OIL PRICE FORECAST: A COMPARISON BETWEEN ARIMA, LSTM AND ARIMA-LSTM MODELS

Cássio Nunes dos Anjos* Cássio da Nóbrega Besarria †
Rennan Kertlly de Medeiros ‡ Diego Pitta de Jesus §
Hyolitta Adrielle Costa de Araújo ¶

Abstract

Oil price forecasting is an important and challenging task due to its influence on the world political-economic scenario. In this paper, three models are proposed, the ARIMA model which is a widely used statistical method to forecast time series, while the LSTM model is a recurrent neural network that can capture long-term dependencies in sequential data. The third, on the other hand, is an approach that combines the forecasting models to provide a forecast of the WTI oil price. The WTI oil series is decomposed into its trend and cycle components using the Christiano-Fitzgerald filter and each is predicted separately. With this combination of models and techniques we provide a more accurate and reliable forecast of the WTI oil price to assist in strategic decision making. The data used for the forecast was obtained from the Energy Information Administration and represents a time series with monthly records of the WTI oil price.
Keywords: Forecasting models; Oil; ARIMA-LSTM.

JEL Classification: G0, C52, C53 .
Área Temática: Econometrics and Numerical Methods.

*PhD candidate of the Graduate Program in Economics, UFPB. cassio.nanjos@professor.joaopessoa.pb.gov.br.
†Professor of the Graduate Program in Economics, UFPB. cassiodanobrega@yahoo.com.br.
‡Professor of the Undergraduate in Economics, UFPB. rennanmedeiros22@hotmail.com.
§Professor of the Undergraduate in Economics, UFPB. diegopitta13@hotmail.com.
¶PhD Student of the Graduate Program in Economics, UFPB. hyolitta01@gmail.com.

1 Introduction

Given the importance of oil in the global scenario, there is a growing investigation in the literature about the impact on the world economy in relation to its price. The forecast of oil price series is investigated using different methodologies, however, there is no consensus on which model has the best predictive accuracy. From the development of time series models, we try to explain and predict the price series of this commodity. Crude oil has a high representation in the global economy.

According He et al. (2021) oil price changes have significant impacts on the global financial markets and the real economy (see, e.g., Kilian (2009); Wang, Wu and Yang (2013); Zhang (2013); Yu, Wang and Tang (2015); Hou, Mountain and Wu (2016); Kilian and Vigfusson (2017)). Therefore, the accurate forecasts of oil prices not only are helpful to build reliable investment strategies for market investors and portfolio managers but also contribute to accurately assess macroeconomic risks for central banks and governments. Given this, numerous studies aim to improve oil price predictability by using various forecasting models, which include the vector autoregressive model (see, e.g., Baumeister and Kilian (2014)), predictive regressions with statistical and economic restrictions (Wang et al. (2015)), forecast combinations (see, e.g., Baumeister and Kilian (2015); Zhang et al. (2018)), dynamic model averaging (see, e.g., Naser (2016)), vector trend forecasting method (Zhao et al. (2018)), DFN-AI model (Wang et al. (2018)), hybrid-refined method (Chai et al. (2018)), hybrid model using variational mode decomposition and AI techniques (Li, Zhu and Wu (2019)). In turn, Albuquerquemello et al. (2018) proposed a SETAR model for oil price prediction, Hajiabotorabi et al. (2021) presented a recurrent neural network model based on multifactor wavelet (Multi-WRNN) in the decomposition time series for pricing crude oil in the futures market. Huang and Deng (2021) built a hybrid predictive model, involving variational mode decomposition (VMD) and the long-term memory model (LSTM).

Many forecasting models are proposed to forecast the price of commodities, including oil. Unsurprisingly, predicting the price of these products is still a difficult task, as it involves factors intrinsic to the product itself, in addition to the influence of external factors, such as political uncertainties. However, predictive models are an important alternative in identifying variables that may or may not be added to understand oil price fluctuations.

Given this context, the guiding question of this work is based on: Which among the ARIMA, LSTM and ARIMA-LSTM models presents the best oil price prediction? From the question presented, the objective of this work is to compare a classic model with a deep learning model or their hybrid form of the oil price and verify which one has the least error in the short and long term forecasts. Thus, the hypothesis that guides this research is that the ARIMA model is a better price predictor to the linear series, LSTM predicts better in the non-linear series and the hybrid model predicts better the original series.

Considering that the price of oil is also influenced by external factors, additionally, its price variation is capable of resulting in important changes in the world economic scenario (VO, 2011). Given the importance of the price of this commodity and for being a product that offers high macroeconomic risk to the situation of any country, impacting various sectors of economic activity, this work is highly relevant. It is worth mentioning that this study has a differential in relation to other works done, besides the hyperparameters that are being simulated, a combination of several hybrid models to predict the price of oil is also performed.

Besides this introductory section, this paper is divided as follows: section 2 presents the empirical literature, section 3 explains the concepts of the models that were used and the data, section 4 presents the results obtained, section 5 the comparisons between the models and finally the conclusion section with our final considerations.

2 Empiric literature

2.1 Importance of oil and its derivatives

Oil represents the most important source worldwide in the energy matrix, however, despite leading the energy matrix, it has been showing a reduction in its use. Oil, like other energy sources, has time and space as variants in determining its use (D'ALMEIDA, 2015). The explanations for the variation in oil prices in the international market are numerous, ranging from the quality of the product to those resulting from the criteria established by the producing countries (ROTSTEIN, 2016).

In addition, like any commodity, oil has its prices defined by supply and demand, however, its derivatives end up having higher market value and add this value to the refining process. Derivatives are products obtained in the refinement process, from the distillation of crude oil. Some examples are gasoline, kerosene, fuel oil, naphtha, lubricants, solvents, among others. Some have energy application, while others have their use directed to other types of industry, such as petrochemical and steel (D'ALMEIDA, 2015). Gasoline is one of the most important petroleum refining products, corresponding to approximately 40% of the yield of a barrel of crude oil (CHANTZIARA; SKIADOPOULOS, 2008).

Among the external factors that influence the price of oil, Aiube (2013) highlights high volatility, futures contracts, seasonality and demand shocks. The costs of production, storage, transport and hedging are also added. In 2020, the 10 largest oil producers in the world, in descending order, were: United States, Saudi Arabia, Russia, Canada, China, Iraq, United Arab Emirates, Brazil, Iran and Kuwait. As they present greater production, these countries end up having a great influence on the price of this commodity. Consumer countries also have a notable share in oil prices, with the top five countries being the United States, China, India, Russia and Japan. Furthermore, the Organization of Petroleum Exporting Countries (OPEC) also generates important power over the price of oil (EIA, 2021).

