing e_1.

End user's intended usage of the forecasted values differs greatly, and optimal model selection heavily depends upon their individual utility functions, which for this thesis remains unknown. For this reason, a model's forecast accuracy performance may differ among various end user's who employs different loss functions tailored to their preferences. These loss functions are defined as economic loss functions and will not be feasible to adopt within this thesis. In response to this, statistical loss functions which are aimed at evaluating average forecast errors serve as convenient measures and emerges as the dominant practice in the literature. However, among the statistical loss functions proposed there exists a variety of differences. The most essential differences are whether a loss function is asymmetric or symmetric with respect to under- and overpredictions given equally-sized absolute errors, and whether the function is linear or non-linear. Certain end users may have strong preferences to whether an underprediction is more or less preferred than an overprediction, and may thus, employ an asymmetric loss function. While if under- and overpredictions is weighted equally bad, a symmetric loss function is appropriate to apply. The other aspect of a loss function is whether the absolute value of forecast errors should be penalized linearly or non-linearly. A linear loss function will use a linearly proportional weighting scheme to penalize growing errors, whereas a non-linear loss function will typically use some different non-linear weighting scheme to penalize the errors.

To properly evaluate any forecast model we will employ a set of loss functions with different characteristics. One of the simplest loss metrics is the Mean Error (ME), which is the average error or bias representing the systematic error of a forecast model to under- or overpredict the target values. This metric penalizes over- and underpredictions equally and applies a linear weighting scheme. A disadvantage when analysing this metric is situations where positive and negative errors simply cancel out by summing to zero, which may be misinterpreted as similar to a perfect forecast where $S_t = f_t$. Perhaps the most commonly adopted loss function, the RMSE, which represents the standard deviation of the forecast errors. This loss function gives more weight to extreme outliers and penalizes under- and overpredictions equally. RMSE is computed by taking the square root of the underlying MSE function, which then yields a loss metric in the same units as the quantity measured. In addition to the fundamental ME and RMSE loss functions, we will employ three loss functions derived from following the proposed family of loss functions by Patton (2011). Patton developed a family of parametric loss functions nested from two of the most widely used loss functions in the literature, namely the MSE and QLIKE loss functions. The proposed class of loss functions depends on a shape parameter, b, that allows for asymmetric penalties to be applied for underpredictions ($b < 0$) and overpredictions ($b > 0$), as well as a symmetric penalty ($b = 0$). Where the latter case turns out to be equivalent to the MSE loss function. To evaluate our models' forecast performances, variations of Patton's suggested subset of loss functions will be used, and is given by the formula:

$$L(S_t, f_t; b) = \begin{cases} \frac{1}{(b+1)(b+2)}(S_t^{b+2} - f_t^{b+2}) - \frac{1}{b+1} f_t^{b+1}(S_t - f_t), & \forall\ b \notin \{-1, -2\} \\ f_t - S_t + S_t \log \frac{S_t}{f_t}, & \text{for } b = -1 \\ \frac{S_t}{f_t} - \log \frac{S_t}{f_t} - 1, & \text{for } b = -2 \end{cases}$$

Following the proposed family of loss functions formula, we will employ the QLIKE loss function, in addition to two extra loss functions. One with a penalizing effect on underpredictions ($b = -1$), and one on overpredictions ($b = 1$), which will be denoted as L4 and L5, respectively. All five loss functions to be considered throughout the thesis are presented in Table 4.5. Whereas in Figure 4.2 their underlying functions' shapes are plotted for imaginary data points with respect to their absolute loss.

Table 4.5: *Loss functions employed to evaluate forecast performance. This table presents the loss function's abbreviated name, Patton's shape parameter (b), loss function formula and a short description of the function's behaviour. Table inspired by work from Lorentzen and Sharma (2015).*

Loss Function	Name	Parameter	Formula	Description
L1	ME	NA	$E[S_t - f_t]$	Symmetric and linear
L2	RMSE	0	$E\left[\sqrt{(S_t - f_t)^2}\right]$	Symmetric and non-linear
L3	QLIKE	-2	$E\left[\frac{S_t}{f_t} - \log\frac{S_t}{f_t} - 1\right]$	Asymmetric and non-linear
L4	NA	1	$E\left[\frac{1}{6}(S_t^3 - f_t^3) - \frac{1}{2}f_t^2(S_t - f_t)\right]$	Asymmetric and non-linear
L5	NA	-1	$E\left[f_t - S_t + S_t\log\frac{S_t}{f_t}\right]$	Asymmetric and non-linear

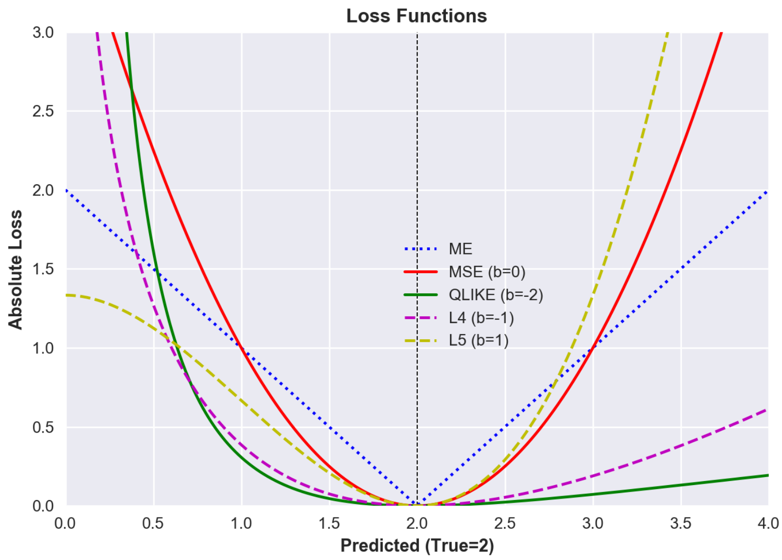


Figure 4.2: *Visualization of utilized loss functions characteristics with respect to their absolute loss, against predicted values ranging from 0 to 4 versus a true value of 2. For the RMSE loss function, the underlying MSE function is plotted.*

Finally, in addition to the aforementioned loss functions, an SR metric will also be used. That is, to what degree are the forecasts able to predict the true price change's directions. This is not technically a loss function, but rather a descriptive statistic. The SR for the forecast window is computed from the following relationship:

$$s_i = \begin{cases} 1, & \text{if } (S_{t+i} - S_t)(f_{t+i} - f_t) \geq 0 \\ 0, & \text{if } (S_{t+i} - S_t)(f_{t+i} - f_t) < 0 \end{cases}$$

Then by taking the sum of s_i for all steps i, and dividing by the number of price changes during the forecast window (h), the SR becomes:

$$SR = \frac{\sum_{i=1}^{h} s_i}{h}$$

As stated by Baumeister and Kilian (2012) "Under the null hypothesis of no directional accuracy, the success rate of the model at predicting the direction of change in the price of oil should be no better than tossing a fair coin with success probability 0.5". Hence, for a model to have any statistically significant evidence of directional accuracy, we should expect an SR above 0.5. The significance is also determined by the test sample length, as for relatively short forecast windows a high SR is in general not indicative of true directional accuracy. The SR will be further evaluated based on the test statistic developed by Pesaran and Timmermann (1992). This test, which from now on is referred to as the PT-test, is formally defined under the null hypothesis that a given candidate forecasting model is equally directional accurate as the No-Change benchmark model or equivalent to tossing a coin to predict the sign of actual price change. The null is tested against the alternative hypothesis that the candidate model is more directionally accurate than the benchmark model. For all forecast windows considered, suitably constructed p-values will be reported based on the rejection of the aforementioned null hypothesis.

4.3.1 Diebold-Mariano Test

Testing various models' forecasting performances requires some evaluation test of whether the observed improvements are significantly or not compared to the benchmark model. Consider two models' forecasts to be $f_1, \ldots, f_h$ and $g_1, \ldots, g_h$, with the former being that of the benchmark model and the latter of any competing model. The obvious approach is to select the forecast model that produces the smallest error measurements based on the considered loss functions. This approach does not contemplate whether this difference in predictive accuracy is significant, or simply due to the specific choice of data sample values and pure chance. To eliminate this confusion a test statistic proposed by Diebold and Mariano (1995) will be applied to the forecasted results. The test will be referred to as the DM-test, and its procedure to evaluate the significance level of predictive accuracy between the benchmark model and all competing models will now be outlined.

Let ε_t and r_t denote the residuals of each model compared to the actual observed prices (S_t) in the out-of-sample data set, i.e.

$$\varepsilon_t = S_t - f_t \qquad\qquad r_t = S_t - g_t$$

and let the difference between those residuals, d_t, be defined as:

$$d_t = \varepsilon_t^2 - r_t^2$$

The time series of d_t is called the loss-differential, and the formula is clearly related to the MSE loss function, although other loss functions could be employed as well. Further, let the average of the loss-differentials equal:

$$\overline{d} = \frac{1}{h}\sum_{t=1}^{h} d_t$$

Under the null hypothesis, this test measures the significance of $E[d_t]$ being equal to zero for all t, which would imply equal predictive accuracy of the two forecasting models in consideration. Usually, multi-step forecast errors exhibit some degree of autocorrelation, and where an efficient s-step forecast will have forecast errors following an MA($s-1$) process. To incorporate this into the test, the autocorrelation function γ_k, being the estimated autocovariance at lag k is defined as:

$$\gamma_k = \frac{1}{h}\sum_{t=k+1}^{h}(d_t - \overline{d})(d_{t-k} - \overline{d})$$

Finally, the DM-test statistic is computed as:

$$DM^* = \frac{\overline{d}}{\sqrt{\left[\gamma_0 + 2\sum_{k=1}^{s-1}\gamma_k\right] / h}} \tag{4.1}$$

Under the assumption of the null hypothesis that $E[d_t] = 0$, the DM statistic follows an asymptotic normal standard distribution i.e. $DM^* \sim \mathcal{N}(0,1)$. Hence, the two competing forecasts have different predictive accuracy if $|DM^*| > z_{crit}$, where z_{crit} is the two-tailed critical value of a standard normal distribution.

