E.J. M. Gutierrez et al. 2019, 7:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Forecasting Crude Oil Price using Polynomial Regression and Autoregressive Integrated Moving Average (Arima) Model

Erros Josephus M. Gutierrez Karen C. De Pablo
gutierrezerrosjosephus@gmail.com karen.capiraldepablo@yahoo.com
Mathematics Department
Navotas National High School Dr. Arcadio Santos National High School
Philippines

Abstract- The researchers aim to formulate a model to forecast the crude oil price using polynomial regression and Autoregressive Integrated Moving Average (ARIMA) model. The researchers used an equally weigh price between Brent Crude Oil Price and the West Texas Intermediate (WTI) from January 2001 to December 2018 with a total of 216 observations. The Crude Oil Price has undergone logarithmic transformation in formulating the model. Statistical tests are conducted within the study to be able to come up with the best polynomial regression model and ARIMA model. In polynomial regression, quadratic regression turned out to be the best regression model with a mean absolute percentage error of 7.8203. On the other hand, ARIMA (6,1,6) turned out to be the best ARIMA model with a mean absolute percentage error of 7.1210. By Testing the forecasting accuracy of both polynomial regression of 2nd degree and ARIMA (6,1,6), in terms of Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), ARIMA (6,1,6) outperformed the regression model with Mean Square Error of 0.0913, Root Mean Square Error of 0.3021 and Mean Absolute Percentage Error of 7.1210. The analysis shows that ARIMA (6,1,6) has the best forecasting power to forecast the crude oil price. This study can help the government in reviewing and implementing policies with regards to crude oil prices in the Philippines.

Keywords: Polynomial Regression, ARIMA, Crude Oil Price, Forecasting Accuracy, Logarithmic Transformation.

I. INTRODUCTION

Crude oil is vital in economic growth of a country and is considered to be one of the most important commodities in the world. Although it is a non-renewable energy, its importance cannot be denied since it is being consumed every day. The change in its price affects significantly all other things including production, investment and future economic growth which may impact not only the stability and sustainability of a country but also the global economy. An increase in oil price induces a higher inflation which leads to a higher price in domestic products and commodities.

The tensions that take place in the international market among the Organization of Petroleum Exporting Countries (OPEC) coupled with the deepening political tensions between the two major producers of crude oil, USA and Middle East influenced the price surge in crude oil which directly hit the country [1]. On top of the impact of world oil prices, the weakening of the Philippine peso against the US dollar can be attributed to the higher domestic oil prices [2]. For decades, crude oil prices have affected the Philippines economy. During the administration of former president Ramos, the oil deregulation was signed into law which made the price of fuel an unstable subject. It did not only take away the government control on its pricing but also the subsidies on oil products. It was during the reign of former president Arroyo, when the price of crude oil reached its peak at around 132 US Dollars per Barrel and the cost of almost everything went up and led her to issue an executive order mandating oil companies to reduce their prices in Luzon and if not should face consequences [3].

During Aquino regime, the prices of crude oil has continued to rise which was attributed to an increased political tension between Middle East and North Africa and the call to suspend the VAT on oil was rejected because it may be drastic for the government collections needed for various social projects in the country [4]. During the Duterte administration when the Tax Reform for Acceleration and Inclusion (TRAIN) was signed into law on December 2017 and took effect on January 2018. This imposes an additional excise tax on petroleum products in the country. The government reported though with some hesitation that the P2.00 increase in fuel excise tax set on January 2019 would be suspended. However, at that point, world oil prices noticeably dropped, thus the suspension of TRAIN law was lifted [5].

That is why, many analysts seek to find a model that can forecast the crude oil price for future decisions and policy implementation that can have a great impact in the economy. This study is dedicated to forecast the Crude Oil Price in the Philippines for the next five years from 2019 to 2023. The forecasted values of Crude Oil Price are to be computed using the best model among polynomial regressions and ARIMA.

II. OBJECTIVE OF THE STUDY

The main objective of the study is to forecast the monthly Crude Oil Price of the Philippines from 2019 to 2023 by using the best fitted model among polynomial regression models and ARIMA. The study aims to forecast the Crude Oil Prices of the Philippines for the next 5 years from 2019 to 2023 with the application of MATLAB software. The purpose of the study is to answer the following questions:

1.What is the behavior of the graph of the annual Crude Oil Price?

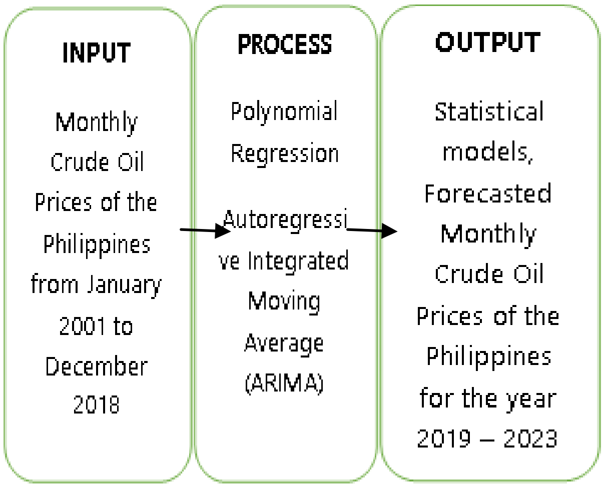
2. Which of the following statistical models best fit to forecast Crude Oil Price?

