Copenhagen Business School

M.Sc. Business Administration & Data Science

Forecasting Monthly Brent Crude Oil Closing Prices

- ARIMA, ETS, and NNAR Forecasting Models -

Predictive Analytics

Submitted By: Eirik Egge (141164)

Normal Pages: 15 (17)

Attached: "Forecasting Brent Crude Oil Appendix"

Abstract

Brent Crude Oil is one of the two main benchmark prices for purchasing oil in the world market. Forecasting crude oil prices is important as it affects a wide range of significant sectors of the economy. This paper proposes the autoregressive integrated moving average (ARIMA), exponential smoothing (ETS), and neural network autoregression models (NNAR) for forecasting Brent Crude Oil prices. Also, Naïve, Drift and Mean are included as benchmark models. The study adopts the data on Brent Crude Oil closing price from the Federal Reserve Bank of St. Louis (FRED). In the model identification process, three models are selected: ARIMA(0,1,0)(2,0,1); ETS(A,N,N); and NNAR(23,1,13). To conclude, the models are evaluated by RMSE, MAE, MAPE, and MASE. Subsequently, this study recommends ETS(A,N,N), as it outperforms the others in terms of the given forecast accuracies.

Keywords: Brent Crude Oil Price, ARIMA, ETS, NNAR, Covid-19 Impact

1. Introduction

Crude oil is a highly significant commodity as nearly the entire world population is relying on it. Today, it is going through unprecedented times as the demand has fallen drastically. Covid-19 has led to a suspension of economic activity worldwide as the world population was forced to remain at home. Consequently, severe economic ripple effects have arisen in oil-dependent industries as industrial production has decreased, global transportation is minimized, supply chains have been disrupted or worst case gone bankrupted. Crude oil, which is sometimes referred to as the 'contemporary engine of growth', was hit the hardest [1]. Furthermore, the conflict between two of the main oil producers, Saudi Arabia, and Russia, has exacerbated the circumstance for oil prices as the two parties cannot come to an agreement regarding cutting oil prices as a response to the crisis [2].

Additionally, the nature of crude oil production is such that the production cannot be stopped as a response to lower demand. The reopening of shut oil wells is of higher cost than selling at low. In April 2020, the price of WTI per barrel reached an all-time low at negative $ 37.63.[1] Although crude oil is known for its highly volatile behavior, the previous year's behavior is unprecedented. Also, highly volatile environments are likely to remain as such in the future. However, the future for petroleum is uncertain as industries are facing increased pressure from the population due to environmental effects as crude oil is responsible for a quarter of the annual global green gas emission [3].

This paper is based on time series forecasting on the closing price of Brent crude oil prices, Europe [DCOILBERNTEU]. Brent is one of the two main benchmark prices for purchasing oil in the world market, the other being WTI. Moreover, it is a highly volatile stock. The studied time period, 1968 - 2021, includes several significant changes in oil prices along with an experienced surge in price fluctuations. Both the Gulf War (1990-1991) and US Terrorist Attack (2001) had a direct impact on supply and demand for oil. Furthermore, the Asian Crisis (1997-1998) and the Global Financial Crisis (2007-2009) affected the oil market through the financial markets. Also, the Oil Price Plunge (2014-2016) and the Covid-19 Pandemic (2019-CDT) are present in the studied time series [4].

[1] https://www.ft.com/content/a5292644-958d-4065-92e8-ace55d766654 (access date 28.07.2021)

This project aims to forecast the Brent Crude Oil Price Europe from 2015(1) till 2021(5). The investigation will use the mean, drift, and naïve forecast approach as baseline forecasts followed by an ARIMA, ETS, and feed-forward Neural Network model. Complementary to the main research question are three propositions:

(1) The forecasting models will find difficulties forecasting the time period of the Covid-19 Pandemic from its outbreak in 2019(11) until the last observation.

(2) The feed-forward neural network model will outperform both ARIMA and ETS based on the chosen accuracy evaluation metrics.

(3) Both the ETS and ARIMA models will outperform all three baseline forecasts based on the chosen accuracy evaluation metrics.

Focusing on the research question and propositions the paper is structured as followed. In section 2 related work investigating forecasting crude oil including forecasting during the Covid-19 pandemic is presented. In section 3, a time series description is presented followed by the methodology with applied processes and methods in section 4. Further, forecasts will be presented and evaluated including a discussion in section 5 followed by a conclusion in section 6. All conducted tests and visualizations, are to be found attached in the appendix, referenced in-text by section and page number.

