aurum

AURUM MÜHENDİSLİK SİSTEMLERİ VE MİMARLIK DERGİSİ
AURUM JOURNAL OF ENGINEERING SYSTEMS AND ARCHITECTURE

Cilt 9, Sayı 1 | Yaz 2025
Volume 9, No 1 | Summer 2025

West texas Crude oil forcasting using ARIMA and Holt winter models using R

Ellaf saleh Rustam[1]

[1]Industrial engineering, Altinbaş university, İstanbul,Türkiye
ellafrostam@gmail.com
(https://orcid.org/ 0000-0002-7960-3861)

Received: 24.10.2020 | Research Article
Accepted: 26.06.2025 | pp.141-149
Published: 30.06.2025
*Corresponding author | DOI: 10.53600/ajesa.799705

Abstract

Investors have been hoping to benefit from the capital markets for hundreds of years by seeking to foresee their potential movements. To this end, various strategies and techniques designed to help market participants produce income have been created. This research explores the efficacy and the feasibility of technological analysis and time series models for predicting the movements of the crude oil price West Texas Intermediate Views are split on the utility of technical analysis. The presence in stock markets is almost omnipresent and is commonly used by experienced and novice traders. Time series were forecast using Auto-Regressive Moving integrated Average (ARIMA) and Holt-winter method.

Keywords: Forecasting, ARIMA, R programming, Holt-winters, west Texas crude oil .

ARIMA ve Holt-Winters modelleri kullanılarak R programı ile Batı Teksas ham petrolü fiyat tahmini

Özet

Yatırımcılar, yüzlerce yıldır sermaye piyasalarının potansiyel hareketlerini öngörmeye çalışarak bu piyasalardan fayda sağlamayı ummuşlardır. Bu amaçla, piyasa katılımcılarının gelir elde etmesine yardımcı olmak üzere çeşitli stratejiler ve teknikler geliştirilmiştir. Bu araştırma, Batı Teksas Petrolü (West Texas Intermediate) fiyat hareketlerinin tahmin edilmesinde teknik analiz ve zaman serisi modellerinin etkinliğini ve uygulanabilirliğini incelemektedir. Teknik analiz konusunda görüşler farklılık göstermektedir. Hisse senedi piyasalarında neredeyse her yerde bulunmaktadır ve hem deneyimli hem de yeni başlayan yatırımcılar tarafından yaygın olarak kullanılmaktadır. Zaman serileri, Otoregresif Hareketli Ortalama (ARIMA) ve Holt-Winters yöntemi kullanılarak tahmin edilmiştir.

Anahtar kelimeler: Tahmin, R programlama, Holt-winters

1. INTRODUCTION

Oil is a commodity that unlike any other, affects everyone's daily life in a plethora of ways. Oil prices and availability affects transportation be it every day driving or flights, as well as economic growth, since goods have to be transported and oil is used almost everywhere in the secondary industry. Presently economic activity and oil consumption are fairly correlated. It is not only for the aforementioned reasons that the interest in oil prices and in particular, in the ability of being able to forecast oil prices is crucial [1].

The main approaches to predict the oil price is a time series approach, which includes linear and nonlinear time series analysis. Modelling the price of oil is difficult because of the changing variability over time. The wide range of oil production cost results in increased medium-term volatility. Therefore shifts in demand will cause a much more drastic change in price than before [2].

A time series plot of the deflated price for crude oil Emphasis has to be put on the end of 2017 and the beginning of 2019. A price change like this has never been seen before and provides quite a challenge to the ambitious statistician. Because such price movements might seem unpredictable, it is a challenge to find a model that performs relatively well, when being confronted with such obstacles [3]

Modeling the price of oil is difficult, because of many fluctuating variables over time. oil price can dramatically change over a short time, which makes it very difficult to predict. Oil demand and supply are quite inelastic in short time, meaning when the demand for oil exceeds supply, price will rise extremely high. On the other hand oil price is affected heavily by political turbulences. Such as 1999 the Asian Financial Crisis and Iraq deciding to increase oil production, which caused oil prices to reach a bottom [4]. More recently and during 2008 financial crisis, market volatility sky rocketed to $147 and dropped to below $40 level less than a year, showing how difficult predicting oil price can be [5]. In general there are three different approaches in forecasting oil prices: long term, medium term and short term [6].

The computer-based model is based on macroeconomic and financial models and also includes many inputs and as¬sumptions. It consists of many integrated modules that interact with each other as part of an equilibrium calculation [7].