I. Polynomial Regression

II. ARIMA

3. What are the forecasted values of the Crude Oil Price for the years 2019 – 2023?

Research Paradigm



III. STATEMENT OF THE PROBLEM

1. Scope and Limitation of the Study

The researchers limit this study for 18 years. It considered years from 2001 up to 2018 for a total of 216 observations. The data were gathered from Index Mundi. The researchers applied Polynomial Regression, and ARIMA to come up with the best fitted model that can forecast the Crude Oil Prices of the Philippines for the next five years.

2 .Significance of the Study

This research focuses on further monitoring of the crude oil price of the Philippines through forecasting. The result of this study will be relevant to the Philippine Government to address the problem of the increasing price of crude oil in the international market and in the country due to the additional taxes imposed in it. The forecast will be a basis for policy makers to hold and suspend the implementation of the additional tranche of tax on the crude oil to alleviate the problems of the consumers.

IV. REVIEW OF LITERATURE AND STUDIES

Oil prices play a vital function to a country's economic development and advancement. In the Philippines, the volatility of crude oil price can cause a major impact in the inflation which affects consumers. This area exhibits a review of related literature and studies that would be useful to the investigation which are outlined, uncovering significant facts and results obtained by the pioneer in the diverse field of study.

Hamilton (2008) presented an analysis of the different variables that can affect oil price. He made an outline to demonstrate the oil price with different factors and tried to figure out which of these factors are the most significant. He examined the statistical predictability in oil prices then concluded that it is not possible to predict oil price which the leading estimate of the real oil price within the future may be its current price [6].

Afia Malik (2016) believed that oil is the second main source in the energy landscape of Pakistan in terms of consumption demand and is heavily dependent on oil import. He also believed that Pakistan, like almost all other oil importing countries in the world is quite vulnerable to oil price volatility. He reiterated that efficiency in the use of energy is the cost-effective response to high energy prices and efficiency in the use of energy can produce considerable improvements in supply in Pakistan, thus reducing reliance on imports [7].

Akira Yanagisawa (2012), "Impact of Rising Oil Prices on the Macro Economy" in his study, he underscored that oil's solid attributes as an indispensable product, especially for the time being, is inelastic versus price, that is the reason when oil prices rise, the value of oil imports by importing nations will increase at about the same rate as the rise in price [8].

Banhi Guha (2016) and Gautum Bandyopadhyay (2016) in their study on the 'Gold Price Forecasting Using ARIMA Model", he stated certain restrictions in forecasting a data using ARIMA Modeling. He believed that ARIMA modeling must only be utilized for just a short run to detect small variations in the data. In case of an unexpected change in the data set when the variation is large, it becomes difficult to capture a definite change, henceforth the model ends up ineffective in forecasting. [9]

Mensah, Emmanuel K. (2015), in his study, "Box-Jenkins Modeling and Forecasting of Brent Crude Oil Price", he believed that the famous instrument to modeling and forecasting time series is the ARIMA model. While it can foresee well in some series, its forecasting performance can be bad especially in the presence of outliers, measurement errors and volatilities in the series. In his investigation, he found out that the best fitted model in forecasting the Brent Crude Oil Price is **ARIMA (1,1,1)** however it may not forecast well in the period of high volatilities. [10]

In a study done by Shabri A. (2013), she proposed a hybrid wavelet linear regression (WLR) that combine both wavelets transform and linear regression for crude oil forecasting. He divided the data into two parts; training and testing data. He compared the prediction accuracy of WLR to the other models like linear regression, ARIMA and GARCH. He found out that the forecasting accuracy of the hybrid WLR is the best as compared to the other models. He also underscored that the forecasting abilities of the linear regression model can be improved if the wavelet transform technique is adopted for pre-processing. [11]

Ahmed R.A., and Shabri A.B (2014), used the Western Texas Intermediate (WTI) crude oil price in their experiment since many researchers have already used the data. They proposed a forecasting technique based on support vector machine. RMSE and MAE were used to evaluate the performance of ARIMA, GARCH and the proposed method. The study revealed that the proposed method outperforms the other methods in terms of forecasting accuracy. The study also recommended the use of the proposed SVM method in forecasting crude oil price. [12]

Quan L. (2014) believed that one of the factors that affect oil price is its complexity. He used the historical data of the crude oil price to come up with the ARIMA model. He reiterated that there are a lot of factors that can affect the crude oil price and he just considered its historical data, and that data is limited, so there may be some prediction errors. From the Daqing oil price forecasting, ARMA model is appropriate, the static model to better simulate historical values. He also underscored that the study does not consider the impact of other external factors. [13]

Urrutia, Alano, Aninipot, Gumapac and Quinto (2014) used multiple linear regression and stepwise regression analysis to determine significant factors affecting foreign trade and found that exchange rate, monthly domestic crude oil prices. Inflation rate and interest rates were significant predictors of imports while exchange rate, monthly domestic crude oil prices, and inflation rate were significant on exports. They also found that the domestic crude oil has uni-directional granger causal relationship with import and export. After considering all the assumptions for SARIMA modeling, they were able to obtain a SARIMA (5,0,8) x (0,1,1) 12 for import, and SARIMA (7,2,3) x (0,1,1) 12 for export in forecasting import and export for 7 years. [14]