2. Literature Review

The study of Kumar (1992) forecasts the WTI crude oil price (1985-1990). This study finds that an ARMA (1,2) model has the best goodness of fit [5]. Furthermore, the study of Azevedo & Campos (2016) presents combinations of ARIMA, ETS, Dynamic Regression, and NNAR to forecast both WTI and Brent closing price (1994-2013). The study demonstrates that an ARIMA(2,1,8) produces the best result for Brent Crude Oil. Further, the study finds that a combination of ARIMA and ETS outperformed all other suggested models. Also, this study emphasizes the promising capabilities of combining models for predicting Brent and WTI [6]. The study of Bildirici et al. (2020) analyzes Brent, WTI, and Dubai Crude Oil under the impact of Covid-19 utilizing a combination of an LSTM and GARCH model. This paper highlights that the chaotic structure of oil prices in the studied time period resulting in the poor performance of traditional forecasting models [2].

3. Data Description

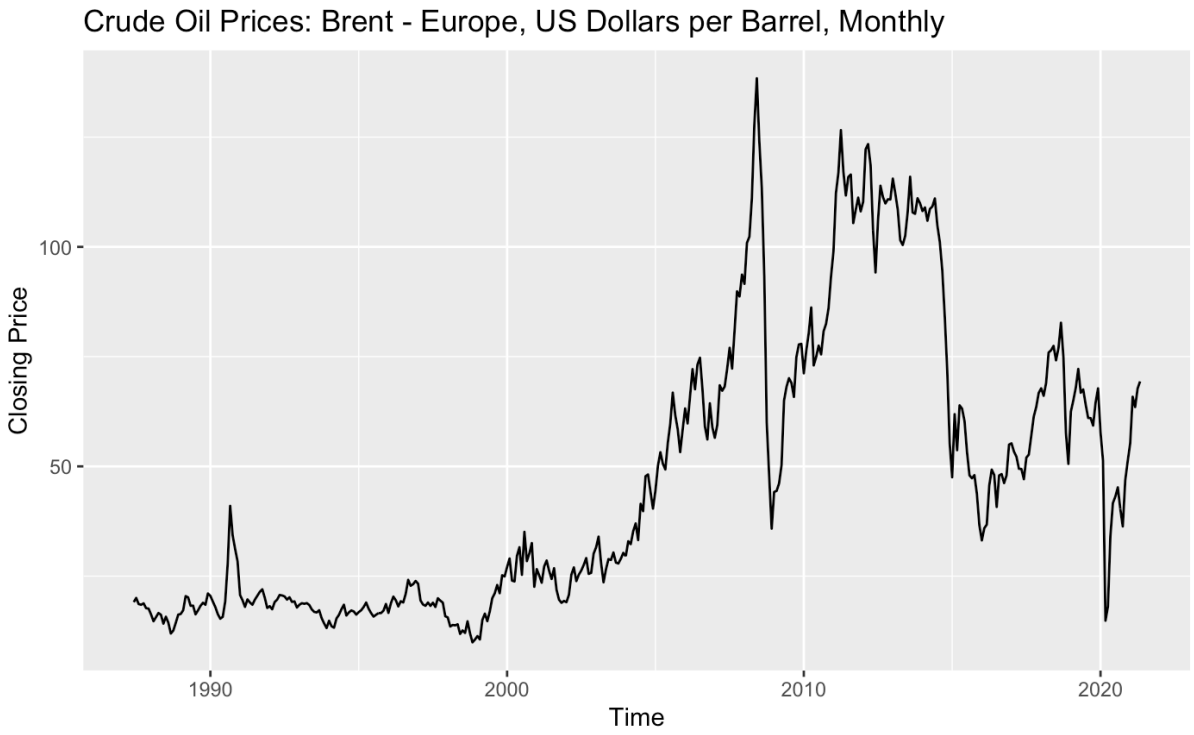

Figure 1 - Brent Crude Oil Time Series (Original Data)

Presented in figure 1 is the closing price time series of Brent Crude Oil, Europe. The studied time period ranges from 1987(6) to 2021(5). Data are displayed on a monthly average frequency, resulting in 408 observations which of none are missing values. The time series is retrieved from FRED, Federal Reserve Bank of St. Louis[2] on 07.06.2021. All observations are measured in US Dollars per Barrel, and the data are neither seasonally adjusted nor adjusted for inflation. The time series allows for forecasting the time period 2015(1) till 2021(5), which accounts for approximately 20 percent of the obtained dataset.

At first look, the time series appears to be non-stationary as the values of the time series do not fluctuate around a constant mean or with constant variance (appx. 1.1, p. 3-5) [5]. Furthermore, there appears to be an exponential increase from the year 2002 until 2008 followed by high fluctuation. In cases where the data shows inconstant variance, a mathematical transformation can simplify the patterns in the data by making the variation more stable and consistent throughout the whole dataset.