Medium term oil forecasting models focus on few years window. Central banks also use medium term models of oil price forecasting for Macroeconomic decision making .Baumeister and Kilian research has shown that the real price forecasts are more accurate than the forecasts based on future prices. Quarterly vector autoregressive models forecast estimate on monthly data. There are different approaches for the quar¬terly forecast. One way is to forecast the monthly real price of oil for each month and then convert them to quarterly average based on [8].

In the literatures, there are several different models used to predict and forecast short term oil prices. Historically linear structural models have not performed well for oil price forecasting and nonlinear time series models have per¬formed much better in forecasting oil prices [4,9,10]. F. Basler examined a time series approach, which includes linear and nonlinear time series analysis and also structural models [4]. He compared linear ARIMA model and neural network autoregressive model for nonlinear time series analysis and confirmed that the nonlinear models forecasts perform the better and follow the volatility of the oil price. In another work, D. Lam modeled oil prices based on univariate time series using the Box-Jenkins methodology. Based on the ACF and PACF techniques ARIMA model was chosen, and followed by GARCH and APARCH as model residuals [9]. He also built a regression model to compare with his nonlinear model. The regression model was based one high explanatory variables, including production, consumption, net import, ending stock, refinery utilization rate, U.S. interest rate, NYMEX oil futures contract 4 and S&P 500 index. But at the end the conclusion was that GARCH and APARCH perform the best.

2.METHODOLOGY

The aim of this study was to predict market movement, not in terms of absolute values but more in terms of general direction. The trading algorithms developed in this study were based around a daily time period. The data used was daily data containing open, high, low and closing prices.

2.1 Data Collection and Quality

The data used in this study was freely collected from the Yahoo web site (www.yahoo.com). It is of high quality with no missing values and represents the opening, high, low and closing prices for each day that the particular market indice was open for trading.

Experimental results and graphs were produced with the open source programming language R version 3.0.2. For help in the creation and organization of the R code for this thesis the open-source development environment R Studio version 0.98.490 was used extensively.

2.2 Holt-winter Models

The single exponential smoothing extended to linear exponential smoothing for forecasting the data with trend. Winters (1960) extended the Holt's method and proposed the Additive and Multiplicative seasonality methods.

The Holt-Winters forecasting procedure has achieved great popularity in practical time series analysis because it is easy to use and understand, straightforward and computationally efficient

2.3 ARIMA Models

The use of Auto-Regressive Integrated Moving Average (ARIMA) models, was explored in order to forecast future prices for financial markets. The process of cutting an ARIMA model to a time series is quite challenging and involves the following general steps:

1. Plot the data to get a general feel for the time series and to establish if it is stationary.

2. Stabilize any variance in the data with a transformation process such as the Box- Cox method.

3. ARIMA models work with stationary data, so if necessary, take differences of the data until it is stationary.

4. Examine the auto-correlation and partial auto-correlation (ACF/PACF) plots in order to determine if an AR (p) or MA(q) model is appropriate.

5. Test the chosen model(s), using the AICc to determine if a better model is available.

6. Check the residuals from the best model by plotting the ACF, and doing a port- portmanteau test on them. If the results from these tests do not look like white noise, a modeled model may be required.

7. Finally, once the residuals have a similar pattern to white noise, the model can be used to generate forecasts. In recent years automatic forecasting algorithms have become available and are widely used .

These are necessary in a variety of circumstances, especially when organizations are faced with the need to repeatedly carry out a large number of forecasts and the human effort required renders manual means impractical.

The auto. ARIMA function found in R's \forecast" package is an example of an automatic algorithm for ARIMA models. This function automates steps 3, 4, and 5 of those outlined previously, in the general steps required for ARIMA modelling. In the following sections, the general steps are used to generate an ARIMA model manually, and then the automatic algorithm is utilized to build one.

3.RESULTS AND DISCUSSION

In this chapter, we analyze a time series of West Texas Intermedi (DCOILWTIC) . The importance of this study came from the application of Box and Jenkins (Holt winter and ARIMA models) approach on the analysis of this monthly series data using the R statistical program. More specifically, we will examine the time series and identify the appropriate seasonal ARIMA model for these data and estimate the models parameters. After that, we will perform diagnostic checks of the residuals for auto correlation, variance and normality assumptions.