Urrutia, de Guzman, Mercado, Bautista and Baccay (2015) made use of logarithmic transformation of the data to satisfy all the assumptions of a multiple regression analysis and had used a normal estimation equation using matrices to come up with a model that can forecast the exchange rate of the Philippines. they also found that only two independent variables namely: interest rate, and labor force participation rate had significance in the dependent variable which is the exchange rate. [15]

Urrutia, Mingo, and Balmaceda (2015) used multiple linear regression to identify significant factors that can influence income tax revenue of the Philippines and these are: employment population, annual domestic crude oil prices, and inflation rate. In forecasting the income tax revenue for the year 2014 – 2020, autoregressive integrated moving average (ARIMA) was used and obtained an ARIMA (0,1,0) as the best fitted model. They also found that the predicted values obtained from the model have no significant difference to the actual value of the income tax revenue. [16]

Urrutia, de Guzman, Olfindo, Mercado, Bautista, and Baccay (2015) aimed to formulate a mathematical model for imports and exports of the Philippines. They made use of regression analysis in formulating a model and found that out of four factors that are said to be affecting imports and exports, only two are significant namely: exchange rate, and monthly domestic crude oil. However, the researchers recommend to look for other independent variables to assess imports and exports of the Philippines more accurately. [17]

Urrutia and Olfindo (2015) in their study on "Modeling and Forecasting the Exchange Rate of the Philippines", they discovered that out of the five independent variables namely: interest rate, inflation rate, labor force participation rate, import and export only the interest rate and labor force participation rate were significant factors of exchange rate when regression analysis was applied. They also obtained an ARIMA (0,1,0) as the best-fitted model for the 6-year forecast of the exchange rate of the Philippines after considering all the assumptions of ARIMA modeling. They suggested adding more factors like crude oil prices, money supply, gross domestic product, domestic currency and central bank intervention to formulate a more accurate model to be used in forecasting. [18]

Urrutia, and Tampis (2017) aimed to formulate a model in estimating the GDP of the Philippines and tried to determine which among the different factors namely capital formation, total trade, interest rate, inflation rate, unemployment rate and stock exchange index can predict GDP and found that capital formation, total trade, interest rate, unemployment rate, and exchange rate were significant predictors of GDP using multiple linear regression. They suggested to do further studies that investigate economic growth in the Philippines. [19]

Tularam and Saeed (2016) looked at the performance accuracy of exponential smoothing, Holt Winter, and autoregressive moving average (ARIMA) with the actual data. They discovered that ARIMA (2,1,2) model gave forecasts that were more accurate than those of ES or HW models. They additionally clarified that an appropriate time-series model will help policy makers and advertising strategist settle on decisions and devise appropriate strategic plan in oil industry. [20]

Xin James He (2018), utilized three unique methods such as simple exponential smoothing, moving averages, and autoregressive integrated moving average (ARIMA) model in forecasting the crude oil price. In his study, the simple moving averages and simple exponential smoothing can provide a reasonable acceptable forecasting accuracy. Second, the more advanced autoregressive integrated moving average such as ARIMA (0, 1, 1) can offer more accurate forecasting results for most time series data, with reasonable computational multifaceted nature, and its model parameter(s). [21]

Siddhivinayak Kulkarni and Imad Haidar (2009) in their study, they presented a model based on multilayer feed forward neural network to forecast crude oil spot price direction in the short-term up to three days ahead. They also tested the relation between crude oil future prices and spot price, and if futures are good predictors to the spot applying nonlinear ANN model. They also found out that applying 3 days simple moving average to the original data then transform it into relative change is the best method amongst the other means tested. [22]

Zhang, G.P. (2003) In his study, he believed that autoregressive integrated moving average (ARIMA) is a well-known yet powerful linear model used in time series forecasting in the past decades while the artificial neural network (ANN) can be an option to the conventional linear methods. He proposed a hybrid model combining ARIMA and ANN taking advantage of their strengths in linear and nonlinear modeling capability. He believed that a mix of these strategies can be a successful method to improve forecasting performance. [23]

The different studies showed how forecasting is done in a variety of ways and processes. Most of the studies presented above used ARIMA and the researcher found no studies using polynomial regression in forecasting the crude oil price. Studies related to crude oil price forecasting used the Brent Crude Oil Price and the West Texas Intermediate however, this study uses an equally weigh price of the two.

V. METHODOLOGY

The researchers used an equally weigh price between Brent Crude Oil Price and the West Texas Intermediate (WTI) from January 2001 to December 2018 with a total of 216 observations. The data were divided into two. The period from January 2001 through December 2015 served as the training data with 180 observations for model development and the period from January 2016 through December 2018 served as the testing data with 36 observations to test the model's forecasting accuracy and consistency. The Crude Oil Price has undergone logarithmic transformation in formulating the model.

1.Multiple Linear Regression

Multiple Linear Regression is a form of linear regression analysis that is statistically used for estimating relationship among variables which have reason and result. The general form of multiple linear regressions is given by:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (1)$$

where $\hat{y}$ is the predicted values of the dependent variable, X is the independent variable, β is the parameter, and ε is the error. [24]

2. Polynomial Regression

Polynomial Regression is a special case of multiple linear regressions that fits the data into a model as a polynomial of nth degree. This likewise estimates the relationship between the independent variable x and the dependent variable y. [25]

The general polynomial regression equation is

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots + \beta_n x^n + \epsilon \quad (2)$$

where y is the predicted values of the dependent variable, x is the independent variable, β is the parameter, and ϵ is the error.