The originally proposed test of Diebold and Mariano actually tends to reject the null hypothesis too often for small samples. Therefore, Harvey et al. (1997) proposed a modified test where improved small-sample size properties can be achieved by; (i) making a bias correction to the test statistic and (ii) comparing the corrected statistic with a Student-t distribution with (n-1)[11] degrees of freedom instead of the standard normal distribution. Their improved and modified test statistic is computed as:

$$DM = \sqrt{\frac{h+1-2s+s(s-1)/h}{h}} DM^*$$

Where DM^* is the original test statistic from equation 4.1. For the forecast results later on reported, p-values constructed from this modified test statistic are reported as DM-test p-values. Any p-values less than a reasonable significance level implies the rejection of the null hypothesis of equal predictive accuracy. In order to choose the model with significantly better performance, the model with best scores among the loss functions employed is chosen.

[11]With n being the number of forecasted values.

5 Forecast Results

For the imminent chapter forecast results of the models in Sections 3.2 Selected Deterministic Forecasting Models and 3.3 Selected Stochastic Forecasting Models will be evaluated. As all models were tested with both daily and monthly price frequencies and 48 combinations of train and test sample lengths, forecast results will be reported in Appendix A Forecast Results Tables. All selected models will be compared against the parsimonious No-Change model. In the research, this model is typically used as a benchmark and more often than not performs better than most complex models and professional survey forecasts for short forecast horizons.

5.1 Benchmark Model: Random Walk

The No-Change model has proven to serve as a superior forecast compared to the competing models for very short horizons of typically one to a few steps into the future. However, as this is a static forecast, and oil prices are prone to fluctuate vastly over relatively short time spans, this model becomes less and less accurate as the forecast horizon is extended. Also, if future oil price levels undergo major changes due to market events, this model may become impractical to use for any forecast horizons. In Tables 5.1 and 5.2 the actual values for respectively daily and monthly forecast results are reported for all horizons and loss functions. It becomes clear from the ME loss function, that this model's accuracy oscillates between under- and overpredicting future prices at the considered forecast horizons.

For daily prices and the L train sample length, the No-Change model serves as most appropriate at forecast horizons of 5 and 22 days. Whereas for the M train sample, it additionally outperforms all other models at a forecast horizon of 130 days. Finally, for the S train sample, it performs best at horizons of 22 and 130 days, whereas using the R train sample, only at a horizon of 1 day. However, using L and S train sample lengths the other models only appear to have significantly worse performance compared to the benchmark model at the forecast horizon of 22 days. While using the M train sample, there is a significantly worse performance at a forecast horizon of 130 days. Whereas for monthly prices and the L, M and S train sample lengths, none of the tested models beats the benchmark model at forecast horizons of 1 and 6 months, considering any loss function. Although the benchmark model seems to be superior at forecast horizons of 1 and 6 months into the future, there is not enough data points to determine whether the other models' performances are actually significantly worse or due to pure chance.

Table 5.1: *Forecast results of the No-Change model for daily frequented prices for all train sample lengths.*

Daily Prices: No-Change

Forecast Horizon	Loss Function				
	ME	RMSE	QLIKE	L4	L5
Train Samples: L, M, S and R					
1	-0.610	0.610	0.00007	9.7	0.004
5	-1.394	1.557	0.00043	64.8	0.023
22	-3.456	4.901	0.00349	705.4	0.205
64	4.213	4.950	0.00390	687.3	0.218
130	-1.855	3.327	0.00169	317.0	0.097
256	3.676	4.990	0.00393	702.0	0.221

Table 5.2: *Forecast results of the No-Change model for monthly frequented prices for all train sample lengths.*

Monthly Prices: No-Change

Forecast Horizon	Loss Function				
	ME	RMSE	QLIKE	L4	L5
Train Samples: L, M, S and R					
1	-2.200	2.200	0.00069	143.1	0.041
3	4.237	4.409	0.00315	540.4	0.175
6	-0.632	2.030	0.00064	117.0	0.036
9	-6.843	7.184	0.00692	1578.8	0.422
12	6.129	6.729	0.00774	1228.0	0.418
24	-2.990	7.016	0.00657	1512.7	0.402

To evaluate the forecast results and significance of the RW with Drift model, the reader is referred to Tables A.1 and A.6. From which it is seen that the model outperforms the No-Change model at forecast horizons of 64 and 256 days for all train sample lengths and daily prices. Whereas for monthly prices it outperforms the benchmark at 3 and 6 months horizons. For both price frequencies, there is statistical evidence that the model is more accurate than the benchmark at the aforementioned horizons. Although, as the model predicts prices with an upward sloping trend, the success rate of this model is typically fixed at near 0.5, which is equivalent to the accuracy of predicting a random coin toss. Hence, for all horizons and both price frequencies, there is never seen any statistical significance in the directional accuracy.

5.2 Futures Models

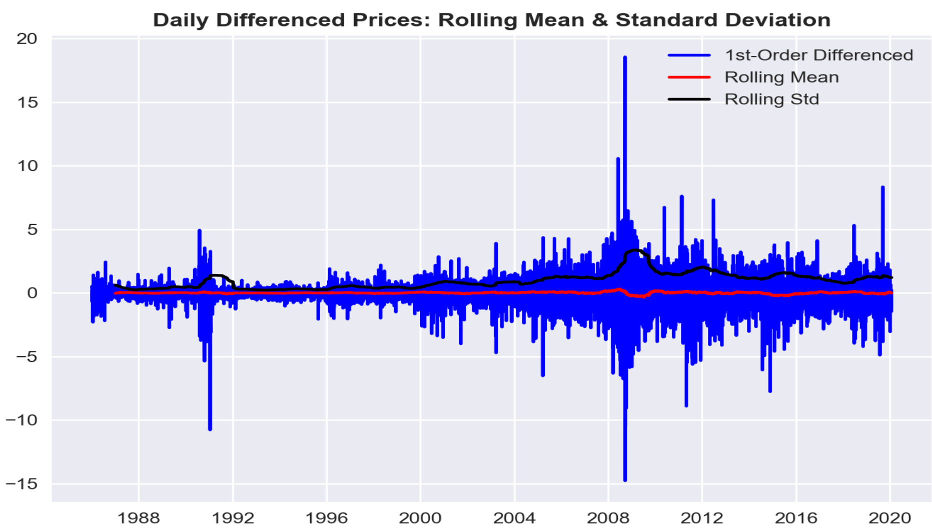
Ever since the 1970s energy crisis which led to the introduction of derivatives contracts for petroleum-related products, speculators and hedgers have increasingly traded actively through this financial layer. With speculators attempting to 'bet against' the market's perception of future price changes in order to pursue economical profits. Whereas hedgers attempt to reduce the amount of risk associated with price changes of the underlying commodity. This is achieved by the hedger from taking offsetting positions contrary to what they already have, and thus balances out any gains and losses of the underlying. Crude oil derivatives markets induce market transparency and liquidity in trading. With the effect of leverage and low trading costs, these markets attract speculators looking for any potential of arbitrage deals, and as their activity increases, so does the information impounded into the derivative's market price. The combined effects from derivatives markets ultimately impact the commodity price through arbitrage activity, leading to a more extensively based market in which current price corresponds more to its true value. Because this price influences production, consumption and storage decisions, derivatives markets contribute to an efficient allocation of resources in the economy (Fleming and Ostdiek, 1999). The evolution of crude oil derivatives markets has transformed the pricing of oil as a physical commodity more towards that of an investable asset. Resulting in its pricing becoming more intertwined with financial market dynamics and phenomena unrelated to supply and demand.

Historical prices of futures contracts with varying lengths of maturity were extracted from Refinitiv Eikon's financial database. For all forecast horizons, a continuous monthly time series of futures prices were obtained. Both the Futures and Spread models provide statistically more accurate forecasts than the No-Change model at forecast horizons of 3, 9, 12 and 24 months. Their directional accuracy is generally worse than that of a random coin toss, and thus neither significant. Both models' forecasts results are presented in Tables A.7 and A.8. The two models produce very similar results, although the Spread model is slightly more accurate. Especially if attention is drawn to the RMSE loss function, i.e. if there is no asymmetric loss weighting towards under- and overprediction.