[2] https://fred.stlouisfed.org/series/DCOILBRENTEU (access date 07.06.2021)

If so, the prediction task might be simplified which can give more accurate forecasts [6]. There are several varieties of mathematical transformations, such as logarithmic and power transformations. A logarithmic transformation has the advantages of being easily interpretable and can constrain the data to stay in a positive space. Although in many cases mathematical transformations make little difference to the forecasts, they can significantly improve the prediction intervals. A commonly used method is Box-Cox transformations, which includes both log and power transformations, depending on the parameter lambda (λ) [6]. The method can be expressed as follows:

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln(y_i) & \text{if } \lambda = 0, \end{cases} \tag{1}$$

As expressed in equation 1, the natural log transformation is a special case of the Box-Cox transformation in cases where $\lambda = 0$. As the variance of the studied time series is inconsistent, this project applies a Box-Cox transformation to achieve a more stable series. The generated lambda parameter is approximately -0.414, resulting in a power transformation followed by some simple scaling. After the transformation, the variance of the time series looks more stable, and the patterns of the time series are now simplified - displaying a more linear trend (appx. 1.2, p. 6-9). For this reason, the Box-Cox transformed data will be used for further analysis.

Furthermore, this project applies the Webel-Ollech test (WO) to investigate the presence of seasonality. Although the seasonal figure of the decomposition plot indicates marginal seasonality, this is not conclusionary evidence as the decomposition assumes existing seasonal fluctuations which are the same each year. Moreover, as the WO test returns a p-value of 0.4092 the null-hypothesis of a non-seasonal time series is rejected (appx. 1.1, p. 5). Further time series characteristics and patterns will be elaborated on in the following section.

4. Methodology

4.1. Non-stationarity

Forecasts and predictions intervals can be strongly misleading for certain models in cases where time series are non-stationary. In this project, ARIMA is the only model which require non-stationarity. A time-series $\{y_t\}$ can be defined as stationary if and only if the distribution of $\{y_t, \ldots, y_{t+s}\}$ is independent on t for all s. Hence, the properties of the time series do not depend on the time at which the series is observed [7]. Properties leading to non-stationarity can be a random walk with or without drift, trends, seasonality, or structural breaks. White noise, on the other hand, is the simplest form of a stationary process. This signal can be defined as a stochastic process such that (i) the expected value is equal to zero, (ii) a homoscedastic variance, and (iii) the autocorrelation equals zero for any t unequal to s, so it does not change over time. The properties of white noise can be expressed as:

$$\begin{aligned} &\text{i.} && E[X] = 0 \\ &\text{ii.} && Var[X] = \sigma_x^2 \\ &\text{iii.} && Corr(X_t, X_s) = 0, \forall t \neq s \end{aligned} \tag{2}$$

Two commonly used methods can be utilized to obtain a stationary time series, differentiation, and detrending. Differentiating can be defined as computing the difference between each consecutive observation, thereby one observation is lost every time the time series is differentiated. Time series can be differentiated to eliminate the presence of unit roots, which can be defined as a feature of some stochastic processes, such as a random walk, that cause complication in statistical inference concerning time series models. If there are d number of unit roots present, then the time series will have to be differentiated d times to make it stationary. Differencing can also be used for detrending by stabilizing the mean of a time series by subtracting changes in the level of a given series. This project applies the Augmented Dickey-Fuller (ADF) test, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, and a visual inspection of the ACF plot to confirm and deal with the issue of non-stationarity. The ACF of a stationary time series will drop to zero relatively quickly, whereas the ACF of a non-stationary time series decreases slowly [8]. As the ACF plot of the original time series displays the latter-mentioned properties, it indicates a non-stationary process. On that account, the following paragraphs are

dedicated to the procedure and performance of the two conducted tests concerning non-stationarity. The ADF testing procedure can be expressed as:

$$\begin{aligned} &\text{i.} && \Delta y_t = \gamma y_{t-1} + \varepsilon_t \\ &\text{ii.} && \Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t \\ &\text{iii.} && \Delta y_t = a_0 + a_2 t + \gamma y_{t-1} + \varepsilon_t \end{aligned} \tag{3}$$

Here, equation 1 is a pure random walk model; equation 2 includes a drift term, whereas equation 3 includes both drift and a linear time trend component. The parameter of interest is gamma (γ), which its value determines the presence of unit root(s). Furthermore, the KPSS testing procedure can be expressed as:

$$\begin{aligned} &\text{i.} && y_t = r_t + \varepsilon_t \\ &\text{ii.} && y_t = r_t + \beta_t + \varepsilon_t \end{aligned} \tag{4}$$

The KPSS test decomposes the times series into three parts (eqn. 4.ii): a deterministic time trend (β_t), a random walk (r_t), and a stationary error (ε_t) [9]. Equation 4.ii tests if the time series is stationary around a deterministic trend, whereas equation 4.i tests for stationarity around a deterministic constant. The null hypothesis of the KPSS and ADF test is stationarity and presence of unit root, respectively. As obtained from the decomposition plots (appx. 1.2, p. 7-9), the time series present a clear trend, hence the KPSS test will be conducted with a deterministic constant and trend component whereas the ADF test will be conducted with a trend and intercept component.