3.1 Preliminary Investigations of the Data

The time series plots display observations on the y-axis against equally spaced time intervals on the x-axis. They are used to evaluate patterns, knowledge of the general trend and behaviors in data over time. The time series plot of West Texas Intermediate is shown in figure 1

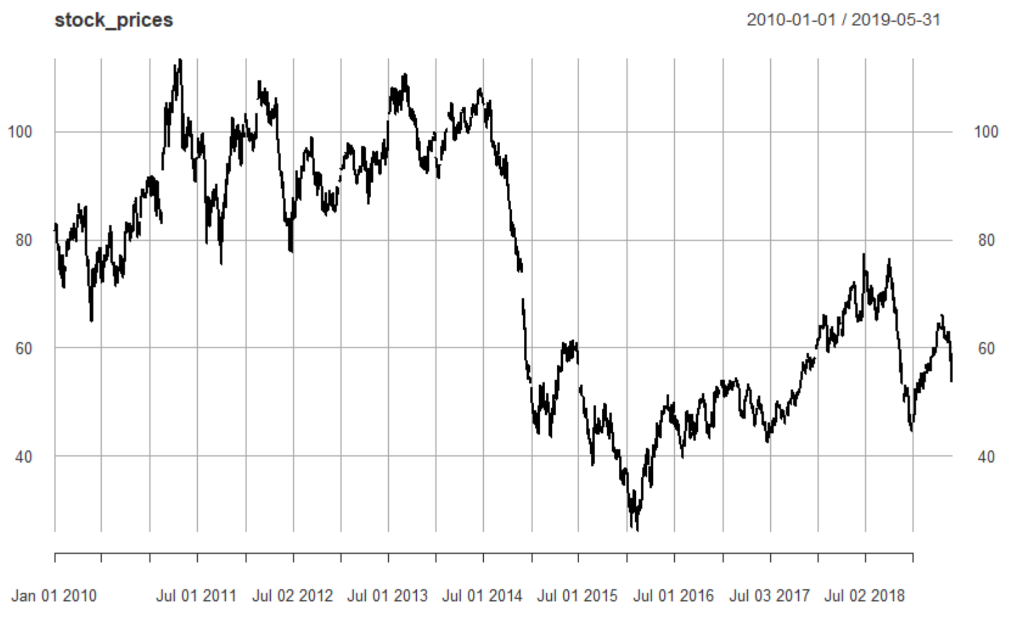

Figure 1:Time series plot of monthly West Texas Intermediate

3.2 The ARIMA models

All the above results and plots confirm that the original time series data is not stationery, and need some treatments to be transformed to a stationery series. Therefore, we used many transformations and we found that the most suitable transformation is by differencing the series. The first differences of the original series to detrend the series and achieve stability of the series. The time series for the first differenced series in Figure 2 indicates the presence of seasonal patterns.