The rise in the price of oil is capable of affecting the cash flow of companies, interest rates and inflation (VO, 2011). A basic example of how the increase in the price of oil directly impacts the economy is the influence it generates on the price of gasoline and diesel oil, which in turn increase transport costs and consequently of all goods. In refineries, the dollar and oil prices are the main variables influencing fuel prices. In Brazil, until 2017, international fluctuations were passed on in a delayed manner to the domestic market, that is, consumers did not suffer so much from the readjustments adopted, which was one of the main reasons for Petrobras indebtedness. Since then, oil has shown large increases, which are significantly reflected in gasoline prices.

2.2 Time series modeling

This section was based on Lazzeri (2020). A time series can be defined as a sequence of observations over a given period of time. In everyday life, we have a simple example of a time series that we all encounter, the change in temperature over a period of a day, week, month or year. Univariate time series consist of the values of a single variable in successive situations over a period of time, while multivariate time series consist of the values of several variables in the same successive situations over the same period.

Time series analysis can provide useful insight into how a variable changes over time, or how it depends on the changing values of other variable(s). From time series prediction, it is possible to analyze the relationship of a variable to its previous values and/or the set of other variables, moreover, it has numerous uses in artificial intelligence. A time series has 4 components: trend defined as the increasing or decreasing behavior of a variable over time; cycle: which consists of an undulating movement that usually lasts several years and is usually periodic; seasonality, the cyclical behavior of time series, is a short component that usually lasts up to 1 year; and noise which is the error in the aggregate observations due to environmental factors, moreover, it is the part of the data that we cannot explain. Time series modeling methodologies make it possible to capture these components and predict them.

2.3 Difficulties in forecast analysis

The concepts presented in this section are based on ROSSI and NEVES (2014) and Granger (1969). There are numerous difficulties in evaluating predictions, so it is important to highlight them so as not to jump to premature conclusions. Some of them are briefly discussed in the following paragraphs.

Influence of the test period - it is agreed that the evaluation of the predictions of a model must be made based on new information. Due to the influence of the evaluation period on the results of forecast analyses, the forecast evaluation process must be an ongoing activity if one intends to use a particular model frequently, and when new observations are available, the analysis must be repeated.

Choice of forecast type - conditional or unconditional, this choice is linked to the modeling objective. When it comes to analyzing the potential predictive performance of an econometric model, ex-post conditional predictions may be more suitable, due to the ability to isolate the effect of forecast errors due to extrapolations of exogenous variables. If the purpose is to analyze the global effective predictive performance of the model, unconditional ex-ante predictions, in which the value of exogenous variables are not known, are more recommended.

Influence of errors in observations - measurement errors in data can distort the performance of predictions, making a good prediction seem unreliable, or else, approaching a bad prediction to the realized value. Thus, it is clear the need for a good analysis regarding the quality of the data used for any type of use of a model, whether predictive or explanatory.

Very wide confidence intervals - the first idea to test the predictive performance of a model is to define an $\alpha\%$ confidence interval of these predictions using the statistical properties obtained in the estimation and observe if the number of realizations outside the interval does not exceed $(100-\alpha)\%$. However, in practice this approach is very limited and can only be applied to extreme cases, when, in fact, there is a clear divergence between predicted and actual values. In addition, the sample size is usually small, causing loose limits.

The evaluation of forecasts can be performed in two ways: mean square error and forecast-fulfillment diagram.

When the prediction error cost function is quadratic, this statistic is appropriate for measuring predictive performance. The Forecast x Realization Diagram is a very suggestive type of chart, it is a simple way to visualize the forecast results of a model. The graphic visualization has the advantage of identifying anomalous points, very distant from the graphic. From the analysis of these points, important issues can be clarified, since data contamination may occur, or the achievements may be affected by variables of an occasional nature; which can in turn impair the analysis of the model's predictive power.

The cost of forecast error is a relevant aspect to evaluate. When evaluating a forecasting methodology or model, two types of cost must be considered for decision making. The model elaboration cost refers to the costs involved in the modeling; while the costs associated with the use of forecasts are related to losses arising from forecast errors. Clearly, the choice of methodology or model should be made with the lowest total cost.

The function that associates a loss with a given forecast error is called the forecast error cost function, also known as the loss function. This function is closely related to the decision-making process associated with forecasting. From a practical point of view, there are serious difficulties in defining this type of function and the usual thing is to make use of conventional functional forms. The loss function is designated by $g(e)$, and must satisfy the following conditions:

1. $g(0) = 0$;
2. $g(e_i) \geq g(e_j); e_i > e_j \geq 0$;
3. $g(e_i) \geq g(e_j); e_i < e_j \leq 0$.

where the e's are the forecast errors.

3 Data and methodology

3.1 Database and models

The database is composed of monthly variables. The database period corresponding to January 1986 to January 2022. The size of the sample is justified by the availability of data. The data used refer to WTI crude oil prices. Appendix **??** presents descriptive statistics for different specifications of the variable of interest.

3.2 ARIMA

Autoregressive Integrated Moving Average (ARIMA) models provide the general framework for time series forecasting techniques. They are based on the transformation of a time series into a stationary series by a process of differentiation. The ARIMA equation is a linear equation in which the predictors comprise the lags of the dependent variable and the error term, combining an autoregressive process (AR), the moving average (MA) and an order of integration (I) (CHOU; TRAN, 2018).

According to Hyndman and Athanasopoulos (2018), in a general way, this model can be written by:

$$y'_t = c + \phi_1 y'_{t-1} + ... + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_p \epsilon_{t-p} + \epsilon_t \quad (1)$$

An autoregressive model indicates that the variable y'_t is a linear function of its past values, where ϵ_t is white noise. It is defined as an AR(p) model, autoregressive of order p. Whereas the moving average process uses the terms of past residual errors, and is defined as MA(q).

For a practical approach to the ARIMA model, the interactive methodology proposed by Box et al. (2015) is used. The first step consists of identifying the model, and the authors propose the use of autocorrelation (ACF) and partial autocorrelation (PACF) functions, from which it is possible to identify the autoregressive order (p) and the moving average (q). To find an optimal model, it is necessary to ensure that the series are stationary, so, if necessary, the series can be transformed and taken in their differences. To build a stationary series, the order of differentiation (d) of the ARIMA model can be used.

Then, the parameters are estimated and the models can be chosen from the information criteria: Akaike Information Criterion (AIC), corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC). In this study, the AIC and BIC criteria were used. The last step is to carry out the diagnosis of the residuals of the adjusted model through statistical tests. In this way, when the residuals are similar to white noise, the predictions can be calculated, otherwise, it goes back to the identification step.

3.3 LSTM

The long-term memory model (LSTM) is a recurrent neural network that is used for time series to account for long-term dependencies. It can be trained with large amounts of data to capture trends in multivariate time series. Said modeling techniques are used for time series regression.