Firstly, we reject the KPSS null hypothesis indicating that the data are non-stationary. Furthermore, we fail to reject all three ADF null hypotheses, suggesting that a unit root is present. Here, both the intercept (a_0) and the coefficient (γ) for y_{t-1} are statistically significant on a 0.1 % level. As the tests indicate that the process might be a higher integration order, the data is differentiated by the first order, $d = 1$. A visual inspection of the ACF plot of the differentiated series drops to zero relatively quickly - a characteristic of stationary time series [7]. Moreover, both tests with type set to trend are performed again on the differentiated time series, in addition to a visual inspection of a time plot. The returned test statistics and time plot suggest that no trend is present.

Secondly, a KPSS test with a constant and an ADF test with a drift component is conducted. We fail to reject the null hypothesis of the KPSS test, indicating that the data are stationary. Moreover, we reject both ADF null hypotheses which indicate the absence of unit root and drift. Here, the intercept (a_0) is no longer significant. Furthermore, gamma (γ) is statistically different from zero, which supports the assumption of stationarity [10]. Lastly, an ADF test with neither a trend nor an intercept component in the regression is conducted. As the test statistic is greater than the critical value at 1 pct we reject the null hypothesis of non-stationarity, indicating that a stationary time series is achieved.

4.2. Forecasting Models

4.2.1. Theoretical Framework of Time Series Models

In this section, the methodological choices regarding model identification will be explicated in addition to an explanatory section of the chosen ones. Some forecasting methods are rather unsophisticated, but they can work remarkably well given the right circumstances. Moreover, these methods can be used as benchmark forecasts, which is how this project applies the average, naïve, and drift methods. Firstly, the average method sets all future values equal to the mean of the historical data. Secondly, the naïve method sets all future values equal to the last observation of the time series. Lastly, the drift method is a derivative of the naïve method allowing the forecasts to increase or decrease over time, where the drift is set to the average change among all observations [11].

Exponential smoothing (ETS) is highly reliable for a wide range of time series with different characteristics. Furthermore, ETS decomposes the time series into three components: Error, Trend, and Seasonality. Selecting the ideal mechanisms of this model is based on recognizing the trend and seasonal components of the time series in addition to how these components enter the smoothing method. The error term can either be additive or multiplicative, the trend component can be absent, or (damped) additive, whereas the seasonal component can be absent, additive, or multiplicative [12].

Another commonly used forecasting approach is the ARIMA model. While ETS models are constructed by the type of trend and seasonality, the ARIMA models' try to describe the autocorrelations in the data. An ARMA model consists of two polynomial terms used to describe a stationary stochastic time series, where the AR term is for autoregression, and the MA term is for moving average. The I component is relevant in non-stationary processes as it describes the integrated

order. Moreover, ARIMA models can also model seasonality in which the seasonal part of the model needs to be specified [13].

Furthermore, this project applies a feed-forward artificial neural network - a more advanced forecasting method than the earlier mentioned methods. A neural network can be defined as a collection of neurons organized in layers. The neurons in the input layer feed the information received forward to other neurons calculating the weighted sum of their input. Then, an activation function is applied to modify the received information which is used as the input in the next layer. The models' parameters are the weights and biases which are learned from the historical data. These are initially guessed before they are updated through a learning algorithm called backpropagation which utilizes the stochastic gradient descent optimization technique. With time-series data, lagged values are used as inputs in the network, whereas the more recent observations are weighted higher. This project utilizes neural network autoregression, NNAR(p, P, k)$_m$, where p is the number of non-seasonal lags used for input, P is the number of seasonal lags used for input, and k is the number of nodes in the hidden layer [14].

4.2.2. Model Identification & Selection

This section is dedicated to the identification of the forecasting models used to predict the time series period from 2015(1) to 2021(5). The models will be selected by Akaike's Information Criteria (AIC). After conducting the KPSS and ADF test, the data has been differentiated of the first order, and the time series is now stationary. Furthermore, the ACF and PACF plots will be evaluated to determine the appropriate order (p) of the autoregressive part and the order (q) of the moving average part in the ARIMA model. The ACF shows the autocorrelation which measures the relationship between y_t and y_{t-k} for different values of k., whereas the PACF corrugates for the effects of the previous lags [13].