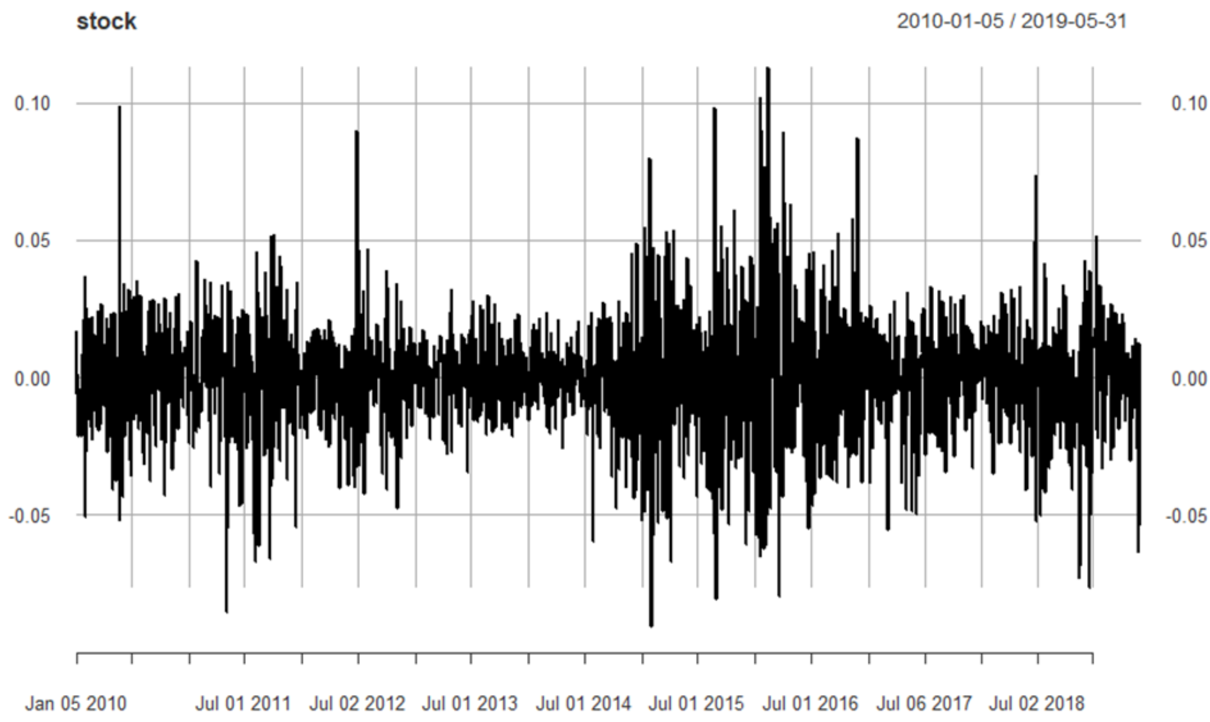

Figure 2: time series data after differencing

By estimating the autocorrelation function (ACF) for the first differenced series in Figure 3, the result of the calculated lag order and p value is shown below

Data: stock

Data: stock

Dickey-Fuller = -13.303, Lag order = 13, p-value = 0.01

Alternative hypothesis: stationary

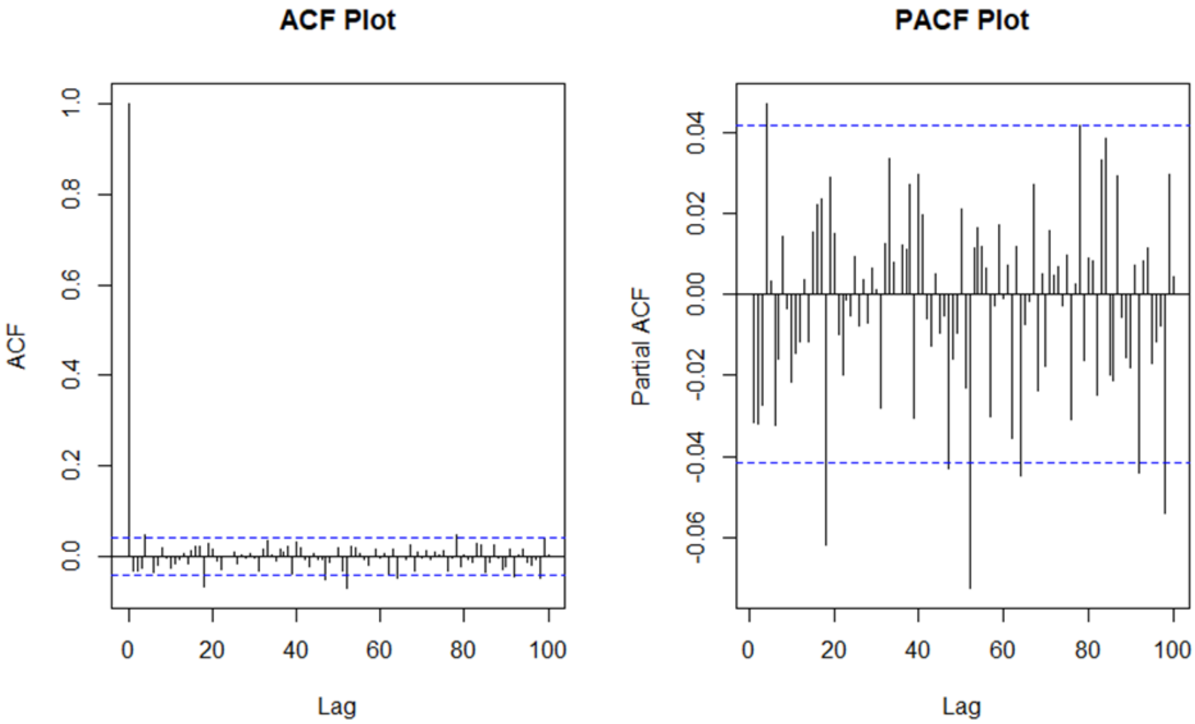

Figure 3: ACF and PACF results

3.3 Parameters Estimation:

Since we concluded in the previous section that the ARIMA model is the best model with the smallest value of AIC, BIC, MSE, RMSE, MAPE and MAE criteria, the parameters had been estimated using the method of maximum likelihood estimation as it is the best and most appropriate method of estimation. The results of the parameters estimation of the model are shown in below

Series: log10 (stock_prices)

ARIMA (0,1,1)

sigma^2 estimated as 7.907e-05: log likelihood=7788.81

AIC=-15573.62 AICc=-15573.62 BIC=-15562.01

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-7.402734e-05	0.008888651	0.006410147	-0.00191589	0.3601521	0.9959237	0.003038883

Which concluded that an ARIMA (0,1,1) is a good fit.