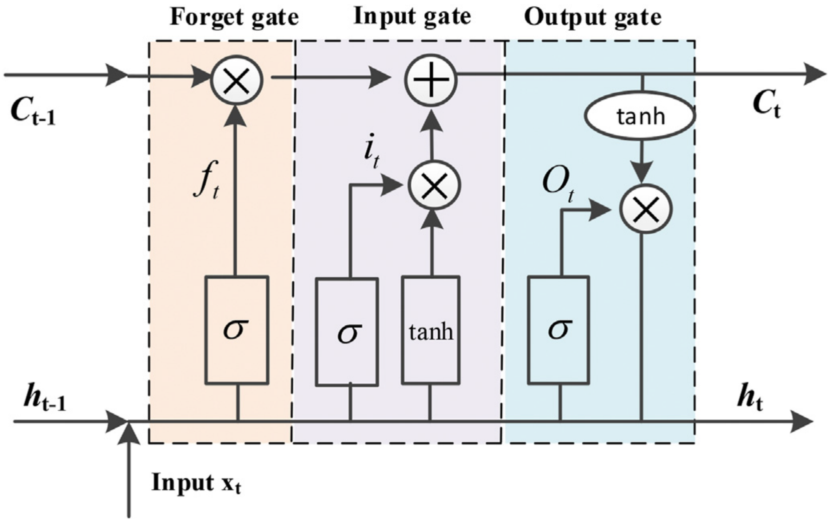
To better understand the LSTM model, it is essential to understand neural networks and recurrent neural networks. An artificial neural network is a layered structure of connected neurons, inspired by biological neural networks. It is not an algorithm, but combinations of several algorithms that allow us to perform complex operations on the data.

Recurrent Neural Networks (RNN) are a class of neural networks adapted to deal with temporal data. RNN neurons have a cellular state/memory, and input is processed according to this internal state. There are recurring modules of 'tanh' layers in RNNs that allow them to retain information. However, it doesn't hold for long, so we use the LSTM models. Long-term LSTM learning is derived from the model's recurrent module, which has a combination of four layers interacting with each other.

LSTMs have an architecture organized around gates, that is, internal mechanisms that use additions, multiplications and an activation function to regulate the flow of information in the LSTM cell, a flow similar to an electrical circuit.

Figure 1, taken from Fan et al. (2021), illustrates the internal structure of an LSTM cell, with the four neural network layers represented by four rectangles and point operators represented by white circles. The LSTM module has one cell state and three gates that provide memory capabilities, the orange gate allows for forgetting, the purple gate for storing information and the blue gate for learning. The input gate is responsible for controlling the flow of information to the current state of the cell using dot-to-dot multiplication operations of "sigmoid" and "tanh" respectively. The forgetting gate uses the sigmoid function to decide which information from the cell's previous state should be forgotten. Finally, the output port decides which information will pass to the next hidden state.

Figure 1



Since x_t is an input and h_t is an output, follow that:
$i_t = \sigma(x_t U^i + h_{t-1} W^i)$
$f_t = \sigma(x_t U^f + h_{t-1} W^f)$
$O_t = \sigma(x_t U^O + h_{t-1} W^O)$
$\tilde{C}_t = \tanh(x_t U^C + h_{t-1} W^C)$
then $C_t = \sigma(f_t \otimes C_{t-1} \oplus i_t \otimes \tilde{C}_t)$
and therefore $h_t = tanh(C_t) \otimes O_t$.

3.4 ARIMA-LSTM (hybrid model)

Due to the development of artificial intelligence, the neural network has been gaining ground in predictions, resulting from its excellent quality, namely strong non-linear mapping, self-learning, self-adaptation and fault tolerance (SHUI-LING; LI, 2017).

Hybrid models combine machine learning techniques with optimization algorithms. Compared to individual models, hybrid models compensate for weaknesses in individual techniques and thus improve the accuracy of predictions. According to the goals defined for problem solving, hybrid models can contain one or more phases (CHOU; TRAN, 2018).

The functioning of a hybrid model that we will use can be briefly explained as follows: first a filter of the series is made by applying the Christiano-Fitzgerald filter, then trend, cycle and noise predictions are made using the ARIMA and LSTM models. At the end, we sum the forecasts to obtain the final forecast value. The ARIMA model is ideal for the trend, while the LSTM model is effective for the cyclical component and both models are eficients to noise component.

The model applied in this work was based on the model created by Fan et al. (2021). Oil price data are in the form of time series data, so they can be assumed to consist of a linear portion and a non-linear portion, expressed as follows:

$$x_t = T_t + C_t + \varepsilon_t \tag{2}$$

where C_t denotes the time cycle of the data, T_t the trend and ε_t the noise component of Christiano-Fitzgerald filter. While the ARIMA model can successfully model linear relationships in the time series, the LSTM can successfully model nonlinear components. Thus, the ARIMA-LSTM hybrid model can achieve better results for the forecast, considering the advantages that each one of the models presents.

The hybrid model can be divided into the following steps:

1. recording of raw data with the Christiano-Fitzgerald filter
2. linear prediction of the trend component T_t of the series and the noise component ε_t
3. nonlinear prediction from modeling the cyclical component
4. take the best results for both metodologies to the final model
5. coupling and evaluation of the final result of the ARIMA-LSTM hybrid model.

This result can be obtained by adding the prediction results of the ARIMA model to the predicted results in the LSTM network, taking the best results for each component for the model. Finally, the performance evaluation of the model is carried out.

Considering that the crude oil market is one of the most volatile in the world, there is a certain tendency to consider that the ARIMA model is not capable of describing the nonlinearity components of oil price time series; since it is a linear model that captures linear characteristics of time series.

3.5 Forecast performance evaluation

There is no doubt that the more observations you have, the better your prediction will be. However, the choice of methodology does not depend only on this factor, it is also necessary to consider, based on the past, whether the methodology used was able to explain the path of the series, or even if it will be able to make good predictions about the future, a since good future forecasts are the objective of time series. Furthermore, it is essential to choose the methodology that has the least error in your past estimates or paths already taken. For us it is important to quantify the performance of a model to use it as a feedback and comparison. Error measurements are commented below Sooares (2020).

The *Mean Squared Error* (MSE) is the mean squared of the difference between the predicted and true values, has the same units as the squared true and predicted values, and is always positive. From the equation below, we can see that the mean squared error is more penalizing for larger errors or outliers.

$$MSE = \frac{\sum_{t=1}^{T} (e_t)^2}{T} \tag{3}$$

Since the MSE is a quadratic measure, we will use the square root of the error, which has the same unit of measure as the sample data. Therefore, it is more interpretable compared to MSE.

The *Root Mean Square Error* (RMSE) is the square root of the mean square error. It is also always positive and is in the range of the data. It also penalizes more for bigger errors.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T} (e_t)^2}{T}} \tag{4}$$

We will use this measure to compare the performance of models when analyzing medium and long-term forecast periods.