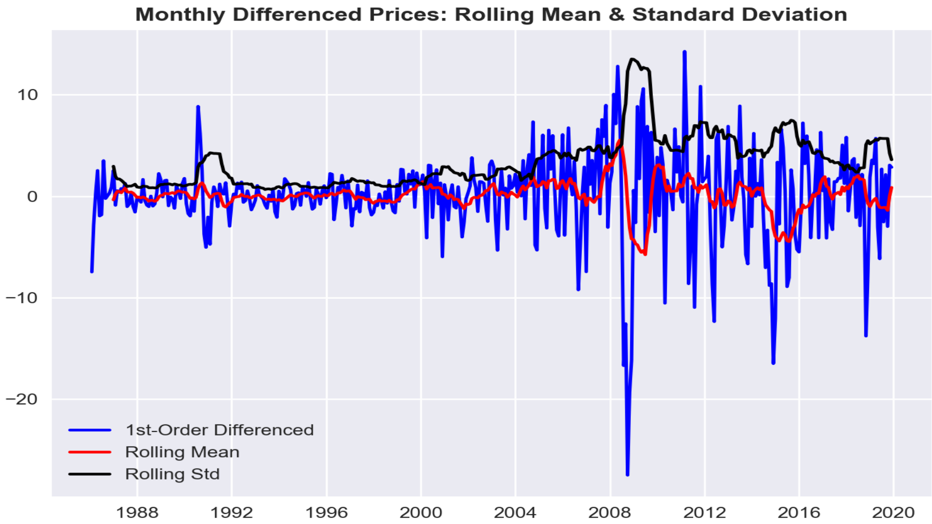
5.3 ARIMA Model

In order to apply the ARIMA model to crude oil prices, the time series needs to be transformed into a stationary one. This was done by performing a first-order differencing to the price series. As seen in Figures 5.1a and 5.1b, the daily and monthly differenced prices exhibits traits as expected for a stationary time series for the majority of the duration. However, during periods with extreme volatilities in the market such as during the financial crisis of 2008/2009, the rolling mean and rolling standard deviation fluctuates greatly. It is not possible to fully uncover whether the price series are stationary solely based on these plots. Hence, an Augmented-Dickey-Fuller test was performed, which states that under the null hypothesis a time series has a unit-root, and is thus non-stationary. The alternative hypothesis states that there is no unit-root, and the time series has no time-dependent structure and is therefore stationary. The test statistics of both first-order differenced daily and monthly frequented prices suggest there is no unit root, and the null hypothesis is rejected at less than a 0.1% significance level.

Because the ARIMA model is dependent upon the specified order of (p,d,q), there is potential to test various model specifications, and which yields best out-of-sample forecast accuracy is not clear. A common approach to identify the optimal order is based on performing a grid search over a range of parameters and their combined orders. To determine which order best fits the data, a score based on the Akaike Information Criteria (AIC) was used. This score is a widely used measure as it quantifies not just the model's goodness of fit, but also the parsimony of the model, into a single statistic. Hence, there is a penalty for adding more complexity to the model. Due to limiting computing software, the AIC score was calculated for all parameter combinations of p and q in the range of [0, 5], and the model order with lowest AIC score was chosen. The Model selected from this approach is referred to as the ARIMA model.



(a) Daily Prices



(b) Monthly Prices

Figure 5.1: *Visualization of differenced daily and monthly WTI price series, with rolling mean and rolling standard deviation. For both series a rolling window of one year was used, hence a window of 252*[12] *and 12 was used for daily and monthly prices, respectively.*

The optimal specified ARIMA model is well capable of fitting the time series in-sample, but using the same order for out-of-sample forecasting does not produce huge improvements compared to the benchmark model. All forecast results and significance of the ARIMA model is presented in Tables A.2 and A.9. Considering daily prices, the model shows significant improvements mainly at forecast horizons of 64 and 256 days. Although, using L and S train samples, resulted in improvements at horizons of 130 days and 1 day, respectively. While for monthly prices the results were significantly better but for less of the horizons considered. For all train sample lengths, a significant improvement at the horizon of 12 months was obtained. By using L and M train samples, there was in addition improvement at the 24 months horizon, and for the R train sample improvement at the 3 months horizon. Using either daily or monthly frequented prices resulted in the same way as the RW with Drift model rarely success rates above 0.5, and based on the PT-test statistics these were neither significant.

Out-of-sample forecasting accuracy of a specific ARIMA model is not comparable to the in-sample fit. Hence, the optimal specified model based on the AIC, may not yield satisfying forecast results. For this reason and based on the autocorrelation and partial autocorrelation plots of the stationary differenced price series, an AR(3) model

[12]There are on average 252 tradings days in a year, thus this was used as an approximation.

was also tested for its forecasting ability. Forecast results are reported in Tables A.3 and A.10. For daily prices, there were little improvements from this model compared to the ARIMA model, except for the M train sample and a forecast horizon of 22 days. Otherwise, it produced, similar to the ARIMA model, improvements at horizons of 64 and 256 days for all train samples. While for monthly prices, all train samples resulted in improvement at the horizon of 12 months, with the addition of the 3 months for the R train sample. Nonetheless, this model showed larger reductions in most loss functions at the horizons it improved upon the benchmark model's scores. There was also no significant evidence of this model being directional accurate.

5.4 Geometric Brownian Motion Model

Despite econometric theory suggesting a rather simple relationship between the price of a product and its supply and demand. The price of a commodity like crude oil is influenced by multiple irregular factors and shows strong evidence of randomness and stochastic behaviour. With that said, if crude oil prices follow the same path as a stochastic process, there would be no way to accurately predict its future price using a stochastic model, and if so, it would at least partially be a result of pure luck. Because the GBM model includes a diffusion term, each time the model is estimated with all parameters equal, different results would be produced. For this reason, it is very hard to conclude on a stochastic model's forecasting performance. However, an approximation of its predictive power can be deducted by performing multiple simulations and interpreting the mean of all simulations as the expected 'on average' forecast performance. One clear drawback by averaging the simulations is that the diffusion term σW_t in equation 3.10 to a large degree vanishes due to the Brownian increments being centred around a zero mean. This results in shocks of the predicted crude oil prices that are not as significant in size as one would expect for the commodity.

Daily and monthly 'on average' forecast results achieved by the GBM model are reported in Tables A.4 and A.11, respectively. The GBM model shows significantly better performance than the No-Change model when using daily prices for all train samples at a forecast horizon of 64 days. Using the L, M and S train samples, the model is also outperforming the benchmark model at a forecast horizon of 256 days. For the L and S train samples, improvement is also seen at the 1 day horizon. Whereas by using monthly sampled data, the GBM model shows significant improvements forecasting at the 3 and 12 months forecast horizons. Neither using daily nor monthly sampled prices resulted in higher success rates out-of-sample than what would be required to deem the model as significantly directional accurate.

As noted previously, averaging multiple simulations of this model reduces the stochastic behaviour of the predicted price path. If we were to choose the simulation with best out-of-sample accuracy the model follows the behaviour seen in crude oil prices quite well. However, this would not be the case of any true forecast, as the prices out-of-sample would have had to be known to the model. To show the potential of the model if the correct simulation was chosen, and the approach of averaging multiple simulations would not be required, the model shows some promising trends and shocks. Just for visual ratification of model behaviour, the best simulated price path is seen in Figure 5.2 for a forecast horizon of 256 days. In addition, if the result from the best simulation was to be used, it only shows a reduction in the RMSE loss function of 22.8% compared to the No-Change model at the selected forecast horizon. Considering that the prices of the forecast horizon would have had to be known for the model to yield this good simulation results, it is not a justified large reduction. Especially as the true forecast of the mean simulated path also yields a reduction of 17.5% compared to the benchmark model at this horizon. The major difference between the best simulation and the 'on average' simulation is how well the best simulation follows the general trend of actual prices in the forecast horizon. This is often considered as more crucial to an organization for planning and strategic reasons than an overall smaller loss metric score.

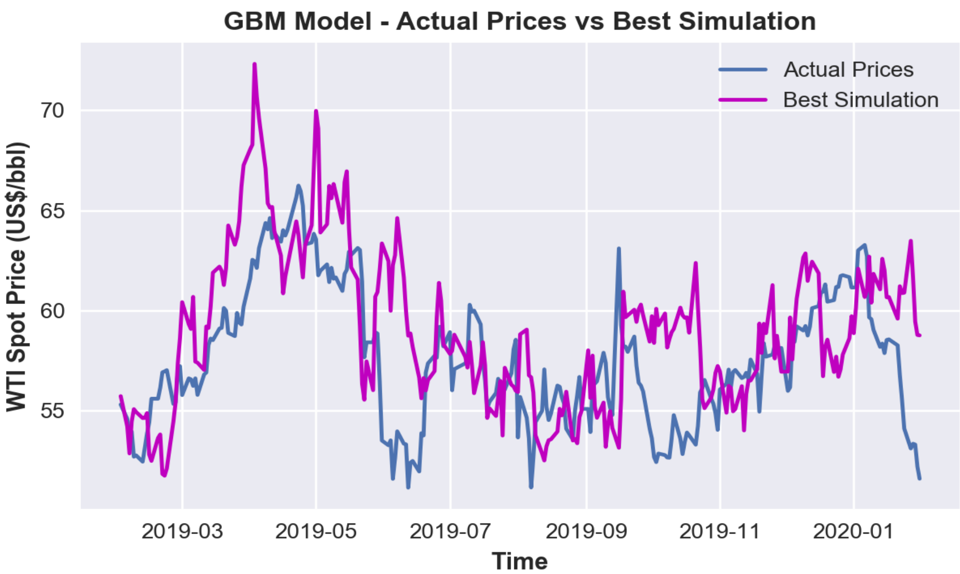


Figure 5.2: *Best simulated price path of the Geometric Brownian Motion model with daily frequented prices, the Long train sample and for a forecast horizon of 256 days.*

5.5 Mean Reverting Model

The theoretical assumptions underlying the MR model are sound from an economical perspective. Looking at historical crude oil prices, one may argue that the main feature of an MR process is observed; that prices tend to revert to a long-term equilibrium price level. This mean reversion feature is supported by e.g. Pindyck and Rubinfeld (1991) and Routledge et al. (2000). Pindyck and Rubinfeld examined over 100 years of oil price data and found a slow mean reversion, and a Dickey-Fuller unit root test was used to reject the hypothesis of being a simple random walk process. The MR model also describes the attraction to revert back as increasingly strong as current price levels move further away from the long-term level. Being a stochastic model with a diffusion term also incorporates the randomness expected from price shocks over time. However, in the same manner as the GBM model, this complicates the evaluation of its performance and makes it far less consistent in terms of forecasting. The GBM model has often been useful when making predictions about stock price fluctuations, while the MR model has been the natural choice for modelling commodity markets which have a strong reversion mechanism due to business cycles and seasonal effects influencing the demand and supply sides. Basic microeconomic theory states that, in the long-run, the price of a commodity is bounded to its long-run marginal production cost or "in case of a cartelized commodity like crude oil, the long-run profit-maximization price sought by cartel managers" (Laughton and Jacoby, 1995, p. 188).