3.4 Forecasts

Based on the two different ARIMA models a recursively expanding one-step-ahead out-of-sample forecast is created for both models. The forecasts are visually displayed in Figure 4. It can be concluded that the ARMA (0,1,1) model provides a better forecast. It can be seen that the models are able to capture most of future price movements in two month periods.

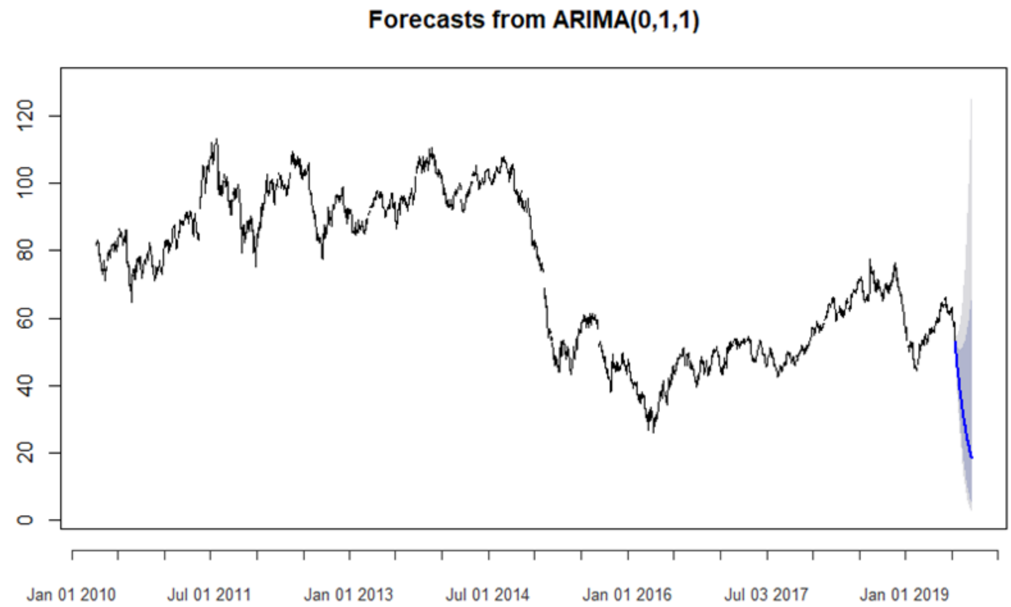

Figure 4: Forecasting the West Texas Intermediate using the ARIMA

Dickey-Fuller Test was preformed

Dickey-Fuller = -2.0871, Lag order = 4, p-value = 0.5404

The algorithm of the forecasting was tested and the testing result are shown in figure

3.5 forecasting using the Holt winters

The simple forecasting using the Holt winter method is shown in figure 6. It clearly that the result from the ARIMA is more realistic than the Holt winter method due to the fact that the oil price is declining.