Appendix section 2.4. displays the ACF and PACF plots for Box-Cox transformed data differentiated by the first order. The ACF plot shows no clear pattern, significant spikes for lag 2, 4, 13, and 24. Conversely, there is no clear pattern suggested in the PACF but significant spikes for lags 2 and 4. Moreover, there is neither exponential decay nor a sinusoidal behavior [13]. Therefore, this project will start with an ARIMA(0,1,0), as the data has been differentiated by the first order. Additionally, as the 2nd and 4th lag are significant in both plots as well as the 24th lag in ACF, it suggests a seasonal AR(2) and MA(1) component. These two models will be compared followed by an iterative

process for exploiting potential improvement. The ARIMA(0,1,0)(2,0,1) model generates a lower AIC score compared to ARIMA(0,1,0), suggesting the model has a better fit. However, neither of the models' residuals are well behaved. The Ljung-Box test suggests the presence of autocorrelated residuals for both models including the once explored in the tuning process there might be room for further improvement as the model does not seem to capture all the variation in the data. Therefore, this project iterates through further combinations utilizing the automatic ARIMA function with no restricted parameters. Finally, an ARIMA(7,1,4)(2,0,1) with a lower AIC score is generated. Moreover, the residuals are better behaved as they are uncorrelated and fluctuate around a zero mean. However, the aim of model identification is to approximate the true data generating process, known as the principle of parsimony. Each time the time series is differentiated, or a lag is added to the AR or MA component, an observation is lost. Conversely, the general principle of parsimony states that a model described by a fewer number of parameters will give have better abilities predicting the new data [15]. Therefore, this project selects ARIMA(0,1,0)(2,0,1).

A similar procedure as for the ARIMA selection will be conducted to find the best fitting ETS model. As the time series has been box-cox transformed the error component is additive. Furthermore, after the transformation, the time series decomposition plots do not display a clear. As previously mentioned, the WO-test indicates that no seasonality is present. Therefore, an ETS(A,N,N) model will be evaluated. Similarly, the automatic ETS function suggests this model. However, several combinations of components are investigated. After an iterative process for improvement is performed, the ETS(A,N,N) has the lowest AIC score. However, the residuals are not well behaved as they are correlated and have a non-constant variance. ETS models are often more prone to non-white noise compared to ARIMA models as they decompose the time series and extract the components rather than exploiting the autocorrelation function [13]. However, further tuning has been explored in addition to validating ETS models with log-transformed and original data. As none of the explored models has the desired properties of residuals including higher AIC scores, the ETS(A,N,N) model will be used for the prediction task.

The first NNAR model has been selected through the automatic function as a starting point followed by an iterative process for exploring potential improvement. First, an NNAR(25,1,13)[12] is selected. Here, 25 non-seasonal observations and 1 seasonal observation are used as predictors, and

there are 13 neurons in the hidden layer. The optimal number of lags is selected according to the AIC score. Other models have been explored through the iterative process, however, the NNAR(25,1,13)[12] appears to capture the variance in the data best. Nevertheless, all the desired properties of residuals are not met, which will be further discussed in the following subsection.

4.2.3. Desired Properties of Residuals

The residuals of a time series model are what is left after the models are fitted. A diagnostic of the residuals is valuable in the sense of verifying whether a model has adequately captured the information in the studied time series. A sufficient model will yield (1) uncorrelated residuals in addition to (2) fluctuating around a zero mean. If the residuals are correlated indicates that there is information left out which ideally should be captured by the model. Also, if the residuals' mean is different from zero the forecast is biased. Although these properties are important for verifying if the model has captured the information available, it is an unideal approach for selecting forecasts methods. Furthermore, it is also beneficial if the residuals have (3) a constant variance and (4) normally distributed. The 3rd and 4th desired properties are beneficial as the prediction interval calculation is simplified. In cases where these are not met, there is not necessarily much that can be done to improve them. However, mathematical transformations might contribute positively [16].

This project investigates the 1st property by conducting a Ljung-Box test for ARIMA and ETS and a Box-Pierce test for NNAR. Furthermore, property 2 is investigated through mean calculation. Property 3 is investigated through a time plot of the residuals, whereas a Shapiro-Wilk normality test is performed to examine property 4. All visualizations and conducted tests can be found in the appendix (appx. 3.2, p. 22-34). Table 1 below displays whether the fitted models have met the desired properties.