Parameters of the MR model are estimated on the basis of historical oil prices, hence the different train samples yield great variations among the parameters and thus, also the forecasted prices. This is one of the reasons why practitioners and academics often rely on the most recent data to make model calibrations for forecasting. This becomes evident from the forecasts results in Tables A.5 and A.12, which is for daily and monthly oil prices, respectively. Also, the long-term equilibrium price depends greatly on which historical oil prices are included. For example, employing the L train sample returns a long-term price of $39.88, whereas the S train sample yields a price of $70.76. Using the R train sample and daily prices, the MR model outperforms the benchmark model for all forecast horizons, except for the 1 day horizon. For the L and M train samples, the model is significantly better than the benchmark at horizons of 64 and 256 days. While using the S train sample, the model is only significantly better at the 64 days forecast horizon. When using monthly prices and the three longest train samples, the model is generally outperforming the benchmark model at horizons of

3 and 12 months. Whereas using the R train sample, this model is superior to the benchmark model at all forecast horizons. In similar ways as the other models did not show particularly evidence of being directional accurate, the MR model is neither very directional accurate. However, at certain forecast horizons and train samples, the model has on average slightly better success rates.

Forecasting with the MR model yield different results for each simulation, for the same reasons as the GBM model does. Hence, the forecast results are obtained by multiple simulations and then using the 'on average' forecast performance as the representative for the model. This eliminates much of the random price shock's magnitudes. Although looking at the best simulated out-of-sample price path in Figure 5.3, the model shows both the effect of mean reversion and that it has large price fluctuations. This result should not be deemed as very representative of the model's performance in general, as this is just the best (out of thousand simulations) price path. The best simulation is only presented to show the model's stochastic behaviour when simulated once.

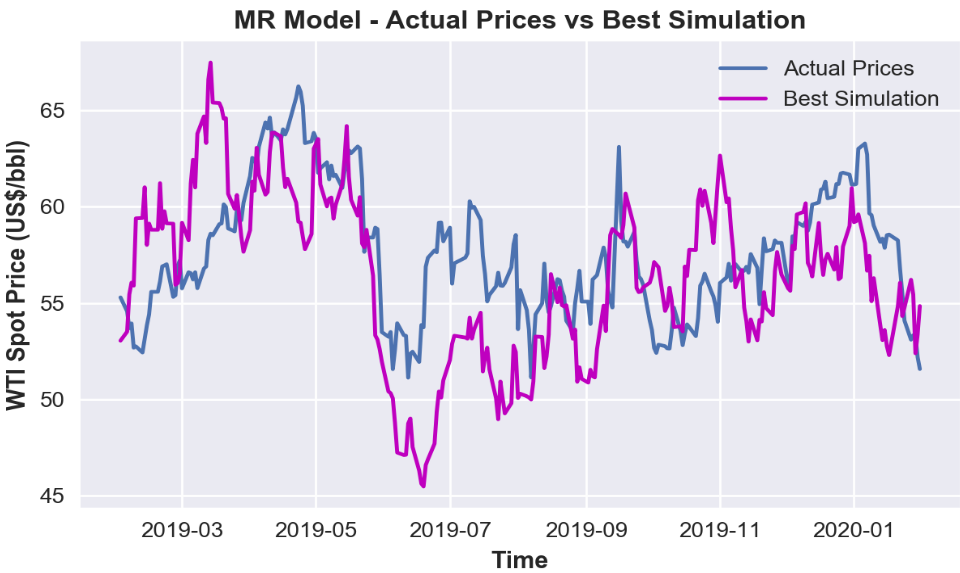


Figure 5.3: *Best simulated price path of the Mean Reverting model with daily frequented prices, the Long train sample and for a forecast horizon of 256 days.*

6 Conclusion

Despite the vast number of papers dealing with the task of forecasting crude oil prices, it is problematic to reconcile the seemingly conflicting results in the literature. The problem is not only due to the definition of the oil price variable, but whether the oil price is expressed in nominal or real terms, what horizons are chosen for model estimation and evaluation, whether the analysis is conducted in-sample or out-of-sample, how forecast accuracy is evaluated, whether tests of statistical significance is included or not, whether the methods are parametric or nonparametric. The most prevalent problem in the literature is that forecast results are sensitive to the choice of sample period and the sample split point. The price of oil has been proven to be predictable in population, which is consistent with economic theory. However, this does not need to translate into out-of-sample forecast accuracy which has been claimed to be inherently unpredictable in the literature (Alquist et al., 2011).

In Tables 6.1 and 6.2 the forecast results provided in Appendix A are summarized for daily and monthly sampled prices, respectively. For each table the best forecasting model is listed for all the considered forecast horizons, train sample lengths and various loss functions employed. The results show the pivotal effect the train sample length has on the forecast outcome. For daily prices and the L train sample, the GBM and No-Change models appears as the dominant forecasting models across the different horizons. However, when utilizing more recent price data in the R train sample, the MR model appear as the most superior model. The forecast performance of the most parsimonious No-Change model typically dominates in two out of the six forecast horizons for both daily and monthly prices. This indicates that, although other models such as the GBM and MR models are more complex in terms of economical underlying reasoning than the No-Change model, they do not generally outperform the forecast accuracy of this model. Considering all train samples the RW with Drift, ARIMA and AR(3) models does not seem to perform consistently well across the different forecast horizons. Notwithstanding, the results show that there is no unifying conclusion on which model provides the best forecasts, but rather results implying that the choice of train sample and forecast horizon has the greatest effect on the preferred model. Whereas for monthly prices, the No-Change model is typically superior at short forecast horizons, while both futures-based models show best performances at the longest horizon. In a similar manner as for daily prices, the GBM and MR models appear to perform better than the competing models when only recent market information is utilized. This may be due to better and less biased parameter estimates of the models that better reflect the current state of the crude oil market.

Table 6.1: **Daily Forecast Results**

This table reports the models providing best forecast results using daily sampled oil prices across the various train samples, forecast horizons and loss functions employed. In situations where two models obtain equal results, both models are listed.

Forecast Horizon	Loss Function				
	ME	**RMSE**	**QLIKE**	**L4**	**L5**
			Train Sample: L		
1	GBM	GBM	GBM	GBM	GBM
5	MR	No-Change	No-Change	No-Change	No-Change
22	No-Change	No-Change	No-Change	No-Change	No-Change
64	GBM	GBM	GBM	GBM	GBM
130	No-Change	ARIMA	ARIMA	ARIMA	ARIMA
256	GBM	GBM	GBM	GBM	GBM
			Train Sample: M		
1	MR	MR	MR/No-Change	MR	MR
5	MR	No-Change	No-Change	No-Change	MR/No-Change
22	AR(3)	AR(3)	AR(3)	AR(3)	AR(3)
64	MR	MR	MR	MR	MR
130	No-Change	No-Change	No-Change	No-Change	No-Change
256	GBM	AR(3)	MR	AR(3)	GBM/AR(3)
			Train Sample: S		
1	ARIMA	ARIMA	ARIMA/GBM	ARIMA	ARIMA/GBM
5	MR	MR	MR	MR	MR
22	No-Change	No-Change	No-Change	No-Change	No-Change
64	MR	MR	MR	MR	MR
130	No-Change	No-Change	No-Change	No-Change	No-Change
256	GBM	GBM	GBM	GBM	GBM
			Train Sample: R		
1	No-Change	No-Change	No-Change	No-Change	No-Change
5	MR	MR	MR	MR	MR
22	MR	MR	MR	MR	MR
64	GBM	GBM	GBM	GBM	GBM
130	MR	MR	MR	MR	MR
256	RW Drift/ ARIMA	RW Drift/ ARIMA	RW Drift/ ARIMA	RW Drift/ ARIMA	RW Drift/ ARIMA

Table 6.2: **Monthly Forecast Results**

This table reports the models providing best forecast results using monthly sampled oil prices across the various train samples, forecast horizons and loss functions employed. In situations where two models obtain equal results, both models are listed.