Meanwhile, the *Mean Absolute Error* (MAE) is the mean of the absolute difference between predicted values and actual values. The Mean Absolute Error has the same units as the predicted and true value and is always positive.

$$MAE = \frac{\sum_{t=1}^{T} |e_t|}{T} \tag{5}$$

We will use this measure to analyze the performance of the models when we look at short-term forecast periods.

The *Mean absolute percentage error (MAPE)* is the percentage of the mean of the absolute difference between the predicted values and the true values, divided by the true value.

$$MAPE = \frac{\sum_{t=1}^{T} \left|\frac{e_t}{y_t}\right|}{T} \tag{6}$$

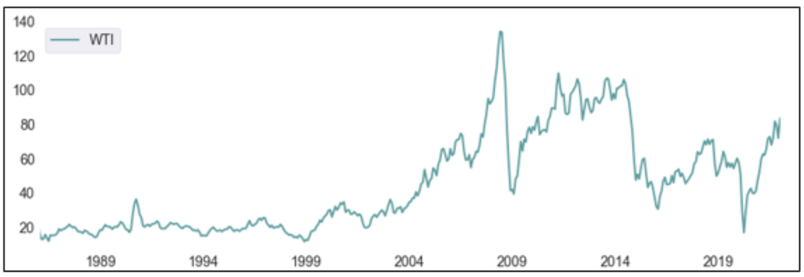
In the latter case, we look at sensitivity, since the absolute average percentage of error makes the magnitude of errors smaller.

4 Results and discussion

4.1 Time-series analysis

The dataset consists of 433 observations for the period from January 1986 to January 2022. Figure 2 illustrates the trajectory of the variable over time. Apparently, the series lacks stationarity and seasonality. Until 2004 it is possible to observe that it is almost constant, where from that moment on it presents high volatility.

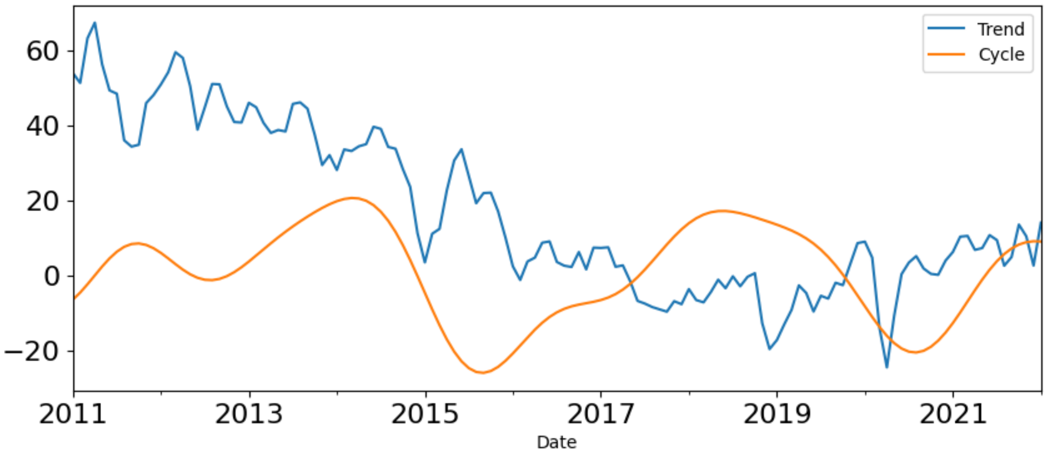
Figure 2 – WTI price trajectory



Source: Authors' elaboration

The Christiano-Fitzgerald filter is similar to the Band-Pass filter, and is applied most effectively if the T is infinite. This filter seeks to find an optimal way, considering both the accuracy and the maintenance of information. It uses all the series information to calculate the filter at each point. Another advantage of this filter is that it does not lead to a large loss of information. Figure 3 presents the application of the filter decomposing the series into trend and cycle.

Figure 3 – Christiano-Fitzgerald filter



Source: Authors' elaboration

Our initial hypothesis is that the ARIMA model better predicts the trend and the LSTM better predicts the cycles of a time series. To verify this hypothesis, the decomposition of the series into trend and cycle was performed using the Christiano-Fitzgerald filter, as we can see in Figure 3.

4.2 Treatment of the series

To apply ARIMA modeling, it is necessary to ensure that the series are stationary. In this way, unit root tests were used to check this condition. Table 1 shows the results for the Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Dickey-Fuller GLS (ADF-GLS) tests. These results suggest that the WTI (Cycle) series is stationary in level while the WTI and WTI (Trend) series have a unit root, but become stationary at first difference.

Table 1 – Unit Root Tests

Variable	ADF	PP	ADF-GLS	KPSS	Result
WTI	-2.31	-2.63	-2.44	-2.20	unit root
WTI (Diff)	-14.06	-13.06	-2.96	0.03	stationary
WTI (Trend)	-1.10	-2.04	-0.99	0.54	unit root
WTI (Trend Diff)	-4.30	-16.41	-2.51	0.05	stationary
WTI Cycle	-1.46	-3.86	-1.08	0.019	stationary

* Diff (series in first difference). Source: Authors' elaboration.

4.3 Evaluation of predictions

As seen in 3.3, the next step consists of identifying the model from the autocorrelation functions and choosing the optimal values of p and q. The selection of these values was based on the minimum criteria of AIC, and the orders chosen can be seen in Table 2. The ARIMA models presented will be used to forecast the parts of the series and the original series.

Table 2 – Optimal values of P,D,Q

Series	p	d	q
WTI	3	1	1
WTI (Trend)	1	1	0
WTI (Cycle)	1	0	0
WTI (Noise)	0	1	1

Source: Authors' elaboration

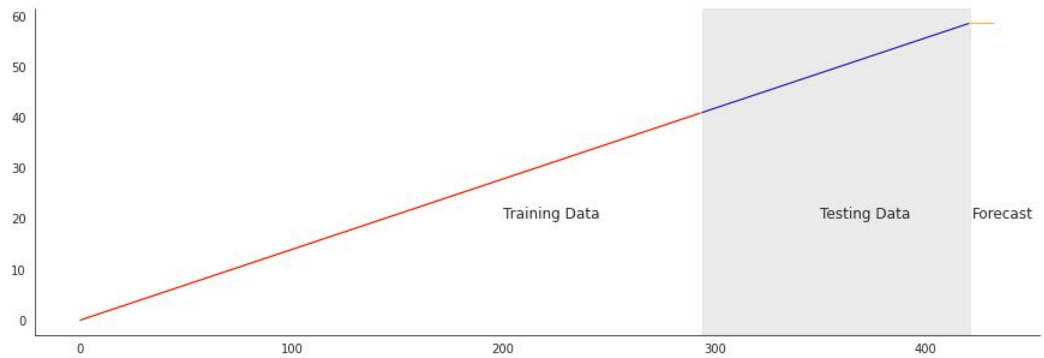
For the chosen models, the residual diagnoses were performed through the Ljung-Box autocorrelation test, the results showed that the residuals are not autocorrelated, therefore, the models were enabled for prediction. After the estimation, the accuracy of the forecasts generated out of the sample was evaluated, and the metrics applied to evaluate the errors and performance of the ARIMA model are presented in Table 3.