3. Evaluation of Accuracy

The mean square error (MSE) is an estimator of the variance S^2, while standard deviation S is estimated by the root mean square error (RMSE). These two important estimators are given by:

$$MSE = \frac{1}{N}\sum_{I=1}^{n}(y_i - \hat{y})^2, RMSE = \sqrt{MSE} \quad (3)$$

Where y_i is actual expected output and $\hat{y}$ is the model's prediction. [26]

Mean absolute percentage error (MAPE) is considered to be the most useful measure to compare forecast accuracy and it also measure relative performance. It is defined as

$$MAPE = \frac{100}{N}\sum_{I=1}^{n}\left|\frac{y_i - \hat{y}}{y_i}\right| \quad (4)$$

If MAPE value is less than 10%, then its forecasting accuracy is excellent, if it is between 10% - 20%, then forecasting accuracy is good, if it is between 20% – 50 %, then forecasting accuracy is just acceptable but if it is over 50%, then forecasting accuracy is inaccurate. [27]

The coefficient of determination R^2 which explains the variation in the dependent variable that is explained by the independent variable is defined as

$$R^2 = 1 - \frac{SSE}{SST} \quad (5)$$

Where SST is the total sum of squares. The value of R^2 always lies between zero and one, that is $0 \leq R^2 \leq 1$. This is an important measure of how the regression fits the data well. An R^2 value of 0.9 or above is very good, a value above 0.8 is good, a value of 0.6 or above may just be satisfactory in some cases and a value of 0.5 or below means that only 50% of the variation in the data, thus prediction becomes poor. [28]

4. Box – Jenkins Method

Box – Jenkins Method is a process of systematically formulating a model which requires a moderately long time series in building a good model and by using the autoregressive integrated moving average (ARIMA). There are three major stages in building a Box – Jenkins model: [29]

- Model Identification. In this stage, a variety of methods is used to identify candidate ARIMA models. The first and foremost step is to test whether a series is stationary or not. Differencing the series to attain stationary, plotting the autocorrelation function (ACF) and partial correlation function (PACF) to determine the number of possible lags in a model.

- Model Checking. Significance tests are performed to estimate the parameters of the model and this includes the maximum likelihood estimation.
- Model Validation. In this stage, forecast of the future values of the time series are calculated and generates confidence interval from the ARIMA model.

5. Augmented Dickey – Fuller Test

Augmented Dickey – Fuller Test is a method used to test the null hypothesis that a unit root exists in an autoregressive model. It is also a test that determines whether you can conclude that a time series is stationary. If a unit root exists, then the time series is not stationary, thus differencing is necessary. It is defined as [30]

$\Delta Y_t = \alpha + \beta t + \gamma Y_{T-1} + \sum_{J=1}^{p} (\delta_J \Delta Y_{t-j}) + e_t$

(6)

Where t is the time index, α is the intercept constant, β is the coefficient on the time trend, γ is the coefficient presenting process root, p is the lag order of the first-difference autoregressive process, e_t is the independent identically distributes residual term.

6.Autoregressive Integrated Moving Average

The Autoregressive Integrated Moving Average also known as the ARIMA process analyzes and forecasts a univariate time series data with equally spaced interval. It is popularized by Box and Jenkins and commonly referred to as the Box-Jenkins model. Once the series is stationary, after differencing, the ACF and PACF are plotted to determine the possible lag order for the autoregressive and moving average term of an ARIMA model.

A no seasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where: [31]

-p is the number of autoregressive terms,

-d is the number of no seasonal differences needed for stationary, and

-q is the number of lagged forecast errors in the prediction equation.

In terms of y, the general forecasting equation is:

$\hat{Y}_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$

(7)

Where ŷ is the forecasted values of the dependent variable, t is the time index, μ is the intercept, ϕ is the autoregressive operator and θ is the moving average operator.

VI. RESULTS AND DISCUSSION

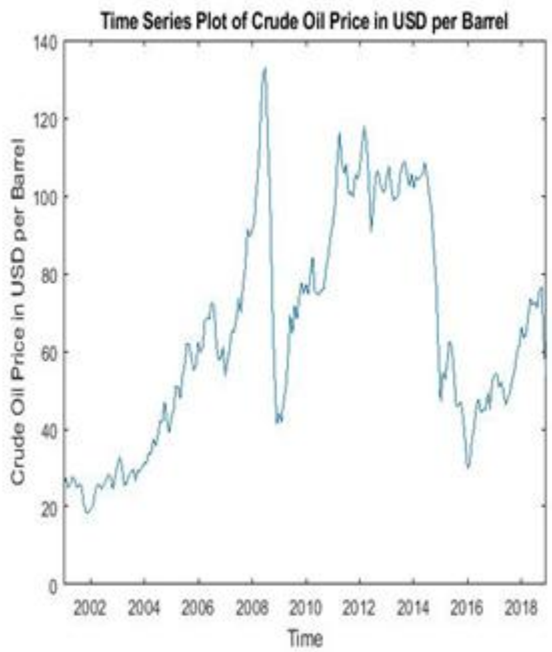


Figure 1: Time Series Plot of the Crude Oil Price in USD per Barrel.