Forecast	**Loss Function**				
Horizon	**ME**	**RMSE**	**QLIKE**	**L4**	**L5**
			Train Sample: L		
1	No-Change	No-Change	No-Change	No-Change	No-Change
3	GBM	GBM	GBM	GBM	GBM
6	No-Change	No-Change	No-Change	No-Change	No-Change
9	Spread	Futures	Futures/Spread	Spread	Spread
12	ARIMA	MR	MR	MR	MR
24	ARIMA	Spread	Spread	Spread	Futures/Spread
			Train Sample: M		
1	No-Change	No-Change	No-Change	No-Change	No-Change
3	MR	MR	MR	MR	MR
6	No-Change	No-Change	No-Change	No-Change	No-Change
9	Spread	Spread	Futures/Spread	Spread	Spread
12	ARIMA	AR(3)	AR(3)	AR(3)	AR(3)
24	Futures	Spread	Spread	Spread	Futures/Spread
			Train Sample: S		
1	No-Change	No-Change	No-Change	No-Change	No-Change
3	MR	MR	MR	MR	MR
6	No-Change	No-Change	No-Change	No-Change	No-Change
9	Spread	Spread	Futures/Spread	Spread	Spread
12	MR	Futures	Futures	Futures	Futures
24	Futures	Spread	Spread	Spread	Futures/Spread
			Train Sample: R		
1	MR	MR	MR	MR	MR
3	GBM	GBM	GBM	GBM	GBM
6	No-Change	MR	MR	MR	MR
9	MR	MR	MR	MR	MR
12	GBM	AR(3)	AR(3)	AR(3)	AR(3)
24	Futures	Spread	Spread	Spread	Futures/Spread

Comparing forecast results of equivalent forecast horizons (e.g. the 22 days and 1 month horizons), shows that the best model in most cases is equal for the two frequented price data sets. This is particularly true for the three longest train samples, but for the most recent train sample, there is no linkage between the best selected models for the daily and monthly forecasts. Using monthly sampled prices implies a limited set of price observations compared to more frequently measured prices. Which in turn may induce bias and weak estimation of the models, and hence not comparable to the forecast results of the daily sampled price forecasts. An interesting point on the loss function's impact on the forecast results is that generally the best forecasting model using one loss function is also the best when another loss function is used. Although, the magnitude of the improvements among the various loss functions compared to the benchmark depends considerably on which one is utilized. But for the ME loss function, the superior model varies more, as this loss metric only measures the bias in the forecast and the positive and negative errors may cancel each other out, and thus results in contradicting and misleading conclusions. Among the models evaluated, none possess any significant evidence of being directionally accurate, with success rates rarely exceeding 0.5, and if so, they are rarely significant in terms of the PT-test statistic. Hence, adopting the considered deterministic and stochastic models to forecast the eccentric processes governing crude oil prices, generally does not produce adequately accurate forecasts, and suggest the use of more advanced and sophisticated models.

6.1 Future Research

Despite the broad interest forecasting the price of crude oil has received, there is still room for future improved understanding of this topic, and in the field of forecasting in general. With forecasting techniques continuously developing and becoming more complex, and better computational tools simultaneously allow for exploiting these techniques' potential to forecast the price of crude oil. Even the most complex models are very simplified versions of reality and do not truly reflect nor capture all influencing factors to the price formation of crude oil.

The application of deep learning machine models is a growing field that spans across numerous cases. These models provide powerful computational tools and algorithms that can learn from and make predictions on data, and may with future development serve as potentially strong models to forecast crude oil prices with acceptable accuracy. These models have the main advantage of being able to capture changing patterns of oil prices as the model continuously updates when new price data is available. Gao and Lei (2017) proposed a model based on ideas and tools from stream learning, a machine learning paradigm for analysis and inference of continuous flow of non-stationary data, to predict oil prices. The experimental results showed that the model outperformed three other popular oil price forecast models, both in terms of higher predictive accuracy and directional accuracy over a variety of forecast horizons. Other deep learning machine techniques include; Multilayer Perceptrons (MLP), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Deep Belief Networks (DBN) and Long Short Term Memory (LSTM) Networks to name a few in the literature. An interesting future work would be to compare the deep learning machine techniques that are used in various fields on the specific task of forecasting crude oil prices.

This thesis focuses on univariate oil price forecasting, hence another interesting research would be to perform a similar analysis while also considering other factors such as financial market indexes, economic growth, dollar exchange rate, demand and supply, global oil inventories, consumer price index, industrial production index, etc. The inclusion of such factors into deep learning machines techniques may also serve as a potentially interesting future research proposal.

Bibliography

Akpanta, A. and Okorie, I. (2014). Application of box-jenkins techniques in modelling and forecasting nigeria crude oil prices. *International Journal of statistics and applications*, 4(6):283–291.

Alquist, R. and Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied econometrics*, 25(4):539–573.

Alquist, R., Kilian, L., and Vigfusson, R. J. (2011). Forecasting the price of oil.

Bahar, A., Noh, N. M., and Zainuddin, Z. M. (2017). Forecasting model for crude oil price with structural break. *Malaysian Journal of Fundamental and Applied Sciences*, 13:421–424.

Bashiri Behmiri, N. and Pires Manso, J. R. (2013). Crude oil price forecasting techniques: a comprehensive review of literature. *Available at SSRN 2275428*.

Baumeister, C. and Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics*, 30(2):326–336.

Baumeister, C. and Kilian, L. (2014). What central bankers need to know about forecasting oil prices. *International Economic Review*, 55(3):869–889.

Baumeister, C., Kilian, L., and Zhou, X. (2013). Are product spreads useful for forecasting? an empirical evaluation of the verleger hypothesis.

Box, G. E. and Jenkins, G. M. (1976). Time series analysis: Forecasting and control, revised ed. *Holden-Day, San Francisco*.

Dbouk, W. and Jamali, I. (2018). Predicting daily oil prices: Linear and non-linear models. *Research in International Business and Finance*, 46:149–165.

Dbouk, W., Jamali, I., and Kryzanowski, L. (2016). Forecasting the libor-federal funds rate spread during and after the financial crisis. *Journal of Futures Markets*, 36(4):345–374.

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

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

Fleming, J. and Ostdiek, B. (1999). The impact of energy derivatives on the crude oil market. *Energy Economics*, 21(2):135–167.

Frey, G., Manera, M., Markandya, A., and Scarpa, E. (2009). Econometric models for oil price forecasting: A critical survey. In *CESifo Forum*, volume 10, pages 29–44. München: ifo Institut für Wirtschaftsforschung an der Universität München.

Gao, S. and Lei, Y. (2017). A new approach for crude oil price prediction based on stream learning. *Geoscience Frontiers*, 8(1):183–187.

Gospodinov, N. and Jamali, I. (2011). Risk premiums and predictive ability of bax futures. *Journal of Futures Markets*, 31(6):534–561.

Granger, C. W. (1999). Outline of forecast theory using generalized cost functions. *Spanish Economic Review*, 1(2):161–173.

Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007-08. Technical report, National Bureau of Economic Research.

Hansen, P. R. and Timmermann, A. (2012). Choice of sample split in out-of-sample forecast evaluation.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291.

IEA (2019). World Energy Outlook 2019. Paris, `https://www.iea.org/reports/world-energy-outlook-2019` Online; accessed 20-January-2020.

Laughton, D. G. and Jacoby, H. D. (1995). The effects of reversion on commodity projects of different length. In: TRIGEORGIS, L. (ed.). *Real options in capital investments: models, strategies, and applications.* p.185–205.

Lorentzen, S. and Sharma, D. (2015). Volatility forecast accuracy in Oslo Stock Exchange. Master Thesis, BI Norwegian Business School.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Pereboichuk, B. (2014). Modeling of crude oil prices with a special emphasis on macroeconomic factors.

Pesaran, M. H. and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4):461–465.

Pindyck, R. S. and Rubinfeld, D. L. (1991). Econometric models and economic forecast (3rd ed.). *New York: McGraw-Hill.*

Reeve, T. A. and Vigfusson, R. J. (2011). Evaluating the forecasting performance of commodity futures prices. *FRB International Finance Discussion Paper*, (1025).

Routledge, B. R., Seppi, D. J., and Spatt, C. S. (2000). Equilibrium forward curves for commodities. *The Journal of Finance*, 55(3):1297–1338.

Schwartz, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of finance*, 52(3):923–973.

Sigman, K. (2006). Geometric Brownian Motion. *USA: Columbia University*.

Stock, J. H. and Watson, M. W. (2015). *Introduction to econometrics*.

Taleb, N. N. (2007). The black swan: The impact of the highly improbable. Vol.2.

Vašíček, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

Appendix

A Forecast Results Tables

This appendix displays the forecast results of all selected models for the different train sample lengths Long, Medium, Short and Recent, which are denoted by L, M, S and R, respectively. Forecast results reported in tables of Sections A.1 Daily Frequented Oil Prices and A.2 Monthly Frequented Oil Prices are obtained by using, respectively, daily and monthly frequented crude oil prices. In all tables, a model's loss function's statistics are reported as ratios relative to the No-Change benchmark model, where its results are previously presented in Tables 5.1 and 5.2. This implies that, for ratios less than 1.0, the specific model performs better than the benchmark model, while for ratios larger than 1.0, the benchmark model is the superior one. For the ME loss function, which may produce both positive and negative values, ratios between −1.0 and 1.0 are indicative of less bias in favour of the competing model. Negative ratios represent opposite bias with respect to over- and underpredictions than the benchmark model, whereas positive ratios represent bias in the same direction as the benchmark model. The Success Rate (SR) of the models is included, which describes to what degree a model is able to predict the direction of the actual price change. As for the benchmark model, the SR is not possible to determine as there is no directional change in the predicted prices. The loss functions' ratios in the proceeding tables are **bold** for values improving the accuracy of the benchmark model. In addition, p-values for the Diebold-Mariano (DM) and Pesaran-Timmermann (PT) tests are included, where **bold** values are used for p-values < 0.05 and *italic* for p-values < 0.1.