Figure 4 – ARIMA Modeling (WTI): (a) ARIMA Forecast

Figure 5 – ARIMA Modeling (WTI - Trend)

Figure 6 – ARIMA Modeling (WTI - Cycle)

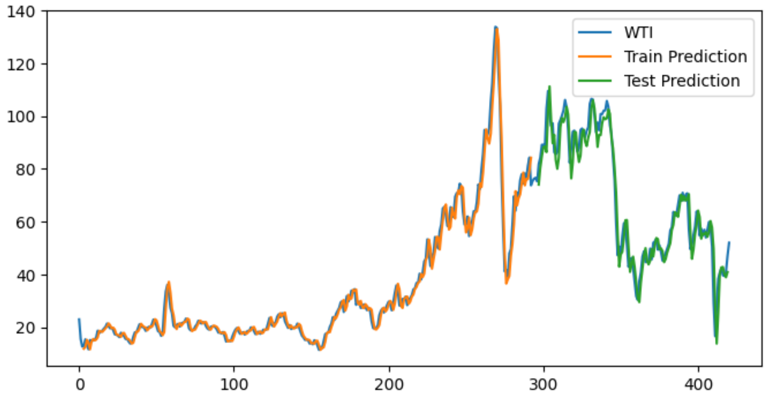
Figure 7 – ARIMA Modeling (WTI - Noise)



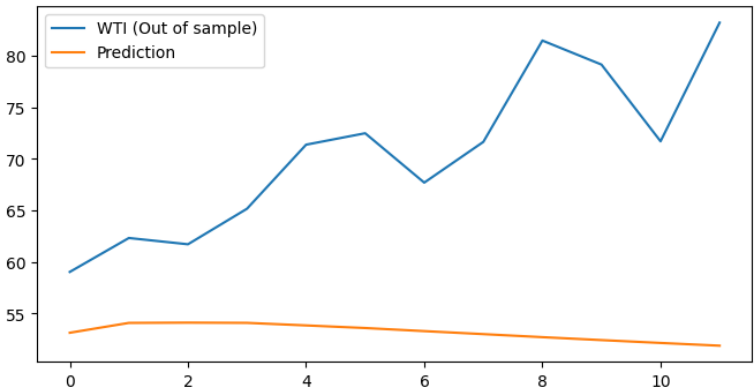
The predicted models were shown in Figures (4, 5, 6 and 7). We used 2/3 of the data for training and fitting using ARIMA modeling, 1/3 for forecasting the test data, and the forecast performed for 12 months out of sample. In the previous figures, the orange line at the end of the series are the forecast values for the ARIMA model.

The ARIMA(3,1,1) model for the original WTI series presented the best performance among the tested models. While for the trend and cycle serie were estimateds an ARIMA(1,1,0) and ARIMA(1,0,0) respectively which are AR(1) models, while for the noise was chose ARIMA (0,1,1) model which is a MA(1) model.

Figure 8 – LSTM Modeling (WTI - Trend):(a) WTI in-sample (b) WTI out-of-sample

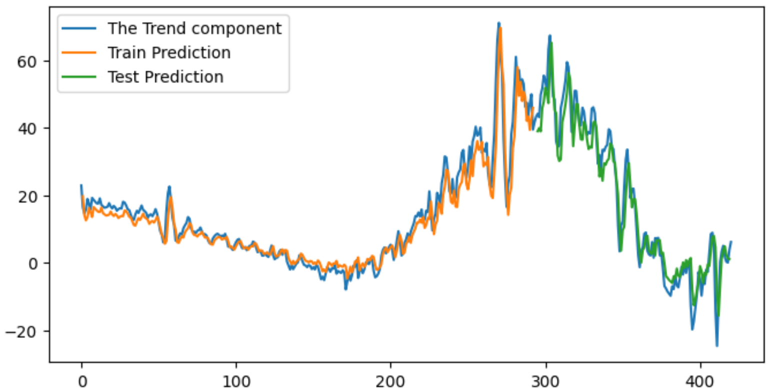


[a]

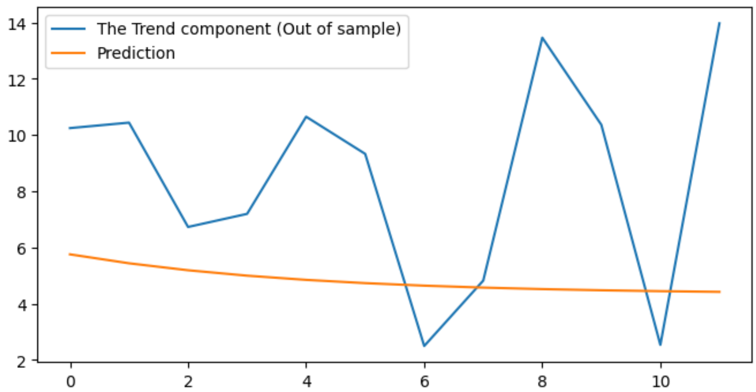


[b]

Figure 9 – LSTM Modeling (WTI - Trend):(a) Trend in-sample (b) Trend out-of-sample

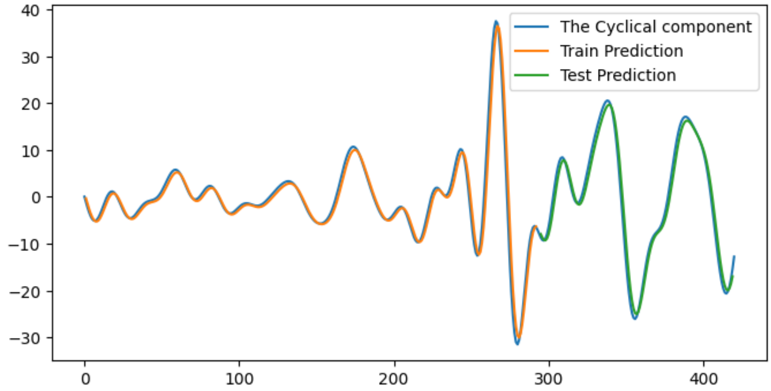


[a]

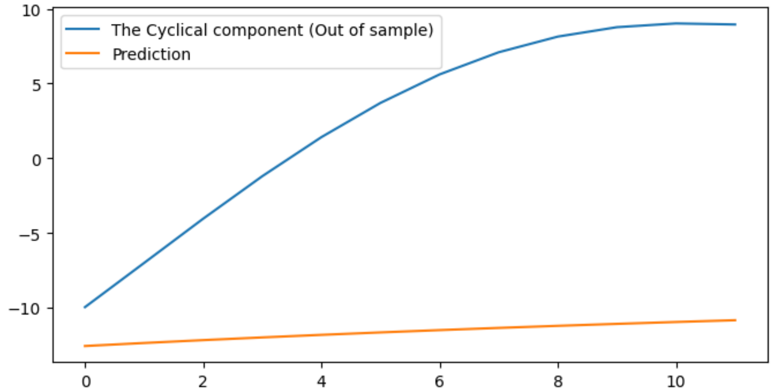


[b]