1.Behavior of the Graph

The graph shows the time series plot of the Crude Oil Price for 18 years from January 2001 up to December 2018. The volatile nature of the crude oil price can be seen as it shows a couple of ups and downs on its price in the past years. It can be gleaned that on December 2001, the crude oil price was only 18.52 USD per barrel and continue to rise until it reached its peak on July 2008 when the price soar high to as much as 132 USD per barrel due to the oil crisis in the international market which was felt worldwide. The graph also shows a non-constant mean and variance which indicates a non-stationary series. Logarithmic transformation is necessary to stabilize the variance.

2. Statistical Models

2.1 Polynomial Regression Model

Table 1Coefficients of Regression Models with 95% Confidence Bounds

Linear Regression Model	Quadratic Regression Model	Cubic Regression Model
$\hat{y} = \beta_0 + \beta_1 x$	$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$	$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

β_0 = -181.9 (-205.2, -158.5)	β_0 = -6.032e+04 (-6.858e+04, -5.206e+04)	β_0 = 1.22e+07 (8.219e+06, 1.618e+07)
β_1 = 0.0002535 (0.0002216, 0.0002853)	β_1 = 0.1642 (0.1417, 0.1867)	β_1 = -49.96 (-66.23, -33.69)
	β_2 = -1.117e-07 (-1.271e-07, -9.64e-08)	β_2 = 6.822e-05 (4.604e-05, 9.04e-05)
		β_3 = -3.105e-11 (-4.113e-11, -2.097e-11)

Table 2 Goodness of Fit

Table 1 Regression Model	SSE	R^2	R^2 adjusted	RMSE
Linear	20.8917	0.5806	0.5783	0.3426
Quadratic	9.6451	0.8064	0.8042	0.2334
Cubic	7.9707	0.8400	0.8373	0.2128

As shown in the table above, the cubic polynomial regression appears to fit the data best and outperforms the other models with lowest error statistics and highest deterministic coefficient. While it's tempting to fit a higher degree

polynomial to get a lower error, this can lead to over-fitting that's why it is important plot the relationships and focus on making sure that the curve fits the nature of the problem. The higher the degree of the polynomial, the more likely it is to produce a weird result on extrapolation.

Table 3 Forecasting Accuracy of the Three – Year Forecast using the Testing Data Set

Regression Model	MSE	RMSE	MAPE
Linear	0.9010	0.9492	19.0876
Quadratic	0.1367	0.3697	7.8203
Cubic	0.8678	0.9315	25.6081

The table shows the forecasting accuracy for the different candidate regression models. By testing the forecasting accuracy of all the models, it appears that quadratic polynomial regression is the best model for future forecast since it has the lowest mean absolute percentage error of 7.8203.

2.2 Autoregressive Integrated Moving Average (ARIMA) Model

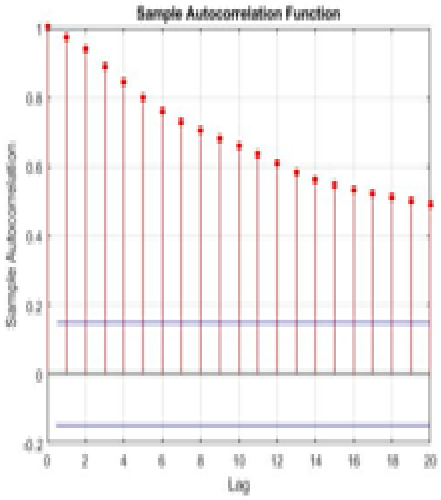


Figure 2: ACF Plot of the Log Return of the Crude Oil Price.

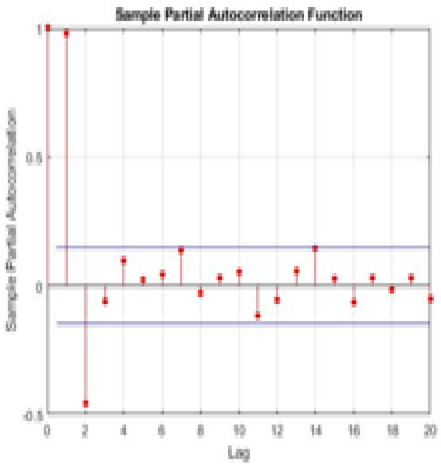


Figure 3: PACF Plot of the Log Return of the Crude Oil Price.

In figures 2 and 3, the downward trend of the plot indicates a unit root. The lengths of the line segment on the ACF plot gradually decay, and continue this pattern for increasing lags. This behavior indicates a no stationary series. To verify the assumption that a unit root exists based on the graph of the autocorrelation Augmented

Dickey – Fuller Test may be used. The Augmented Dickey-Fuller Test of the log return of the crude oil price shows a presence of a unit root, thus differencing is required to attain stationary. The log return of the crude oil price, being differenced once has a p-value of 0.001 which rejects the null hypothesis of having a unit root, thus achieved stationary.

