Implementation of Autoregressive Integrated Moving Average for Time Series Forecasting

Aniket Chari[1], Dr Milind Sakhardande[2]

[1] *Student of Master's in Industrial Engineering, Mechanical Department, Goa College of Engineering, Goa, India*
[2] *Associate Professor, Mechanical Department, Goa College of Engineering, Goa, India*

ABSTRACT

One of the key factors for increasing profits and being competitive in a manufacturing industry is by managing the production cost. Apart from Time and Method Study, which is majorly dependent on manufacturing activities, another aspect of reducing production costs is inventory management. To manage inventory, one needs to know the future demands, material lead time and total inventory cost. Inventory management in terms of inventory cost becomes even more challenging when high-value raw materials are involved whose prices are highly volatile and are fluctuating in a very short span of time. The work presented in this article constitutes an attempt to capture the pattern of fluctuation of raw material price over time using past data and establish a forecasting model to predict the future price of the material by implementing a time series analysis methodology. In our work, we have used past price data and utilized this data to model and forecast future prices. The Box–Jenkins methodology (identification, implementation, Diagnostic and Forecasting) was used on the past data to develop an autoregressive integrated moving average (ARIMA) model. Based on the results obtained from the developed ARIMA model and the model's MAPE we believe that the model can be used for predicting the future price of any raw material. The output of such a prediction model will provide procurement managers of the manufacturing industry with some logical backup for taking decisions related to inventory management.

Keyword: - Time Series Analysis, Prediction Model, Forecasting, Box–Jenkins methodology, ARIMA, Raw Material Price

1. INTRODUCTION

At present time, we are having a very competing manufacturing environment. Manufacturers need to act in response proactively to the fluctuating demand and one of the crucial focuses is trying to capture more accurate demand. This is changing the market to a "pull" environment where the customers are selecting the suppliers based on their demands to the supplier at what price they want the products and what should be the delivered period. As crucial as Demand forecasting is to inventory management so is the price forecasting of raw materials, especially for the high-value raw materials whose prices are fluctuating over a very short span of time. Inventory stock levels can be dependent on such price forecasts. If we forecast that the prices are going to increase then we can go for forward buying, if the prices are going to reduce then hold the raw material procurement and decide on the final product price and delivery commitments to the customers. Such decisions can only be made if we have an accurate price forecast. However, an inaccurate forecast can result in significant costs increase instead of the cost reduction. As a result, many organizations may go for large investments in inventories anticipating the price escalation or may be liable for a penalty on delivery commitments by delaying the raw material procurement process. In some instances, the prices can be intermittent prices, i.e., there are times when the prices remain constant for a long period of time and other times when there is a sudden surge in price attributed to unforeseen calamities like the COVID Pandemic or the Russia-Ukraine War, which is a further complicating issue. Such Intermittent price fluctuations present many challenges for traditional statistical forecasting techniques. Commonly, there are many approaches for forecasting some of which are naive forecasting, exponential smoothing, regression approach, etc. Some more techniques which can be used for the prediction of materials prices are Fuzzy Logic, Statistical Methods (including Regression Analysis, MONTE CARLO method, and ANOVA), Artificial Neural Networks and Trend Analysis. For applying most of these approaches, we need to have historical data.

For real-word situations, the linear statistical forecasting methods have been generally used as they are simple to develop, implement and interpret. However, Linear models have limitations because many real-world problems are nonlinear. Linear models can be used for short-period forecasting where the accuracy is not of high significance. There may be no single linear forecasting technique which individually prevails all data sets over all situations as there is always some extent of nonlinearity in the data which the linear statistical techniques will not be able to capture. Recently, researchers have done a lot of work in the forecasting domain and concluded with many methods among which we found two key approaches largely utilized: artificial neural network (ANN) and time series analysis techniques. Neural Network models have been successfully used in forecasting. Neural Networks can capture the nonlinearity in a data set. Many successful applications have shown that neural networks can be very useful tool for time-series modelling and forecasting. Neural networks are data-driven methods with few prior assumptions about the underlying models. Instead, neural networks have the capability to identify the underlying functional relationship within the data.