Figure 10 – LSTM Modeling (WTI - Cycle): (a) Cycle in-sample (b) Cycle out-of-sample

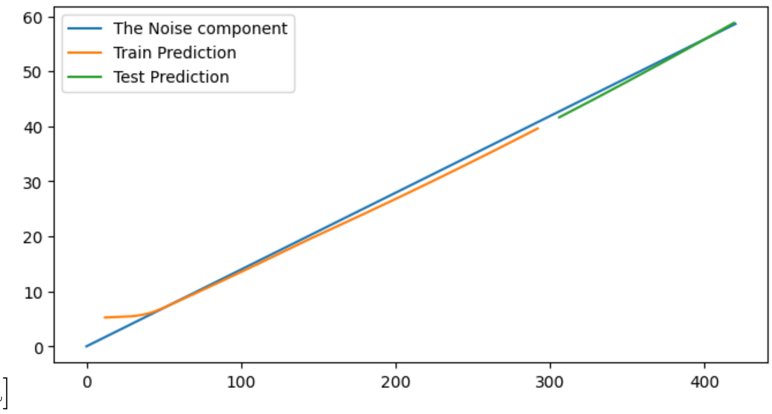


[a]

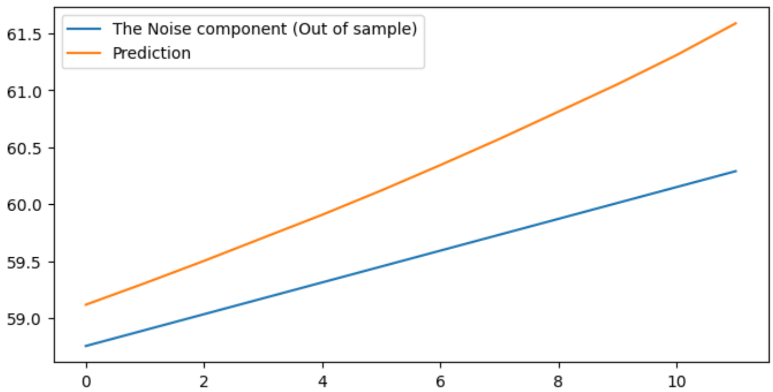


[b]

Figure 11 – LSTM Modeling (WTI - Noise): (a) Noise in-sample (b) Noise out-of-sample



[a]



[b]

Let A be a machine learning algorithm with h hyperparameters. We denote the domain of the h-th hyperparameter by D_h and the general configuration space of the hyperparameter as $D = D_1 \times D_2 \times ... \times D_h$. A vector of hyperparameters is denoted by $v \in D$, and A with its hyperparameters instantiated to v is denoted by A_v. The domain of a hyperparameter can be an integer value (e.g., epochs), binary (e.g., whether or not to use early stopping), or categorical (e.g., optimizer choice). (**??**).

For the numerical hyperparameters we use RandomSearch and GridSearch optimization methods , where initially a RandomSearch method is used to find the non-numerical hyperparameters, together with them we get a set of numerical hyperparameters. The set of hyperparameters is again optimized with GridSearch, where each numerical hyperparameter is tested with the following grid $[\frac{n}{2}, n, n^2]$, where n is the numerical hyperparameter found with RandomSearch. Binary choice is already tested for both cases in the predictions. In this study, we numerical hyperparameters are epochs, batch size, cell units, dropout rate and number of layers, while the non-numerical parameters are Dense and LSTM layer activation function, loss function and optimizer choice.

Table 3 – Error measures for prediction, by selected model

Model	**MAPE**	**MAE**	**RMSE**
WTI			
ARIMA-LSTM-LSTM	8.89%	6.74	8.64
ARIMA-LSTM-ARIMA	10.99%	8.29	10.28
LSTM-LSTM-LSTM	11.81%	8.83	10.56
LSTM-LSTM-ARIMA	14.00%	10.43	12.25
ARIMA-ARIMA-LSTM	20.62%	15.05	16.33
ARIMA-ARIMA-ARIMA	22.81%	16.65	18.06
LSTM-ARIMA-LSTM	23.63%	17.19	18.46
LSTM-ARIMA-ARIMA	25.81%	18.79	20.19
ARIMA	25.01%	18.40	20.47
LSTM	23.70%	17.39	19.17
WTI (Trend)			
ARIMA	55.54%	3.45	3.98
LSTM	50.63%	4.36	5.17
WTI (Cycle)			
ARIMA	28.27%	14.26	15.43
LSTM	122.78%	6.32	7.49
WTI (Noise)			
ARIMA	1.42%	0.85	0.98
LSTM	1.26%	0.75	0.80

Source: Authors' elaboration.

4.4 Forecasts

The forecast of the original oil series was performed using 4 ARIMA models and 4 LSTM models, their results can be seen in table 4. The ARIMA models shown in table 2 were chosen, based on the criteria listed above, in addition to the RMSE analysis of training and testing.

Regarding the LSTM models, the choice was made based on the RMSE of the training and testing. We decided to select the 2 best LSTM hyperparameter sets, for 64 choices of the set, for each of the four LSTM model architectures (each of the architectures has 16 choices of hyperparameter combination), for out-of-sample prediction and finally we chose the best model for hybrid model by analyzing the RMSE again of the two best models.

Table 4 – Predicted values outside the sample (12 months) - Original Series - Hybrid Models

Date	WTI	AAA	AAL	ALA	LAA	ALL	LLA	LAL	LLL
feb/21	59.04	52.85	53.29	56.99	51.83	57.43	55.96	52.26	56.4
mar/21	62.33	53.19	53.82	59.64	51.68	60.27	58.13	52.3	58.76
apr/21	61.72	53.4	54.22	61.21	51.6	62.03	59.41	52.42	60.23
may/21	65.17	53.57	54.6	62.17	51.57	63.19	60.17	52.59	61.19
jun/21	71.38	53.73	54.96	62.76	51.58	63.99	60.61	52.8	61.84
jul/21	72.49	53.89	55.33	63.14	51.62	64.58	60.87	53.06	62.31
aug/21	67.7	54.04	55.7	63.38	51.68	65.04	61.02	53.35	62.68
sep/21	71.65	54.2	56.09	63.53	51.77	65.42	61.1	53.67	62.99
oct/21	81.48	54.35	56.48	63.62	51.87	65.75	61.14	54	63.27
nov/21	79.15	54.5	56.87	63.68	51.98	66.05	61.16	54.35	63.53
dec/21	71.71	54.64	57.27	63.72	52.09	66.35	61.17	54.72	63.79
jan/22	83.22	54.85	57.76	63.74	52.27	66.65	61.16	55.18	64.07

Source: Authors' elaboration.