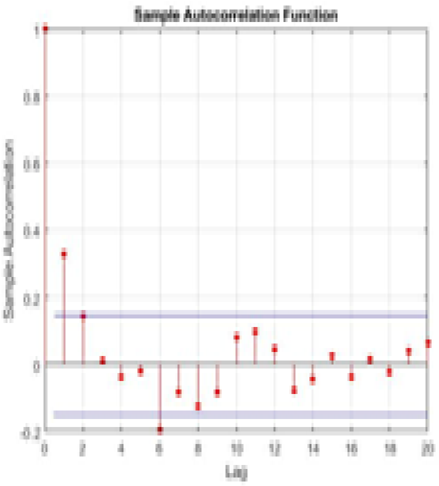


Figure 4: ACF Plot of the Differenced Log Return of the Crude Oil Price.

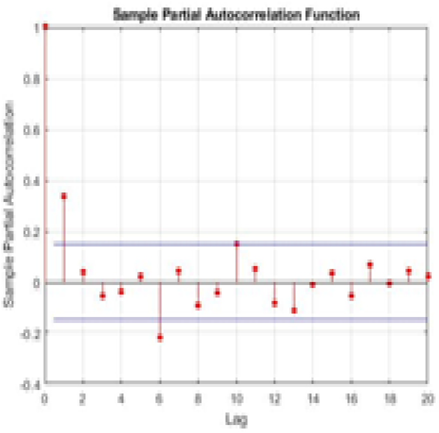


Figure 5: PACF Plot of the Differenced Log Return of the Crude Oil Price.

In figures 4 and 5, The Sample Autocorrelation Function of the differenced log return shows a significant spike at lag 1 and at lag 6 while the Sample Partial Autocorrelation Function shows a sharp cut off after lag 1 and also has a significant spike at lag 6. These significant lags for AR and MA are needed for the identification of the candidate ARIMA models. The Candidate ARIMA models based on these plots are ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (0,1,1) and ARIMA (6,1,6).

Table 4 Goodness of Fit

	ARIMA (1,1,0)	ARIMA (0,1,1)	ARIMA (1,1,1)	ARIMA (6,1,6)
AIC	-391.6858	-388.0498	-389.8978	-396.5974
BIC	-382.1405	-378.4876	-377.1707	-352.4513

Table shows the candidate ARIMA models with their AIC and BIC. ARIMA (6,1,6) has the lowest AIC value suggesting that this model nicely straddle the requirements of the goodness of fit.

Table 5: Forecasting Accuracy of the Three – Year Forecast Using the Testing Data

Model	**MSE**	**RMSE**	**MAPE**
ARIMA (1,1,0)	0.2359	0.4857	12.5986
ARIMA (0,1,1)	0.2001	0.4474	11.3514
ARIMA (1,1,1)	0.2485	0.4985	13.0136
ARIMA (6,1,6)	0.0913	0.3021	7.1210

The table shows the MSE, RMSE, and MAPE on the different models and realized that ARIMA (6,1,6) outperformed the other models at it has the Mean Square Error of 0.0913, Root Mean Square Error of 0.3021 and Mean Absolute Percentage Error of 7.1210. The analysis shows that ARIMA (6,1,6) has the best forecasting power to forecast the log return crude oil price. The parameters are estimated using the maximum likelihood estimation. This is shown in table 6.

Table 6: Estimation Results

Parameter	Value	Standard Error	TStatistic	P-Value
Constant	0.0051476	0.010945	0.47032	0.63813
AR{1}	0.50914	0.80695	0.63095	0.52808
AR{2}	-0.027774	0.097787	-0.28403	0.77639
AR{3}	0.10124	0.10012	1.0112	0.31191
AR{4}	-	0.13302	-	0.4426

	0.10212		0.76772	5
AR{5}	-0.72298	0.090734	-7.9681	1.6111e-15
AR{6}	0.36125	0.56304	0.64161	0.52113
MA{1}	-0.11283	0.83391	-0.13531	0.89237
MA{2}	-0.068942	0.28523	-0.2417	0.80901
MA{3}	-0.085719	0.14044	-0.61036	0.54163
MA{4}	0.047422	0.10096	0.46969	0.63858
MA{5}	0.84561	0.1024	8.2578	1.4834e-16
MA{6}	-0.33142	0.67384	-0.49184	0.62283
Variance	0.0055346	0.00065804	8.4106	4.0788e-17

The figure shows how the predicted price using quadratic regression and ARIMA (6,1,6) fit the actual data. Based on the graph, ARIMA (6,1,6) is better than quadratic regression as it fits the actual data smoothly.

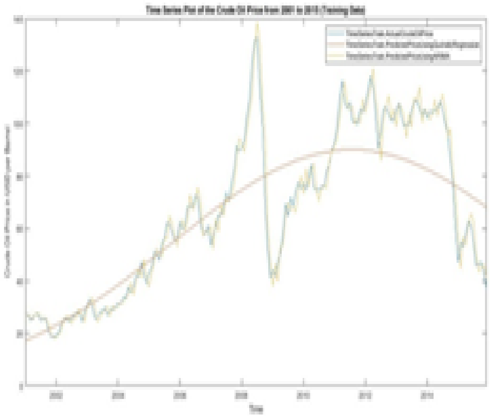


Figure 6: Time Series Plot of the Actual and Predicted Crude Oil Price Using Quadratic Regression and ARIMA (6,1,6).