In a manufacturing company which uses high-value raw materials, the price forecasts of these materials are of great relevance. Certainly, predicting the prices facilitates the decision on how much to procure and when to procure. Such forecasts are of most significance to small-scale industry which uses high-value raw material as procurement of such raw material can directly affect their cash flow, productivity, and profitability. In our case, we have considered an electrocoating industry which deals with coatings and one of the high-value coating raw materials being used is platinum. Hence, we made an attempt to develop an ARIMA model to predict the future prices of platinum.

2. LITERATURE REVIEW

In today's organizations, forecasting is becoming very significant as many crucial decisions are depending on these forecasts. The science of estimating the future level of some variables is known as a forecast. Usually, the variable being forecasted is the demand but, in our work, this variable is the purchasing price of raw materials [1][2].

For a supply chain management system to be robust and efficient there must be a collaboration between different departments of an organization: planning, procurement, manufacturing, and logistics. Based on an accurate forecast one much devise an optimal procurement plan to minimize the total production cost, of which, two of the components are the procurement cost and inventory holding costs. Because of using these forecasts, one can expect benefits like economized inventories, decreased supply chain costs, better return on assets, improved cash flow management, higher customer satisfaction, and reduced final product lead times. We need to also keep in mind that this optimal procurement plan should also be in line with different company policies among others like production capacity, minimum production lots, etc.

Gaafar and Choueiki (2000) applied an ANN model in material requirement planning for a lot-sizing problem in the case of deterministic time-varying demand [3]. Sustrova, T. (2016) used the ANN models in business processes, especially in inventory management. The developed model was utilized for the optimization of inventory levels to better the ordering system and inventory management [4]. Prybutok et al. (2000) conducted a study on electricity demand to compare and assess the performance of ARIMA and ANN methods and forecast the time series [5]. S. L. Ho et al. (2002) conducted a study using ARIMA and neural networks on the simulated failure time of a compressor [6]. They have discussed the predictive performances of the proposed models. Karin K. (2012) utilized Artificial Neural Networks (ANN), support vector machine (SVM) and ARIMA to predict the demand for consumer products. In terms of MAPE, SVM was superior in forecasting demand [7]. Aburto et al. (2007) have described their development of a hybrid intelligent system for demand forecasting by combining ARIMA and neural networks [8]. Mitrea et al. (2009) have compared the forecasting accuracy of Moving Average (MA) and Autoregressive Integrated Moving Average (ARIMA) with Neural Networks (NN) models as Feed-forward NN and Nonlinear Autoregressive networks with Exogenous inputs (NARX) and their results show forecasting with NN offers better forecasting accuracy [9]. Catalao et al. (2006) have proposed a neural network approach to forecast next-week prices in the electricity market of mainland Spain [10]. Contreras et al. (2002) and Conejo (2005) used ARIMA methodology to forecast electricity prices [11] [12].

As shown in the brief literature review above, Artificial Neural Network is a strong tool for the modelling of any time series. However, in this article, we have developed and tested the Autoregressive Integrated Moving Average first to demonstrate its ability to make accurate forecasts of a high-value raw material price as a primary study.

A time series can be defined as an observation that is recorded in chronological order of time [1]. In Time series analyses for developing forecasting models, one applies mathematical techniques that utilize historical data as an input to output the forecasting variable. Time series methodology is founded on the proposition that the future is a chaotic amplification of the past. To predict the future, the challenge lies in capturing the chaotic amplification and interpreting it into a mathematical model using known techniques or developing new techniques. There has been much research to develop a time series forecasting model using various techniques. Their research interest was forecasting the sales of food products, tourism, maintenance repair parts, electricity prices, automobiles, and some other products and services [14] [15] [16] [17] [18] [19] [20].