We made forecasts for each component and for the original series with different ARIMA and LSTM models for each series. Analyzing table 3 based on the MAPE, we observe that the LSTM model is better than the ARIMA model for WTI, the trend and the random component. When we consider the MAE, the LSTM model is better than the ARIMA model for WTI, the cycle and the random component. Finally, when we stop to analyze the RMSE the LSTM model is better than the ARIMA model on the WTI, the cycle and the random component. We see that the LSTM model performs better than the ARIMA model for WTI and the random component regardless of the error metric. Note also that the cycle error values show a significant difference between the models in any of the error metrics by more than doubling from one to the other. The results of the predictions can be seen in tables 6, 7, 8, 9 in the appendix.

Finally generating combinations, we have 8 hybrid models, since there are 3 components and two modeling methodologies. This combination is made through the sum of the components that were separated by the Christiano-Fitzgerald filter. Table 3 shows that of the 8 hybrid models for the WTI, the first 4 show a significant improvement over the pure models, and the first 3 reduce the errors by half, and the third best model is composed only of combinations of LSTM's. It is worth noting that although it is not in the top 4 hybrid models, combining forecasts only from ARIMA models presents an improvement over the pure ARIMA model.

Still on table 3, the 4 best models have the LSTM model as cycle model, which corroborates our initial hypothesis that the LSTM better predicts cycles, and two best models have the ARIMA as model that better predicts the trend, which also validates our other initial hypothesis. We also see that for the random component the ARIMA model always ranks immediately behind the LSTM model, for all combinations of trend and cycle.

5 Comparisons

The focus of the study was to compare traditional models, deep learning, and a hybrid model of these two models to predict the WTI oil spot price time series, which is important due to oil shocks. We employed a Christiano-Fitzgerald filter to split the sequence into linear and non-linear parts, and used eight hybrid models, of combinations of ARIMA's with LSTM's and with each other.

Our forecasts showed better performance of the ARIMA to model the linear part, the LSTM model is more effective modeling the non-linear part, and the random component fits both models well. The out-of-sample forecasting results showed that the best models for the spot price of oil are: ALL, ALA and LLL. The ALL model was superior to these two models with a difference of 1.55, 1.64 and 2.1% in MAE, RMSE and MAPE errors respectively when compared to the ALA model.

We observe that the performance of the ALA is very similar to that of the LLL model. With the ALA and LLL models having a difference of less than one unit. In general, we draw conclusions based on simulation time and errors, which indicates the ALA model as the best and we would have the ALL and LLA models as second and third best respectively due to the computational cost of the LLL model for gains in the magnitude of at most 2 in errors, a fact that is repeated when comparing ALL and ALA. When we do not consider the computation time, the most significant and reliable model is the ALL model, followed by the ALA model.

6 Conclusions

In view of the great importance of oil in the world economy, we searched several models from ARIMAS and LSTM to obtain the best performing model to predict the oil price and the three components: cycle, trend and noise. Thus, this work proves to be a potential tool for financial agents in the commodities market, especially to forecast the WTI and its impacts on the economy and the energy sector.

From the models proposed in this study, we saw that all the hybrid models for WTI oil price, they were better compared to the pure models. For predicting the trend series and the cyclical series generated by the Christiano-Fitzgerald filter, the best models are the ARIMA and LSTM models respectively, corroborating our initial hypotheses, and more applying ARIMA to predict the parts of the series, is better than using the pure ARIMA model for the original series.

In isolation, the LSTM provides good forecasts for the random component, while the ARIMA model generates forecasts with errors close to those of the LSTM, which in itself does not generate a large differential between the models that differ only in the random component, with the errors for the out-of-sample forecast being very close. Therefore, our suggestion for future work is to find a better model for the random component generated by the Christian-Fitzgerald filter and continue improving the LSTM models for the cycle and the ARIMA model for the trend.

References

AIUBE, F. A. L. *Modelos quantitativos em finanças com enfoque em commodities*. [S.l.]: Bookman Editora, 2013.

ALBUQUERQUEMELLO, V. P. de; MEDEIROS, R. K. de; BESARRIA, C. da N.; MAIA, S. F. Forecasting crude oil price: does exist an optimal econometric model? *Energy*, Elsevier, v. 155, p. 578–591, 2018.

BAUMEISTER, C.; KILIAN, L. What central bankers need to know about forecasting oil prices. *International economic review*, Wiley Online Library, v. 55, n. 3, p. 869–889, 2014.

BAUMEISTER, C.; KILIAN, L. Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 33, n. 3, p. 338–351, 2015.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.

CHAI, J.; XING, L.-M.; ZHOU, X.-Y.; ZHANG, Z. G.; LI, J.-X. Forecasting the wti crude oil price by a hybrid-refined method. *Energy Economics*, Elsevier, v. 71, p. 114–127, 2018.

CHANTZIARA, T.; SKIADOPOULOS, G. Can the dynamics of the term structure of petroleum futures be forecasted? evidence from major markets. *Energy Economics*, Elsevier, v. 30, n. 3, p. 962–985, 2008.

CHOU, J.-S.; TRAN, D.-S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, Elsevier, v. 165, p. 709–726, 2018.

D'ALMEIDA, A. L. *Indústria do petróleo no Brasil e no mundo: formação, desenvolvimento e ambiência atual*. [S.l.]: Editora Blucher, 2015.

EIA, E. I. A. International energy outlook 2021. 2021.

FAN, D.; SUN, H.; YAO, J.; ZHANG, K.; YAN, X.; SUN, Z. Well production forecasting based on arima-lstm model considering manual operations. *Energy*, Elsevier, v. 220, p. 119708, 2021.

GRANGER, C. W. Prediction with a generalized cost of error function. *Journal of the Operational Research Society*, Springer, v. 20, n. 2, p. 199–207, 1969.

HAJIABOTORABI, Z.; SAMAVATI, F. F.; GHAINI, F. M. M.; SHAHMORADI, A. Multi-wrnn model for pricing the crude oil futures market. *Expert Systems with Applications*, Elsevier, v. 182, p. 115229, 2021.

HE, M.; ZHANG, Y.; WEN, D.; WANG, Y. Forecasting crude oil prices: A scaled pca approach. *Energy Economics*, Elsevier, v. 97, p. 105189, 2021.

HOU, K.; MOUNTAIN, D. C.; WU, T. Oil price shocks and their transmission mechanism in an oil-exporting economy: A var analysis informed by a dsge model. *Journal of International Money and Finance*, Elsevier, v. 68, p. 21–49, 2016.

HUANG, Y.; DENG, Y. A new crude oil price forecasting model based on variational mode decomposition. *Knowledge-Based Systems*, Elsevier, v. 213, p. 106669, 2021.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts, 2018.