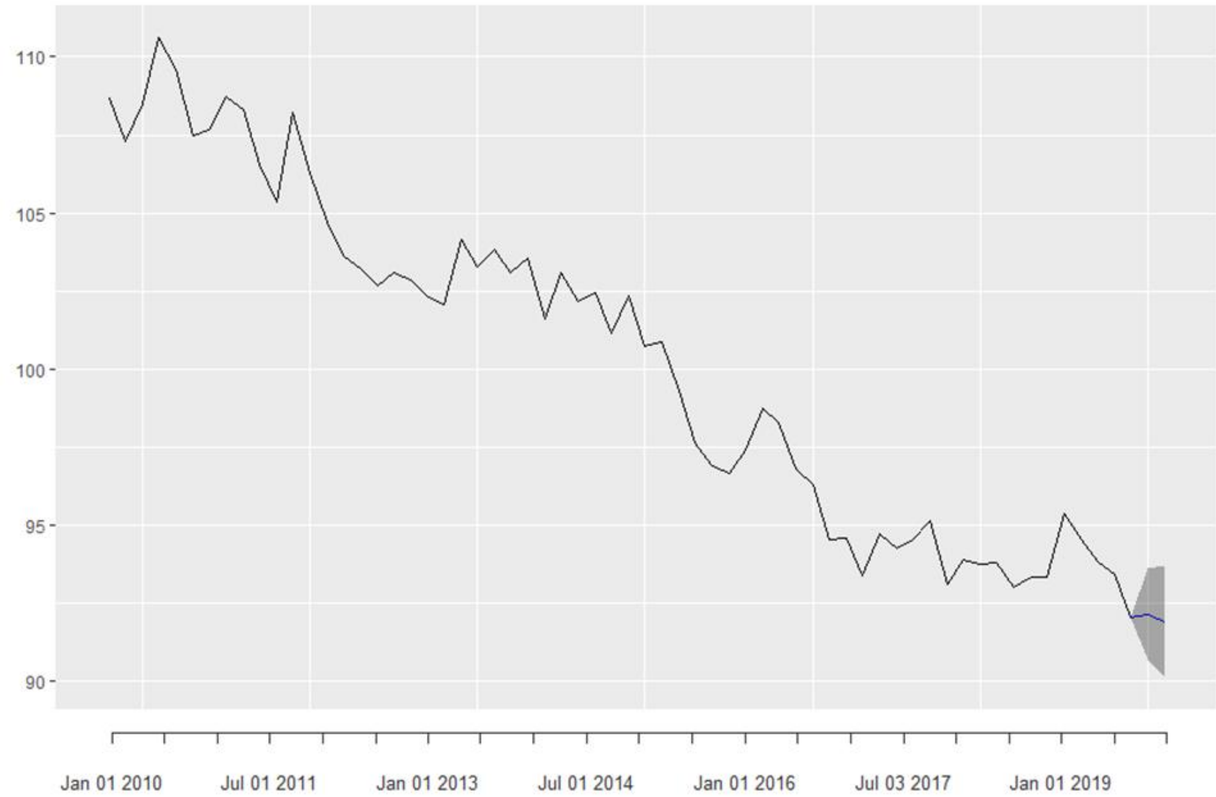

Figure 6: Forecasting of the West Texas Intermediate using the Holt winter method

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.01081923	1.104172	0.878904	-0.01851234	0.8711326	0.8992476	0.03157262

After comparing the different types of error for both method it's seen that the ARIMA has the least Dickey-Fuller Test was preformed

Dickey-Fuller = -2.1571, Lag order = 4, p-value = 0.6404

4.CONCLUSIONS

This work studied and evaluated various forecasting methods for monthly and quarterly West Texas Intermediate oil index data. It evaluated and compared few structural and time-series methods, including ARIMA and Holt winter methods. These variable were later used to regress various length of data to predict up to 2 months in advance. ARIMA model predicted drop of the oil price in 2019 towards the end of the year.

In conclusion, due to high volatility nature of oil price, it is found that ARIMA based forecasting provide the best forecasting. Nevertheless, and due to com¬plicated cost dynamics of the oil, even the best models can have very large error in predicting real price of oil in the future, particularly when it is forecasting for longer stretches of time.

5.REFERENCES

[1] L.J. Bachmeier and J.M. Griffin. Testing for market integration crude oil, coal, and natural gas. The Energy Journal, 27(2):55–71, 2006.

[2] S. D´ees, A. Gasteuil, R. Kaufmann, and M. Mann. Assessing the factors behind oil price changes. European Central Bank Working Paper Series, 855, 2008.

[3] S. D´ees, P. Karadelogou, R. Kaufmann, and M. S´anchez. Modelling the world oil market assessment of a quarterly econometric model. Energy Policy, 35:178–191, 2007.

[4] L. Derek. 2013. “Time series modeling of monthly WTI crude oil returns.” PhD diss.

[5] U. Web. 2017. “Cushing, OK WTI Spot Price FOB.”

[6] A. Ron, L. Kilian, and R. Vigfusson. 2011. “Forecasting of the Price of Oil.” International Finance Discussion Papers, 1022.

[7]. DOE, US. 2017a. “Annual Energy Outlook 2017.”

[8]. Baumeister, Christiane, and L. Kilian. 2014. “Real-time analysis of oil price risks using forecast scenarios.” IMF Economic Review, 62: 119–145.

[9].Bosler, Fabian Torben. 2010. “Models for oil price prediction and forecasting.” PhD diss. San Diego State University, Department of Mathematics.

[10] M. Saeed, and F. Foroutan. 2006. “Forecasting nonlinear crude oil futures prices.” The Energy Journal, 27: 81–96.