Table 7: Comparison of the Forecasting Accuracy for the Three-Year Forecast

Model	MSE	RMSE	MAPE
Quadratic Regression	0.1367	0.3697	7.8203
ARIMA (6,1,6)	0.0913	0.3021	7.1210

The table shows how the two models performed in an out-sample data in forecasting the crude oil price. Analyzing the results of the forecasting accuracy in terms of MSE, RMSE, and MAPE Between quadratic regression and ARIMA (6,1,6), the latter is better in terms of forecasting accuracy with a mean square error of 0.0913, root mean square error of 0.3021, and a mean percentage absolute error of 7.1210.

Forecasted Values for the year 2019 – 2023.

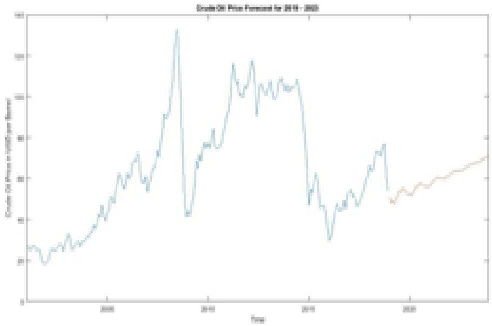


Figure 7: Five-Year Forecast of the Crude Oil Price of the Philippines.

Table 8: Forecasted Values for 2019 – 2023.

Month	Price	Month	Price	Month	Price
Jan-19	50.92	Sep-20	56.41	May-22	63.85
Feb-19	48.09	Oct-20	56.10	Jun-22	63.76
Mar-19	49.18	Nov-20	56.31	Jul-22	64.31
Apr-19	47.59	Dec-20	55.77	Aug-22	65.08
May-19	49.38	Jan-21	56.55	Sep-22	65.53
Jun-19	51.69	Feb-21	57.92	Oct-22	66.13
Jul-19	53.92	Mar-21	58.69	Nov-22	66.91
Aug-19	53.68	Apr-21	59.23	Dec-22	67.15
Sep-19	55.57	May-21	60.35	Jan-23	67.25
Oct-19	54.67	Jun-21	60.33	Feb-23	67.61
Nov-19	53.14	Jul-21	59.84	Mar-23	67.80
Dec-19	52.14	Aug-21	59.90	Apr-23	67.87

Jan-20	52.64	Sep-21	60.03	May-23	68.40
Feb-20	51.74	Oct-21	59.74	Jun-23	69.02
Mar-20	52.85	Nov-21	60.37	Jul-23	69.43
Apr-20	54.69	Dec-21	61.38	Aug-23	70.02
May-20	55.94	Jan-22	61.93	Sep-23	70.70
Jun-20	56.27	Feb-22	62.53	Oct-23	71.01
Jul-20	57.67	Mar-22	63.45	Nov-23	70.70
Aug-20	57.37	Apr-22	63.59	Dec-23	71.01
Jan-19	50.92	Sep-20	56.41	May-22	63.85

VII. CONCLUSION

The researchers were able to formulate a Regression and an ARIMA model for the 5-year forecast of the crude oil price of the Philippines. In regression analysis, it turned out that quadratic regression is the best form of polynomial regression that can forecast the crude oil price more accurately than the rest. On the other hand, the formulated ARIMA model is (6,1,6) after considering all the assumptions in ARIMA modeling as it outperformed all the candidate ARIMA models in terms of its forecasting accuracy. However, in analyzing the forecasting accuracy of the quadratic regression, and ARIMA (6,1,6) it turned out that the latter is the best model in forecasting the crude oil price. This study can help the government in implementing policies with regards to taxes impose on crude oil in the Philippines.

RECOMMENDATION

Since this research aims to forecast the Crude Oil Price of the Philippines, the researchers recommend to add more factors that can influence crude oil price to be able to formulate a more accurate model given the fact that oil price is volatile such as exchange rate, inflation rate, imports, exports, production and consumption of crude oil in the Philippines. It is also suggested for future researchers to use different methods to capture the volatility of the crude oil price.

REFERENCES

[1]. Uson M. (2018). Oil prices soar to record high – not because of TRAIN LAW. The Philippine Star. Retrieved from: http://www.philstar.com/opinion/2018/05/29/1819562/oil-prices-soar-record-high-not-because-train-law

[2]. Villanueva J. (2018). TRAIN's contribution to oil price hikes only around 20%. The Philippine News Agency. Retrieved from: https://www.pna.gov.ph/articles/1047501

[3]. Conde C. (2009). Oil and Gas Price Increases Meet Opposition in Philippines. The New York Times. Retrieved from: https://www.nytimes.com/2009/10/31/world/asia/31iht-oil.html

[4]. (2012) Oil Price Spiral. The Philippine Daily Inquirer. Retrieved from: https://opinion.inquirer.net/25265/oil-price-spiral

[5]. Gimena A.C. (2019). Revisiting the TRAIN Law. Business World. Retrieved from: https://www.bworldonline.com/revisiting-the-train-law/

[6]. Hamilton, James D. (2008) Understanding crude oil prices. University of California Energy Institute Working Paper, #EPE-023. Forthcoming in Energy Journal.

[7]. Afia Malik (2016). The Impact of Oil Price Changes on Inflation in Pakistan. International Journal of Energy Economics and Policy, 6(4), 727-737.

[8]. Yanagisawa, Akira (2012). Impact of Rising Oil Prices on the Macro Economy. Institute of Energy Economics, Japan. Retrieved from: https://eneken.ieej.or.jp/data/4338.pdf

[9]. Banhi Guha (2016) and Gautum Bandyopadhyay (2016). Gold Price Forecasting Using ARIMA Model. Journal of Advanced Management Science Vol. 4, No. 2, March 2016Mensah, Emmanuel K. (2015). Box-Jenkins Modeling and Forecasting of Brent Crude Oil Price. Department of Information Engineering.