KILIAN, L. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, v. 99, n. 3, p. 1053–69, 2009.

KILIAN, L.; VIGFUSSON, R. J. The role of oil price shocks in causing us recessions. *Journal of Money, Credit and Banking*, Wiley Online Library, v. 49, n. 8, p. 1747–1776, 2017.

LAZZERI, F. *Machine learning for time series forecasting with Python*. [S.l.]: John Wiley & Sons, 2020.

LI, J.; ZHU, S.; WU, Q. Monthly crude oil spot price forecasting using variational mode decomposition. *Energy Economics*, Elsevier, v. 83, p. 240–253, 2019.

NASER, H. Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (dma) approach. *Energy Economics*, Elsevier, v. 56, p. 75–87, 2016.

ROSSI, J. W.; NEVES, C. d. Econometria e séries temporais. *Rio de Janeiro*, 2014.

ROTSTEIN, J. *Petróleo: a crise dos anos 80*. [S.l.]: Digitaliza, 2016.

SHUI-LING, Y.; LI, Z. Stock price prediction based on arima-rnn combined model. In: *4th International Conference on Social Science (ICSS 2017)*. [S.l.: s.n.], 2017. p. 1–6.

SOOARES, A. P. A. *Análise de séries temporais: aplicações em economia*. [S.l.]: Editora CRV, 2020.

VO, M. Oil and stock market volatility: A multivariate stochastic volatility perspective. *Energy Economics*, Elsevier, v. 33, n. 5, p. 956–965, 2011.

WANG, M.; ZHAO, L.; DU, R.; WANG, C.; CHEN, L.; TIAN, L.; STANLEY, H. E. A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms. *Applied energy*, Elsevier, v. 220, p. 480–495, 2018.

WANG, Y.; LIU, L.; DIAO, X.; WU, C. Forecasting the real prices of crude oil under economic and statistical constraints. *Energy Economics*, Elsevier, v. 51, p. 599–608, 2015.

WANG, Y.; WU, C.; YANG, L. Oil price shocks and stock market activities: Evidence from oil-importing and oil-exporting countries. *Journal of Comparative economics*, Elsevier, v. 41, n. 4, p. 1220–1239, 2013.

YU, L.; WANG, Z.; TANG, L. A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. *Applied Energy*, Elsevier, v. 156, p. 251–267, 2015.

ZHANG, Y.; MA, F.; SHI, B.; HUANG, D. Forecasting the prices of crude oil: An iterated combination approach. *Energy Economics*, Elsevier, v. 70, p. 472–483, 2018.

ZHANG, Y.-J. Speculative trading and wti crude oil futures price movement: an empirical analysis. *Applied Energy*, Elsevier, v. 107, p. 394–402, 2013.

ZHAO, L.-T.; WANG, Y.; GUO, S.-Q.; ZENG, G.-R. A novel method based on numerical fitting for oil price trend forecasting. *Applied Energy*, Elsevier, v. 220, p. 154–163, 2018.

APPENDIX A – Descriptive statistics and Predictions for the WTI oil

Table 5 – Statistics descriptive

Variable	Count	Mean	Std	Min	Max
WTI	433	44.790508	28.718752	11.350000	133.880000
WTI (Trend)	433	14.871740	17.764958	-24.607009	71.050785
WTI (Cycle)	433	-0.226232	10.204900	-31.493481	37.489334
WTI (Noise)	433	30.145000	17.464644	0.000000	60.290000
WTI Sample	421	44.055226	28.760517	11.350000	133.880000
WTI Sample (Trend)	421	15.052750	17.973425	-24.607009	71.050785
WTI Sample (Cycle)	421	15.052750	17.973425	-24.607009	71.050785
WTI Sample (Noise)	421	29.307639	16.981193	0.000000	58.615278
WTI Out-of-sample	12	70.586667	7.839469	59.040000	83.220000
WTI Out-of-sample (Trend)	12	8.521303	3.817064	2.494576	13.975625
WTI Out-of-sample (Cycle)	12	2.542945	6.681468	-9.962324	9.025381
WTI Out-of-sample (Noise)	12	59.522419	0.503191	58.754838	60.290000

Source: Authors' elaboration.

Table 6 – Predicted values outside the sample (12 months) - Original Series - Pure Models

Data	WTI	ARIMA	LSTM
feb/21	59.04	54.06	53.14
mar/21	62.33	54.64	54.09
apr/21	61.72	54.40	54.12
may/21	65.17	53.76	54.10
jun/21	71.38	53.00	53.85
jul/21	72.49	52.27	53.60
aug/21	67.7	51.63	53.30
sep/21	71.65	51.11	53.01
oct/21	81.48	50.71	52.71
nov/21	79.15	50.42	52.43
dec/21	71.71	50.21	52.15
jan/22	83.22	50.07	51.89

Source: Authors' elaboration.

Table 7 – Predicted values outside the sample (12 months) - Trend

Data	WTI (Trend)	ARIMA	LSTM
feb/21	10.24	6.78	5.76
mar/21	10.44	6.95	5.44
apr/21	6.72	6.99	5.19
may/21	7.19	7	5
jun/21	10.65	7	4.85
jul/21	9.33	7	4.73
aug/21	2.49	7	4.64
sep/21	4.82	7	4.57
oct/21	13.46	7	4.52
nov/21	10.36	7	4.48
dec/21	2.53	7	4.45
jan/22	13.97	7	4.42

Source: Authors' elaboration.

Table 8 – Predicted values outside the sample (12 months) - Cycle

Data	WTI (Cycle)	ARIMA	LSTM
feb/21	-9.96	-12.61	-8.47
mar/21	-7	-12.44	-5.99
apr/21	-4.04	-12.27	-4.46
may/21	-1.19	-12.11	-3.51
jun/21	1.41	-11.95	-2.92
jul/21	3.7	-11.79	-2.54
aug/21	5.61	-11.64	-2.3
sep/21	7.09	-11.48	-2.15
oct/21	8.14	-11.33	-2.06
nov/21	8.77	-11.18	-2
dec/21	9.02	-11.04	-1.96
jan/22	8.95	-10.83	-1.94

Source: Authors' elaboration.

Table 9 – Predicted values outside the sample (12 months) - Noise

Data	WTI (Noise)	ARIMA	LSTM
feb/21	58.76	58.68	59.12
mar/21	58.89	58.68	59.31
apr/21	59.04	58.68	59.5
may/21	59.17	58.68	59.71
jun/21	59.32	58.68	59.91
jul/21	59.46	58.68	60.12
aug/21	59.6	58.68	60.34
sep/21	59.74	58.68	60.57
oct/21	59.88	58.68	60.81
nov/21	60.02	58.68	61.05
dec/21	60.16	58.68	61.31
jan/22	60.3	58.68	61.59

Source: Authors' elaboration.