A.1 Daily Frequented Oil Prices

Table A.1: **RW with Drift and Daily Prices**

Forecast	**Loss Function**						**DM-Test**	**PT-Test**
Horizon	**ME**	**RMSE**	**QLIKE**	**L4**	**L5**	**SR**	**p-value**	**p-value**
				Train Sample: L				
1	1.005	1.005	1.000	1.010	1.000	0.000	NA	NA
5	1.007	1.007	1.000	1.014	1.000	0.200	*0.087*	0.201
22	1.014	1.011	1.020	1.022	1.020	0.364	**0.001**	0.769
64	**0.974**	**0.981**	**0.962**	**0.964**	**0.963**	0.562	**0.000**	0.269
130	1.137	1.029	1.053	1.061	1.052	0.485	**0.000**	0.308
256	**0.882**	**0.943**	**0.880**	**0.894**	**0.887**	0.531	**0.000**	0.290
				Train Sample: M				
1	1.011	1.011	1.000	1.021	1.000	0.000	NA	NA
5	1.016	1.015	1.023	1.031	1.000	0.200	*0.087*	0.201
22	1.030	1.022	1.043	1.047	1.044	0.364	**0.001**	0.769
64	**0.944**	**0.960**	**0.915**	**0.924**	**0.922**	0.562	**0.000**	0.269
130	1.298	1.071	1.142	1.152	1.144	0.485	**0.000**	0.308
256	**0.734**	**0.884**	**0.766**	**0.789**	**0.774**	0.531	**0.000**	0.290
				Train Sample: S				
1	1.003	1.003	1.000	1.000	1.000	0.000	NA	NA
5	1.006	1.006	1.000	1.011	1.000	0.200	*0.087*	0.201
22	1.018	1.014	1.026	1.028	1.024	0.364	**0.001**	0.769
64	**0.978**	**0.984**	**0.967**	**0.969**	**0.968**	0.562	**0.000**	0.269
130	1.163	1.035	1.065	1.074	1.062	0.485	**0.000**	0.308
256	**0.895**	**0.949**	**0.893**	**0.905**	**0.896**	0.531	**0.000**	0.290
				Train Sample: R				
1	1.025	1.025	1.000	1.041	1.000	0.000	NA	NA
5	1.035	1.035	1.047	1.069	1.043	0.200	*0.086*	0.201
22	1.079	1.060	1.115	1.130	1.117	0.364	**0.000**	0.769
64	**0.865**	**0.906**	**0.810**	**0.827**	**0.817**	0.562	**0.000**	0.269
130	1.837	1.265	1.550	1.626	1.567	0.485	**0.000**	0.308
256	**0.248**	**0.803**	**0.623**	**0.657**	**0.633**	0.531	**0.000**	0.290

Table A.2: **ARIMA and Daily Prices**

Forecast Horizon	Loss Function ME	RMSE	QLIKE	L4	L5	SR	DM-Test p-value	PT-Test p-value
Train Sample: L								
1	1.041	1.041	1.000	1.082	1.000	0.000	NA	NA
5	1.143	1.092	1.163	1.194	1.174	0.600	0.259	0.698
22	1.028	1.025	1.049	1.052	1.049	0.318	**0.002**	0.126
64	**0.941**	**0.960**	**0.915**	**0.925**	**0.922**	0.516	**0.000**	0.801
130	1.044	**0.998**	**0.994**	**0.997**	**0.990**	0.477	0.801	0.602
256	**0.903**	**0.954**	**0.903**	**0.914**	**0.905**	0.504	**0.000**	0.927
Train Sample: M								
1	1.102	1.102	1.143	1.216	1.000	0.000	NA	NA
5	1.069	1.109	1.209	1.231	1.217	0.400	0.423	0.365
22	1.013	1.013	1.026	1.028	1.024	0.591	**0.008**	0.583
64	**0.947**	**0.962**	**0.921**	**0.928**	**0.927**	0.469	**0.000**	0.529
130	1.281	1.047	1.083	1.103	1.082	0.483	**0.000**	0.153
256	**0.726**	**0.880**	**0.758**	**0.783**	**0.769**	0.531	**0.000**	0.498
Train Sample: S								
1	**0.916**	**0.916**	**0.857**	**0.835**	**0.750**	1.000	NA	NA
5	1.108	1.094	1.186	1.199	1.174	0.200	**0.035**	0.171
22	1.042	1.024	1.046	1.051	1.044	0.500	**0.005**	0.804
64	**0.961**	**0.972**	**0.941**	**0.948**	**0.945**	0.469	**0.000**	0.564
130	1.118	1.021	1.041	1.044	1.041	0.469	**0.001**	0.490
256	**0.882**	**0.943**	**0.880**	**0.894**	**0.887**	0.500	**0.000**	0.967
Train Sample: R								
1	1.141	1.141	1.286	1.299	1.250	0.000	NA	NA
5	1.132	1.105	1.209	1.222	1.217	0.600	**0.012**	0.698
22	1.103	1.072	1.138	1.157	1.141	0.500	**0.000**	0.902
64	**0.918**	**0.945**	**0.887**	**0.896**	**0.890**	0.500	**0.000**	0.981
130	1.837	1.265	1.550	1.626	1.567	0.485	**0.000**	0.308
256	**0.248**	**0.803**	**0.623**	**0.657**	**0.633**	0.531	**0.000**	0.290

Table A.3: **AR(3) and Daily Prices**

Forecast Horizon	Loss Function ME	RMSE	QLIKE	L4	L5	SR	DM-Test p-value	PT-Test p-value
Train Sample: L								
1	1.095	1.095	1.143	1.196	1.000	0.000	NA	NA
5	1.075	1.061	1.116	1.127	1.130	0.400	**0.018**	0.698
22	1.024	1.016	1.029	1.032	1.029	0.318	**0.000**	0.236
64	**0.961**	**0.972**	**0.941**	**0.947**	**0.945**	0.547	**0.000**	0.892
130	1.121	1.023	1.047	1.049	1.041	0.492	**0.000**	1.000
256	**0.876**	**0.940**	**0.875**	**0.889**	**0.882**	0.535	**0.000**	0.134
Train Sample: M								
1	1.116	1.116	1.286	1.247	1.000	0.000	NA	NA
5	1.053	1.043	1.070	1.088	1.087	0.400	**0.017**	0.698
22	**0.157**	**0.728**	**0.553**	**0.518**	**0.541**	0.364	**0.034**	0.482
64	**0.935**	**0.953**	**0.903**	**0.912**	**0.908**	0.547	**0.000**	0.892
130	1.302	1.073	1.142	1.155	1.144	0.492	**0.000**	1.000
256	**0.719**	**0.877**	**0.753**	**0.777**	**0.760**	0.535	**0.000**	0.134
Train Sample: S								
1	1.064	1.064	1.143	1.134	1.000	0.000	NA	NA
5	1.065	1.055	1.093	1.113	1.087	0.200	**0.026**	0.201
22	1.025	1.017	1.034	1.036	1.034	0.364	**0.000**	0.769
64	**0.968**	**0.977**	**0.951**	**0.956**	**0.954**	0.562	**0.000**	0.269
130	1.154	1.032	1.065	1.068	1.062	0.492	**0.000**	1.000
256	**0.894**	**0.949**	**0.893**	**0.904**	**0.896**	0.535	**0.000**	0.134
Train Sample: R								
1	1.084	1.084	1.143	1.175	1.000	0.000	NA	NA
5	1.225	1.195	1.395	1.432	1.391	0.200	**0.032**	0.201
22	1.098	1.070	1.135	1.152	1.137	0.364	**0.000**	0.769
64	**0.834**	**0.885**	**0.769**	**0.790**	**0.780**	0.562	**0.000**	0.269
130	1.801	1.249	1.515	1.584	1.536	0.485	**0.000**	0.308
256	**0.261**	**0.806**	**0.628**	**0.661**	**0.638**	0.531	**0.000**	0.290