Desired properties	ARIMA(0,1,0)(2,0,1)	ETS(A,N,N)	NNAR(23,1,13)
(1) Uncorrelated	No	No	Yes
(2) Zero mean	Yes	Yes	Yes
(3) Constant Variance	No	No	No
(4) Normal Distribution	No	No	No

Table 1 - Residual properties of selected models

Table 1 shows that neither ETS nor ARIMA satisfies the desired property of uncorrelated residuals, indicating that the model does not capture all the variance in the historical data. As previously discussed in section 4.2.2., all possible ETS models have been explored through the iterative tuning process, in which none of them has uncorrelated residuals. Other box-cox transformations, a log-transformation, and the original data have been explored as the chosen transformation might have over-squeezed the original variation. However, all explored models have higher AIC scores (including AICc and BIC) and the same residual properties. The property of uncorrelated residuals is only met for NNAR. However, the requirement of zero mean is met for all models. Furthermore, none of the models meet the requirement of constant variance. The variance decreases over time for both the ARIMA and ETS model, whereas the NNAR have periods of relatively high and low variance. This indicates the presence of heteroscedasticity for all chosen models, which often occurs if there is a great difference among the observations. The presence of heteroscedasticity can bring consequences, such as inaccurate convergence of the prediction interval [17]. As it is present in all the residuals of the selected models, other mathematical transformations have been explored. Unfortunately, these transformations did not assist with constant variance. Although other models like GARCH and ARCH can deal with this issue, this project will not apply them [18]. Nevertheless, correcting unequal variance should be done if the problem is severe [19]. Therefore, the models selected in section 4.2.2. will be used.

Finally, none of the selected models satisfy the desired property of normally distributed residuals. The ARIMA residual plot indicates a positive skew in addition to a longer right tail. Furthermore, the ETS residual plot indicates a negative skew and a longer right tail. Lastly, the NNAR residuals do not follow a bell curve as there are several outliers and an undesired number of observations equal to zero. Complementary to the visual inspection, we reject the null hypothesis of the Shapiro-Wilk test on a 1% level. Therefore, this project concludes that none of the residuals are normally distributed. All performed tests and visualizations can be found in appendix section 3.2. p. 22-34.

5. Forecasting Results & Discussion

This section is dedicated to the forecasting results, including a discussion of the research question and complementary propositions. The historical data has been spitted into a train and test set, which accounts for 80 and 20 percent of the data, respectively. Due to the long time period, this project will not apply cross-validation [20]. Models selected in section 4.2.2. will be evaluated by RMSE, MAE, MAPE, and MASE test scores. As previously mentioned, the forecasting period is from 2015(1) until 2021(6). Furthermore, a 95 % and 80 % prediction interval will be predicted for all models. The conducted forecasts for all models can be found in figure 2 below.