[10]. Mensah, Emmanuel K. (2015). Box-Jenkins Modelling and Forecasting of Brent Crude Oil Price. Department of Information Engineering.

[11]. Shabri A. (2013). Crude Oil Forecasting with an Improved Model Based on Wavelet Transform and Linear Regression Model. 3rd International Conference on Applied Mathematics and Pharmaceutical Sciences (ICAMPS'2013).

[12]. Ahmed R.A., and Shabri A.B (2014). Daily Crude Oil Price Forecasting Model Using Arima, Generalized Autoregressive Conditional Heteroscedastic And Support Vector Machines. American Journal of Applied Sciences 11 (3): 425-432, 2014
[13]. Quan L. (2014). Daqing crude oil price forecast based on the ARIMA model. Bio Technology an Indian Journal. Volume 10 issue 24.
[14]. Urrutia J.D., Alano E.M.D., Aninipot P.M.R., Gumapac K. A., Quinto J.Q. (2014). Modelling and Forecasting Foreign Trade of the Philippines Using Time Series SARIMA Model. European Academic Research, Vol II, Issue 8/November 2014.
[15]. Urrutia J.D., de Guzman J.D., Mercado J., Bautista L., & Baccay E. (2015). Modeling the Philippines' Exchange Rate: A Regression Analysis. European Academic Research, Vol. III, Issue 3/June 2015.
[16]. Urrutia J.D., Mingo F.L.T., & Balmaceda C.N.M. (2015). Forecasting Income Tax Revenue of the Philippines Using Autoregressive Integrated Moving Average (ARIMA) Modeling: A time Series Analysis. American Research Thoughts, Vol. 1, Issue 9/ July 2015.
[17]. Urrutia J.D., Olfindo M.L.T, Mercado J., Baigan A.V.M., & Baccay E. (2015). A Mathematical Model for Estimating Imports and Exports of the Philippines: A Normal Estimation Equation for Multiple Linear Regression. European Academic Research, Vol. III, Issue 2/May 2015.
[18]. Urrutia J.D., and Olfindo M.L.T. (2015). Modelling and Forecasting the Exchange Rate of the Philippines. A Time Series Analysis. American Research Thoughts, Volume 1, Issue 9/July 2015.
[19]. Urrutia J.D., and Tampis R.L. (2017). Regression Analysis of the Economic Factors of the Gross Domestic Products in the Philippines. Journal of Fundamental and Applied Sciences. Available at: http://www.jfas.info
[20]. Tularam, G.A. and Saeed, T. (2016) Oil-Price Forecasting Based on Various Univariate Time-Series Models. American Journal of Operations Research, 6, 226-235. http://dx.doi.org/10.4236/ajor.2016.63023
[21]. Xin James He (2018). Crude Oil Prices Forecasting: Time Series vs. SVR Models. Journal of International Technology and Information Management Volume 27, Number 2 2018
[22]. Siddhi vinayak Kulkarni and Imad Haidar (2009). Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices. International Journal of Computer Science and Information Security, Vol.2, No.1, June 2009 Retrieved from:https://arxiv.org/ftp/arxiv/papers/0906/0906.4838.pdf
[23]. Zhang, G.P. (2003) Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neuro computing, 50, 159-175. http://dx.doi.org/10.1016/S09252312(01)00702-0
[24]. Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. Procedia - Social and Behavioral Sciences, 106, 234–240.
[25]. IIT Kanpur. Chapter 12. Polynomial Regression. Retrieved from: http://home.iitk.ac.in/~shalab/regression/Chapter12-Regression-PolynomialRegression.pdf
[26]. Drakos G. (2018). How to select the Right Evaluation Metric for Machine Learning Models: Part 1 Regression Metrics. Retrieved from: https://towardsdatascience.com/how-to-select-the-right-evaluation-metric-for-machine-learning-models-part-1-regrression-metrics-3606e25beae0
[27]. Lewis, C.D. (1982) International and Business Forecasting Methods. Butterworth's, London.
[28]. Ostertagová E.(2012). Modeling using polynomial regression. Proscenia Engineering 48 (2012) 500 – 506. Available at: https://doi.org/10.1016/j.proeng.2012.09.545
[29]. The ARIMA Procedure. Retrieved from: https://dms.umontreal.ca/~duchesne/chap7.pdf
[30]. Dickey D. and Fuller W. Likelihood ratio statistics for autoregressive time series with unit root. Econometrical 1981, 4(49) pp. 1057-1072.
[31]. ARIMA Model Retrieved from: https://people.duke.edu/~rnau/411arim.htm?fbclid=IwAR2mKeut76gLjOM8Q49Z8XKzV--6ZbGdw2_qu-ESHh_3Zl0-tJoZ9UqXZqU#pdq

Author's details

1.Secondary School Teacher I, Mathematics Department, Navotas National High School, National Capital Region, Philippines, gutierrezerrosjosephus@email.com

2.Secondary School Teacher I, Mathematics Department, Dr. Arcadio Santos National High School, National Capital Region, Philippines, karen.capiraldepablo@yahoo.com