Table A.4: **GBM and Daily Prices**

Forecast Horizon	Loss Function ME	RMSE	QLIKE	L4	L5	SR	DM-Test p-value	PT-Test p-value
Train Sample: L								
1	**0.907**	**0.907**	**0.857**	**0.825**	**0.750**	1.000	NA	NA
5	1.036	1.043	1.070	1.088	1.087	0.400	0.168	0.698
22	1.126	1.097	1.186	1.214	1.195	0.409	**0.001**	0.906
64	**0.794**	**0.851**	**0.708**	**0.732**	**0.720**	0.562	**0.000**	0.435
130	1.679	1.201	1.408	1.460	1.423	0.492	**0.000**	0.897
256	**0.153**	**0.825**	**0.654**	**0.694**	**0.670**	0.477	**0.000**	0.240
Train Sample: M								
1	1.051	1.051	1.143	1.103	1.000	0.000	NA	NA
5	1.034	1.039	1.070	1.079	1.043	0.000	0.203	0.668
22	1.058	1.049	1.092	1.104	1.093	0.455	**0.001**	0.965
64	**0.832**	**0.888**	**0.777**	**0.795**	**0.784**	0.562	**0.000**	0.498
130	2.434	1.546	2.237	2.472	2.299	0.446	**0.000**	0.156
256	**-0.039**	**0.883**	**0.743**	**0.798**	**0.760**	0.547	**0.001**	0.219
Train Sample: S								
1	**0.962**	**0.962**	**0.857**	**0.928**	**0.750**	1.000	NA	NA
5	1.098	1.092	1.163	1.194	1.174	0.000	*0.068*	0.668
22	1.056	1.032	1.063	1.069	1.063	0.409	**0.000**	0.440
64	**0.861**	**0.906**	**0.810**	**0.828**	**0.817**	0.547	**0.000**	0.585
130	1.615	1.182	1.373	1.411	1.381	0.454	**0.000**	0.291
256	**0.227**	**0.813**	**0.639**	**0.672**	**0.652**	0.500	**0.000**	0.779
Train Sample: R								
1	1.056	1.056	1.143	1.113	1.000	0.000	NA	NA
5	1.034	1.030	1.047	1.062	1.043	0.600	*0.054*	0.698
22	1.157	1.134	1.261	1.300	1.268	0.318	**0.001**	0.232
64	**0.746**	**0.840**	**0.690**	**0.714**	**0.697**	0.500	**0.000**	0.475
130	2.481	1.571	2.308	2.558	2.381	0.469	**0.000**	0.435
256	**-0.440**	1.028	**0.964**	1.108	1.009	0.543	0.540	0.288

Table A.5: **MR and Daily Prices**

Forecast Horizon	Loss Function ME	RMSE	QLIKE	L4	L5	SR	DM-Test p-value	PT-Test p-value
Train Sample: L								
1	1.072	1.072	1.143	1.144	1.000	0.000	NA	NA
5	**0.996**	1.002	1.000	1.003	1.000	0.600	0.841	0.365
22	1.001	1.004	1.009	1.008	1.005	0.500	**0.020**	0.902
64	**0.975**	**0.985**	**0.969**	**0.972**	**0.972**	0.484	**0.000**	0.819
130	1.264	1.078	1.154	1.167	1.155	0.554	**0.000**	0.218
256	**0.916**	**0.973**	**0.941**	**0.948**	**0.946**	0.535	**0.000**	0.280
Train Sample: M								
1	**0.972**	**0.972**	1.000	**0.938**	**0.750**	1.000	NA	NA
5	**0.999**	1.014	1.023	1.028	1.000	0.200	0.490	0.201
22	1.092	1.069	1.132	1.149	1.137	0.364	**0.001**	0.363
64	**0.704**	**0.819**	**0.654**	**0.680**	**0.665**	0.484	**0.000**	0.290
130	1.847	1.287	1.604	1.683	1.629	0.492	**0.000**	0.903
256	**-0.270**	**0.888**	**0.738**	**0.815**	**0.765**	0.488	**0.008**	0.367
Train Sample: S								
1	1.015	1.015	1.000	1.031	1.000	0.000	NA	NA
5	**0.774**	**0.859**	**0.744**	**0.735**	**0.739**	0.200	0.137	0.201
22	1.161	1.114	1.221	1.253	1.229	0.273	**0.000**	*0.083*
64	**0.544**	**0.712**	**0.487**	**0.518**	**0.500**	0.562	**0.000**	0.505
130	2.261	1.455	2.006	2.176	2.052	0.508	**0.000**	0.743
256	**-0.549**	1.033	**0.967**	1.120	1.014	0.516	0.516	0.937
Train Sample: R								
1	1.090	1.090	1.143	1.186	1.000	0.000	NA	NA
5	**0.707**	**0.794**	**0.628**	**0.627**	**0.609**	0.200	0.137	0.201
22	**0.917**	**0.943**	**0.897**	**0.884**	**0.893**	0.591	**0.001**	0.663
64	**0.868**	**0.902**	**0.803**	**0.820**	**0.812**	0.500	**0.000**	0.825
130	**0.937**	**0.980**	**0.959**	**0.959**	**0.959**	0.577	**0.000**	*0.082*
256	**0.590**	**0.830**	**0.669**	**0.700**	**0.679**	0.531	**0.000**	0.375

A.2 Monthly Frequented Oil Prices

Table A.6: **RW with Drift and Monthly Prices**

Forecast Horizon	**Loss Function** **ME**	**RMSE**	**QLIKE**	**L4**	**L5**	**SR**	**DM-Test p-value**	**PT-Test p-value**
Train Sample: L								
1	1.041	1.041	1.087	1.086	1.073	0.000	NA	NA
3	**0.964**	**0.966**	**0.927**	**0.935**	**0.931**	0.667	*0.091*	0.400
6	1.473	1.014	1.031	1.028	1.028	0.500	0.826	0.435
9	1.075	1.069	1.129	1.149	1.135	0.444	**0.001**	0.344
12	**0.924**	**0.940**	**0.872**	**0.889**	**0.878**	0.583	**0.000**	0.299
24	1.442	1.149	1.286	1.337	1.301	0.583	**0.001**	0.425
Train Sample: M								
1	1.085	1.085	1.174	1.180	1.171	0.000	NA	NA
3	**0.924**	**0.929**	**0.854**	**0.866**	**0.857**	0.667	*0.090*	0.400
6	1.992	1.058	1.109	1.122	1.139	0.500	0.641	0.435
9	1.156	1.144	1.277	1.326	1.294	0.444	**0.001**	0.344
12	**0.835**	**0.875**	**0.745**	**0.775**	**0.756**	0.583	**0.000**	0.299
24	1.958	1.344	1.694	1.868	1.749	0.583	**0.000**	0.425
Train Sample: S								
1	1.063	1.063	1.130	1.132	1.122	0.000	NA	NA
3	**0.955**	**0.958**	**0.911**	**0.920**	**0.914**	0.667	*0.091*	0.400
6	1.687	1.029	1.063	1.059	1.056	0.500	0.749	0.435
9	1.132	1.121	1.231	1.271	1.244	0.444	**0.001**	0.344
12	**0.915**	**0.933**	**0.858**	**0.877**	**0.866**	0.583	**0.000**	0.299
24	1.851	1.302	1.604	1.746	1.647	0.583	**0.000**	0.425
Train Sample: R								
1	1.273	1.273	1.609	1.631	1.610	0.000	NA	NA
3	**0.766**	**0.785**	**0.600**	**0.624**	**0.606**	0.667	*0.090*	0.400
6	4.384	1.543	2.281	2.435	2.361	0.500	0.126	0.435
9	1.603	1.583	2.263	2.644	2.382	0.444	**0.001**	0.344
12	**0.420**	**0.654**	**0.401**	**0.442**	**0.414**	0.583	**0.001**	0.299
24	6.578	3.438	7.688	14.895	9.473	0.583	**0.000**	0.425

Table A.7: **Futures and Monthly Prices**

Forecast	**Loss Function**						**DM-Test**	**PT-Test**
Horizon	**ME**	**RMSE**	**QLIKE**	**L4**	**L5**	**SR**	**p-value**	**p-value**
				Train Samples: L, M, S and R				
1	1.536	1.536	2.304	2.393	2.317	0.000	NA	NA
3	**0.944**	**0.946**	**0.889**	**0.898**	**0.891**	1.000	*0.091*	0.168
6	2.745	1.325	1.719	1.774	1.750	0.500	0.197	0.435
9	**0.937**	**0.953**	**0.913**	**0.904**	**0.912**	0.333	*0.056*	0.244
12	**0.458**	**0.602**	**0.329**	**0.379**	**0.347**	0.500	**0.000**	1.000
24	**-0.174**	**0.729**	**0.537**	**0.529**	**0.532**	0.375	**0.005**	1.000

Table A.8: **Spread and Monthly Prices**

Forecast	**Loss Function**						**DM-Test**	**PT-Test**
Horizon	**ME**	**RMSE**	**QLIKE**	**L4**	**L5**	**SR**	**p-value**	**p-value**
				Train Samples: L, M, S and R				
1	1.531	1.531	2.290	2.376	2.317	0.000	NA	NA
3	**0.944**	**0.946**	**0.889**	**0.898**	**0.891**	1.000	*0.091*	0.168
6	2.728	1.321	1.703	1.762	1.750	0.500	0.199	0.435
9	**0.936**	**0.952**	**0.913**	**0.903**	**0.910**	0.333	*0.056*	0.243
12	**0.475**	**0.612**	**0.341**	**0.391**	**0.356**	0.500	**0.000**	1.000
24	**-0.230**	**0.728**	**0.536**	**0.528**	**0.532**	0.375	**0.006**	1.000