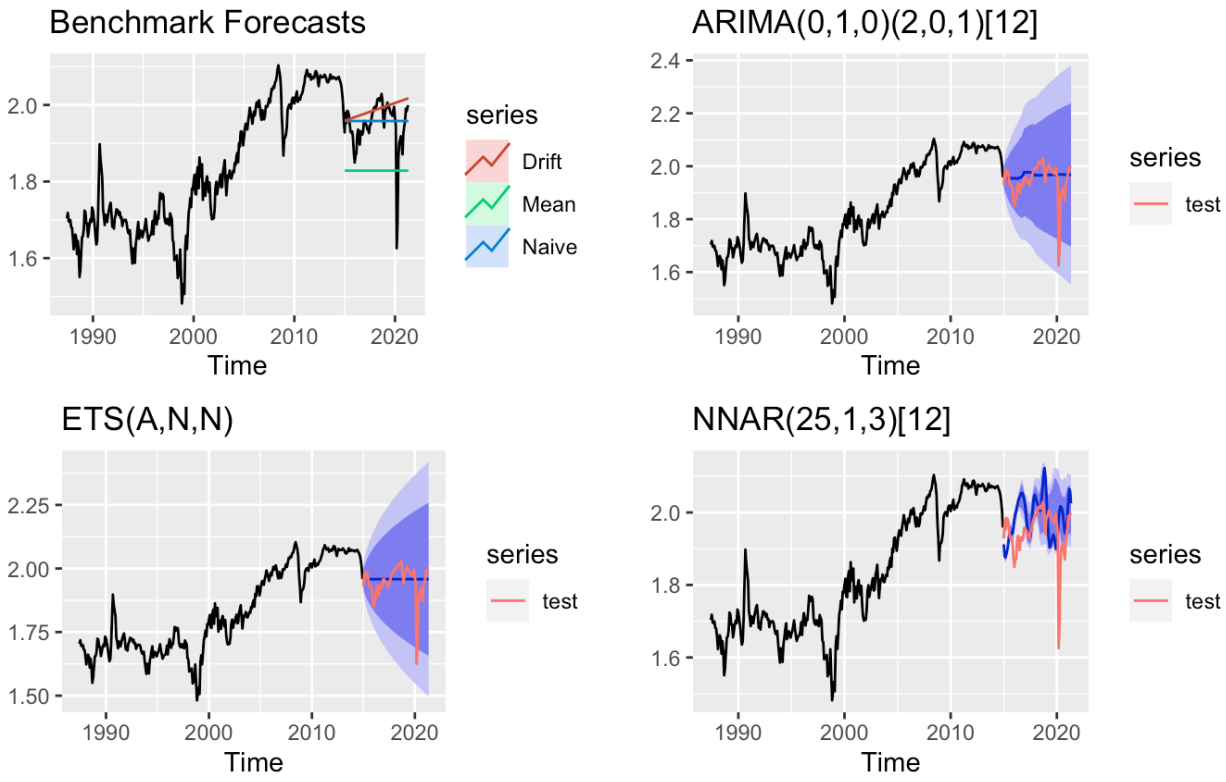

Figure 2 - Combined Forecast Plot

Furthermore, in table 2 below are the evaluation metrics of all forecasting models expressing their performance on the test set. The same evaluation metrics describing how the models fit the training data are attached (appx. 4.2., p. 36-37).

Rank	Model	RMSE	MAE	MAPE	MASE
1	ETS	0.0640	0.0416	2.1339	0.6931
2	Naïve	0.0640	0.0416	2.2138	0.6931
3	ARIMA	0.0652	0.0417	2.2226	0.6934
4	Drift	0.0767	0.0481	2.5761	0.8002
5	NNAR	0.0922	0.0679	3.5741	1.1303
6	Mean	0.1348	0.1280	6.5455	2.1300

Table 2 - Forecast Errors

5.1. Proposition 1

"The forecasting models will find difficulties forecasting the time-period of the Covid-19 Pandemic from its outbreak in 2019(11) until last observation". As displayed in figure 2, the selected forecasting models are not successful in predicting the Covid-19 Pandemic. However, this was anticipated in accordance with the research of Bildirici et al. (2020). Although the ETS model produces the lowest forecasting error during the time period of Covid-19, the differences among all models are marginal (appx. 4.2, p. 37-38). However, the closing price downfall is captured by the 95 % confidence prediction interval of both ARIMA and ETS. Moreover, NNAR and all three baseline models do not capture this downfall.

5.2. Proposition 2

"The feed-forward neural network model will outperform both ARIMA and ETS based on the chosen accuracy evaluation metrics". As the NNAR model is considered to be a more sophisticated forecasting method, this project proposed that this model will outperform the other selected models. In the model fitting process, NNAR yields the highest training accuracy. However, the results show that the NNAR model captures the variance of the test set poorly, with all accuracy metrics considered (appx. 4.2., p. 36-38). Furthermore, there are marginal differences in accuracy measures among the forecasting models. Moreover, the NNAR model is not able to capture the volatility in the 80 % certainty prediction interval, unlike ARIMA and ETS. Nevertheless, proposition 2 is rejected.

5.3. Proposition 3

"Both ETS and ARIMA models will outperform all three baseline forecasts based on the chosen accuracy evaluation metrics". First, the ETS demonstrates the best goodness of fit. However, it shares the accuracy metrics of the Naïve method, except for a marginal difference on MAPE. Both the point forecast and prediction interval of these models are virtually identical. Second, the ARIMA model is outperformed by both ETS and Naïve, but the difference is marginal considering all accuracy metrics. Consequently, proposition 3 is rejected. It should also be noted that as neither ETS nor ARIMA meets the desired property of uncorrelated residuals, there is information left out which ideally should be captured by the models. Furthermore, as heteroskedasticity might be present in these models (including NNAR) it might have caused the prediction interval to converge inaccurately (section 4.2.3.).

6. Conclusion

This paper proposed an ARIMA, ETS, and NNAR model for forecasting the Brent Crude Oil Europe closing price from 2015(1) till 2021(5), fitted on the historical data from 1987(6) - 2014(12). Due to the impact of the Covid-19 pandemic and the current oil conflict between Russia and Saudi Arabia, oil prices have gone through abnormal fluctuations. Subsequently, this project proposed that all models would find difficulties predicting this time period. As this was the case for all selected models, proposition 1 was accepted. Furthermore, propositions 2 and 3 were rejected as the selected models did not yield the proposed accuracy ranking (see table 2). In accordance with the study of Bildirici et al. (2020), the traditional forecasting models (ARIMA & ETS) had difficulties capturing the variance of the forecast period. However, this was also the case for NNAR which yielded a forecasting accuracy among the poorest performing models. Furthermore, it should be noted that the residual properties of all selected models are not well-behaved. Neither ARIMA nor ETS fulfills the desired properties of uncorrelated residuals, constant variance, and normal distribution. The desired property of uncorrelated residuals is only met for NNAR. However, all models' residuals achieve a zero mean. Consequently, this project accepts that (1) the information in the studied time series has likely not been captured adequately by the models and (2) the presence of heteroskedasticity might have caused an inaccurate convergence of the prediction intervals for all selected models. On that account, this project recognizes the limitations of the applied methodologies. Firstly, to deal with the issue of