In our present study, we have used the Box–Jenkins time series approach, specifically the Autoregressive integrated moving average, to model and forecast the price of a raw material price. For this, we have used the price data of Platinum from 7th May 2018 to 29th April 2022 (1 - 1024 Days). Several possible ARIMA models were estimated using AC and PAC and evaluated by four performance criteria: Akaike criterion (AIC), Schwarz Bayesian criterion (SBC), Hannan-Quinn Criterion, and Adjusted R Square Value. The adequate model was diagnosed using a Q-Statistical test and new prices were forecasted for the next 600 days (1024 – 1624 Days).

3. METHODOLOGY

3.1 Autoregressive Integrated Moving Average (ARIMA)

The autoregressive integrated moving average (ARIMA) model is extensively utilized for time series analysis due to its flexibility. Usually, the variable to be predicted in Time series is demand but, in our work, we have used the ARIMA technique to predict the price of a material. Kurawarwala and Matsuo (1998) have used the ARIMA technique to estimate the seasonal variation of demand by using past data and validated the models by examining the forecast performance [21]. Miller and Williams (2003) blended the seasonal factors in their research of improving forecasting accuracy [22]. There are many drawbacks to the classical approach of ARIMA. It becomes very difficult to identify an ARIMA model if the seasonal adjustment order is high or when its diagnostics fail to indicate that the time series is stationary after seasonal adjustment.

Generally, for a stationary time series an ARMA model is developed. ARMA is a combination of autoregressive and moving averages. The autoregressive (AR) components capture the autocorrelation between time series. It is also known as the long memory model, and it represents the value at the current time instant in terms of the values at the previous time instant depending on the order or lag 'p'. The 1st order Autoregressive model, AR (1), is given by equation (1). The moving average (MA) component, also known as the short memory model, corresponds to the deviation of the series at the current time instant from its mean value as a linear combination of errors in the past time instants depending on the order or lag 'q'. The 1st order Moving average model, MA (1), is given by equation (2). ARMA models can be used to describe only stationary time series.

A time series is said to be stationary if the mean and variance of the underlying process remain constant. If the series is non-stationary, then we must use an ARIMA model with order (p, d, q), in which the non-stationary series is transformed into stationary series by taking the difference or log difference between consecutive time instants d times.

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t \quad (1)$$

Where, ε_t is white noise and α_1 is the parameter of the 1st order AR model.

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (2)$$

Where, ε_t is white noise and θ_1 is the parameter of the MA model.

The integrated process: The I (integrated) part of ARIMA is when the time series is non-stationary and needs to be transformed to make it a stationary series. To do this we replaced the raw data values with the difference values of d order from one observation to the next. Usually, the differences obtained are comparatively lower than the raw data or even fluctuate around a mean value for a process observed at various time intervals. A 1st Order differentiation implies that the difference is taken between two successive values of Y from the raw data as shown by equation (3).

$$Y_t = Y_{t-1} + \varepsilon_t \quad (3)$$

Where ε_t is white noise.

The three iterative stages of the Box–Jenkins methodology are – the identification of the model, estimation of the model parameters, and diagnostic checking steps [23]. The principal rule to identify the ARIMA model is that it should have some theoretical autocorrelation properties. By comparing the theoretical and empirical autocorrelation patterns, we try to identify one or several potential ARIMA models for the given time series. For the identification of the ARIMA models, Box–Jenkins suggested the use of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data [23]. Since in almost all situations, the raw data is non-stationary and hence, to identify the ARIMA models, we will have to transform the raw data to make it stationary.

As far as the stationarity of a time series is concerned, the statistical characteristics such as the autocorrelation structure and the mean of the series remain constant over time. Primarily, to remove trends and stabilize the variance from the raw data, we generally implement the differencing and power transformation. After that, we simply estimate the model parameters and then specify the model. For model estimation, usually, the maximum likelihood algorithm is used such that the overall error is reduced. Further, we use, the Akaike criterion (AIC), Schwarz Bayesian criterion (SBC), Hannan-Quinn Criterion, and Adjusted R Square Values to select the best possible model fitting to our sample data. Lastly, in the diagnostic stage, we check if our selected model satisfies the requirements of a stable univariate process. For this, we check if the residuals of the model are white Noise using the Ljung-Box Q Statistic test and we check the ARMA structure using AR and MA roots. If the model is not satisfying the diagnostic benchmarks, then we will have to repeat the process of model selection and parameter estimation by selecting different orders of the model. Box–Jenkins approach lets us reach a high degree of satisfaction with the model due to its iterative process converging to reduced errors. After completing the Box- Jenkins methodology we can use the final model to forecast our variable, which is the price in our case.