Table A.9: **ARIMA and Monthly Prices**

Forecast	Loss Function						DM-Test	PT-Test
Horizon	ME	RMSE	QLIKE	L4	L5	SR	p-value	p-value
Train Sample: L								
1	1.705	1.705	2.812	2.959	2.854	0.000	NA	NA
3	1.101	1.099	1.222	1.201	1.217	0.000	0.130	0.168
6	3.008	1.175	1.359	1.391	1.389	0.833	0.511	0.150
9	1.621	1.601	2.324	2.695	2.441	0.444	**0.000**	0.759
12	**-0.005**	**0.591**	**0.309**	**0.371**	**0.328**	0.583	**0.008**	0.501
24	**0.139**	**0.813**	**0.670**	**0.657**	**0.664**	0.458	*0.055*	0.829
Train Sample: M								
1	1.774	1.774	3.043	3.210	3.073	0.000	NA	NA
3	1.053	1.055	1.121	1.110	1.114	0.000	0.276	0.168
6	4.676	1.629	2.531	2.721	2.611	0.500	0.113	0.435
9	1.656	1.624	2.387	2.779	2.509	0.333	**0.000**	0.344
12	**-0.025**	**0.638**	**0.362**	**0.431**	**0.383**	0.583	**0.015**	0.299
24	**0.852**	**0.936**	**0.881**	**0.873**	**0.878**	0.542	0.150	0.829
Train Sample: S								
1	1.331	1.331	1.754	1.788	1.756	0.000	NA	NA
3	1.146	1.137	1.314	1.282	1.303	0.000	**0.047**	0.168
6	2.810	1.235	1.500	1.538	1.528	0.500	0.288	0.435
9	1.375	1.349	1.718	1.875	1.768	0.444	**0.000**	0.344
12	**0.834**	**0.868**	**0.731**	**0.764**	**0.742**	0.583	**0.000**	0.299
24	2.441	1.473	1.985	2.274	2.072	0.583	**0.000**	0.427
Train Sample: R								
1	1.273	1.273	1.609	1.631	1.610	0.000	NA	NA
3	**0.766**	**0.785**	**0.600**	**0.624**	**0.606**	0.667	*0.090*	0.400
6	4.384	1.543	2.281	2.435	2.361	0.500	0.126	0.435
9	3.187	3.134	7.149	11.624	8.360	0.444	**0.000**	0.759
12	**0.420**	**0.654**	**0.401**	**0.442**	**0.414**	0.583	**0.001**	0.299
24	3.783	2.247	3.960	5.746	4.455	0.583	**0.000**	0.872

Table A.10: **AR(3) and Monthly Prices**

Forecast Horizon	Loss Function ME	RMSE	QLIKE	L4	L5	SR	DM-Test p-value	PT-Test p-value
Train Sample: L								
1	1.729	1.729	2.899	3.043	2.927	0.000	NA	NA
3	1.239	1.234	1.565	1.502	1.543	0.000	*0.099*	0.168
6	4.267	1.571	2.375	2.516	2.444	0.500	0.142	0.435
9	1.443	1.414	1.868	2.069	1.934	0.444	**0.000**	0.628
12	**0.466**	**0.591**	**0.318**	**0.366**	**0.333**	0.583	**0.000**	0.299
24	2.250	1.360	1.729	1.913	1.786	0.625	**0.001**	0.700
Train Sample: M								
1	1.808	1.808	3.159	3.336	3.195	0.000	NA	NA
3	1.198	1.197	1.467	1.416	1.446	0.000	0.116	0.168
6	4.861	1.708	2.781	2.991	2.861	0.500	0.103	0.435
9	1.542	1.508	2.092	2.371	2.182	0.444	**0.000**	0.344
12	**0.323**	**0.523**	**0.244**	**0.289**	**0.258**	0.583	**0.001**	0.299
24	2.833	1.607	2.301	2.742	2.435	0.583	**0.000**	0.427
Train Sample: S								
1	1.425	1.425	2.000	2.053	2.000	0.000	NA	NA
3	1.109	1.106	1.238	1.216	1.229	0.000	0.104	0.168
6	3.709	1.448	2.031	2.129	2.083	0.500	0.165	0.435
9	1.236	1.215	1.423	1.503	1.450	0.333	**0.000**	0.167
12	**0.624**	**0.703**	**0.463**	**0.511**	**0.478**	0.583	**0.000**	0.299
24	2.098	1.364	1.737	1.926	1.796	0.542	**0.000**	0.255
Train Sample: R								
1	1.592	1.592	2.464	2.573	2.488	0.000	NA	NA
3	**0.748**	**0.784**	**0.600**	**0.622**	**0.606**	0.667	*0.066*	0.400
6	6.698	2.211	4.500	5.095	4.722	0.500	**0.029**	0.435
9	1.566	1.541	2.162	2.494	2.265	0.444	**0.001**	0.344
12	**0.281**	**0.593**	**0.323**	**0.367**	**0.337**	0.583	**0.001**	0.299
24	6.453	3.391	7.527	14.444	9.241	0.583	**0.000**	0.427

Table A.11: **GBM and Monthly Prices**

Forecast Horizon	Loss Function ME	RMSE	QLIKE	L4	L5	SR	DM-Test p-value	PT-Test p-value
Train Sample: L								
1	1.172	1.172	1.362	1.379	1.366	0.000	NA	NA
3	**0.851**	**0.861**	**0.730**	**0.747**	**0.737**	0.667	0.116	0.400
6	3.128	1.232	1.484	1.529	1.528	0.500	0.314	0.435
9	1.268	1.251	1.497	1.602	1.533	0.444	**0.001**	0.344
12	**0.677**	**0.765**	**0.558**	**0.599**	**0.572**	0.583	**0.000**	0.299
24	2.945	1.780	2.723	3.427	2.930	0.583	**0.000**	0.427
Train Sample: M								
1	1.215	1.215	1.464	1.484	1.463	0.000	NA	NA
3	**0.715**	**0.738**	**0.530**	**0.553**	**0.537**	0.667	*0.063*	0.113
6	3.998	1.454	2.047	2.152	2.083	0.500	0.155	0.435
9	1.510	1.490	2.038	2.321	2.128	0.444	**0.001**	0.344
12	**0.472**	**0.670**	**0.421**	**0.463**	**0.435**	0.583	**0.001**	0.299
24	3.965	2.233	3.938	5.647	4.415	0.583	**0.000**	0.425
Train Sample: S								
1	1.121	1.121	1.246	1.261	1.244	0.000	NA	NA
3	**0.825**	**0.832**	**0.679**	**0.700**	**0.686**	1.000	*0.085*	0.168
6	2.957	1.200	1.422	1.451	1.444	0.500	0.354	0.435
9	1.341	1.322	1.653	1.798	1.699	0.444	**0.000**	0.344
12	**0.744**	**0.815**	**0.641**	**0.677**	**0.653**	0.583	**0.000**	0.299
24	3.187	1.856	2.921	3.755	3.164	0.583	**0.000**	0.425
Train Sample: R								
1	1.435	1.435	2.014	2.082	2.024	0.000	NA	NA
3	**0.679**	**0.702**	**0.476**	**0.502**	**0.486**	0.667	*0.098*	0.400
6	6.967	2.317	4.844	5.656	5.139	0.500	**0.041**	0.435
9	2.011	2.011	3.386	4.444	3.697	0.444	**0.002**	0.344
12	**-0.006**	**0.688**	**0.417**	**0.503**	**0.445**	0.583	**0.012**	0.299
24	12.880	6.693	18.977	72.704	28.582	0.583	**0.000**	0.427

Table A.12: **MR and Monthly Prices**

Forecast Horizon	Loss Function ME	RMSE	QLIKE	L4	L5	SR	DM-Test p-value	PT-Test p-value
Train Sample: L								
1	1.033	1.033	1.072	1.068	1.073	0.000	NA	NA
3	**0.933**	**0.940**	**0.876**	**0.887**	**0.880**	0.667	**0.048**	0.400
6	1.989	1.068	1.141	1.144	1.139	0.500	0.571	0.435
9	1.071	1.068	1.127	1.147	1.135	0.333	**0.006**	0.759
12	**0.289**	**0.501**	**0.222**	**0.266**	**0.237**	0.583	**0.027**	0.616
24	**0.888**	1.010	1.020	1.021	1.020	0.750	**0.024**	**0.034**
Train Sample: M								
1	1.251	1.251	1.551	1.576	1.561	0.000	NA	NA
3	**0.701**	**0.734**	**0.521**	**0.546**	**0.531**	0.667	*0.095*	0.400
6	4.600	1.608	2.469	2.652	2.556	0.500	0.114	0.435
9	1.282	1.262	1.522	1.630	1.557	0.444	**0.000**	0.344
12	**0.305**	**0.614**	**0.349**	**0.392**	**0.364**	0.583	**0.001**	0.299
24	2.515	1.538	2.135	2.493	2.244	0.667	**0.000**	0.266
Train Sample: S								
1	1.279	1.279	1.623	1.648	1.610	0.000	NA	NA
3	**0.645**	**0.675**	**0.438**	**0.464**	**0.446**	0.667	*0.097*	0.400
6	5.261	1.793	3.031	3.312	3.139	0.500	*0.069*	0.435
9	1.402	1.386	1.795	1.987	1.858	0.444	**0.001**	0.344
12	**0.202**	**0.610**	**0.340**	**0.389**	**0.356**	0.583	**0.002**	0.299
24	3.197	1.813	2.814	3.561	3.035	0.583	**0.000**	0.872
Train Sample: R								
1	**0.951**	**0.951**	**0.913**	**0.903**	**0.902**	1.000	NA	NA
3	**0.737**	**0.765**	**0.568**	**0.594**	**0.577**	0.667	*0.071*	0.400
6	1.263	**0.943**	**0.891**	**0.890**	**0.889**	0.333	0.209	0.237
9	**0.684**	**0.725**	**0.551**	**0.513**	**0.538**	0.556	**0.000**	0.344
12	**0.474**	**0.640**	**0.380**	**0.426**	**0.395**	0.583	**0.000**	0.501
24	-1.252	**0.965**	**0.953**	**0.922**	**0.940**	0.458	0.842	0.569