heteroskedastic residual behavior, GARCH and ARCH models can be explored. Second, more sophisticated models can be explored as both ETS and ARIMA captures the variance of the forecasting period poorly. This has also been suggested in the research paper of Bildirici et al. (2020). Lastly, as elaborated on the in the literature review (section 2), future research can explore combining forecasting models to achieve higher forecasting accuracies.

Bibliography

[1] M. I. Haque and A. R. Shaik, "Predicting Crude Oil Prices During a Pandemic: A Comparison of Arima and Garch Models," *Montenegrin Journal of Economics,* pp. 197-198, 2021.

[2] M. Bildirici, N. G. Bayazit and Y. Ucan, "Analyzing Crude Oil Prices under the Impact of COVID-19 by Using LSTARGARCHLSTM," *Energies,* p. 2980, 2020.

[3] M. I. Haque and A. R. Shaik, "Predicting Crude Oil Prices During a Pandemic: A Comparison of Arima and Garch Models," *Montenegrin Journal of Economics,* pp. 197-207, 2021.

[4] M. Zavadska, L. Morales and J. Coughlanc, "Brent crude oil prices volatility during major crises," *Finance Research Letters,* 2020.

[5] M. S. Kumar, "The forecasting accuracy of crude oil futures prices," *International Monetary Fund Staff Papers, Springer,* p. 432, 1992.

[6] V. G. Azevedo and L. M. Campos, "Combination of forecasts for the price of crude oil on the spot market," *International Journal of Production Research,* pp. 5219-5235, 2016.

[7] G. Athanasopoulos and R. J. Hyndman, "The forecaster's toolbox," in *Forecasting: Principles and Practice*, Monash University, Australia, oTexts, 2018.

[8] G. Athanasopoulos and R. J. Hyndman, "Transformations and adjustments," in *Forecasting: Principles and Practice* , Monash University, Australia, oTexts, 2018.

[10] G. Athanasopoulos and R. J. Hyndman, "Stationarity and differencing," in *Forecasting: Principles and Practice*, Monash University, Australia, oTexts, 2021.

[11] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics,* pp. 159-178, 1992.

[12] E. E. Holmes, M. D. Scheuerell and E. J. Ward, "Dickey-Fuller and Augmented Dickey-Fuller tests," in *Applied time series analysis for fisheries and environmental data*, https://nwfsc-timeseries.github.io/atsa-labs/sec-boxjenkins-aug-dickey-fuller.html#dickey-fuller-test, 2021.

[13] R. J. Hyndman and G. Athanasopoulos, "Some simple forecasting methods," in *Forecasting: Principles and Practice*, Monash University, Australia, oTexts, 2018.

[14] R. J. Hyndman and G. Athanasopoulos, "Exponential smoothing," in *Forecasting: Principles and Practice*, Monash University, Australia , oTexts, 2018.

[15] R. J. Hyndman and G. Athanasopoulos, "ARIMA models," in *Forecasting: Principles and Practice*, Monash University, Australia, oTexts, 2018.

[16] R. J. Hyndman and G. Athanasopoulos, "Advanced forecasting methods," in *Forecasting: Principles and Practice*, Monash University, Australia, oTexts, 2018.

[17] M. B. Seasholtz and B. Kowalski, "The parsimony principle applied to multivariate calibration," *Analytica Chimica Acta,* vol. 277, no. 2, pp. 165-177, 28 May 1993.

[18] R. J. Hyndman and G. Athanasopoulos, "Residual diagnostics," in *Forecasting: Principles and Practice* , Monash University, Australia , oTexts, 2018.

[19] C. Hanck, M. Arnold, A. Gerber and M. Schmelzer, "Heteroskedasticity and Homoskedasticity," in *Introduction to Econometrics with R*, Essen, Germany, University of Duisburg-Essen, 2015.

[20] R. Engle, "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics," *Journal of Economic Perspectives,* pp. 157-168, 2001.

[21] J. Fox, "Applied regression analysis, linear models, and related methods," *Sage Journal,* p. 306, 1997.

[22] R. J. Hyndman and G. Athanasopoulos, "ARIMA vs ETS," in *Forecasting: Principles and Practice*, Monash University, Australia, oTexts, 2018.

[23] R. J. Hyndman, "Some simple forecasting methods," in *Forecasting: Principles and Practice*.