3.2 Case Study on Platinum Price

In this article, the price forecasting of the high-value raw material, i.e., of platinum, for an electro-coating industry is attempted based on real past data. This study investigates the effectiveness of price forecasting using time series analysis - ARIMA. Following the Box–Jenkins approach, our work has been implemented in three parts: identification, estimation, and diagnostic. The model has been developed based on the price of Platinum from 7th May 2018 to 29th April 2022, i.e., 1 - 1024 time periods or days as shown in Fig -1:.

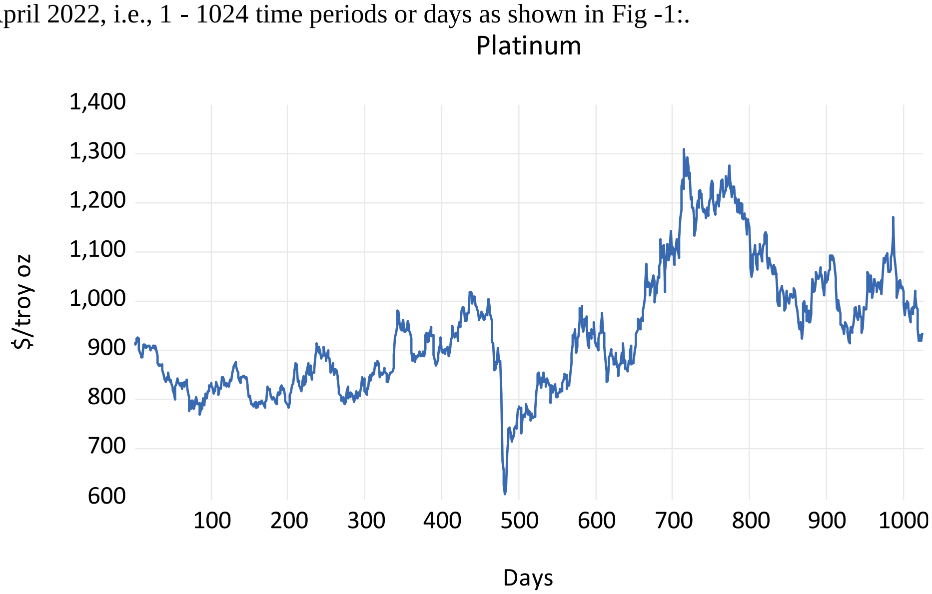


Fig -1: Historical Price Variation of Platinum from May 2018 to April 2022

3.2.1 Identification of Model

In the first step, we checked the stationarity of the raw data followed by the initial pre-processing of the raw data to transform it to stationary data, and then we selected all the possible values of p and q which we would fine-tune as the model fitting progresses. As described in the previous section, the time series is said to be stationary if its mean and variance remain constant across time. If there is a strong trend or seasonality observed in the data, then the data is believed to be non-stationary. To check the stationarity of the data, three checks were done. First was the graphical check – from Fig -1:, we can see from the graph a slightly increasing trend in the data, variance changes across time and also mean is not constant across time which implies data is non-stationary.

Next, a correlogram test was done on the data using EViews Software. For Stationarity, the correlogram test of the series is shown in Fig -2:. We can see that the Autocorrelation (AC) decays but not in an immediate way and there is a significant lag in Partial Autocorrelation (PAC). Also, all the p-values are less than 0.05 suggesting data is non-stationary.

Sample: 1 1024
Included observations: 1024

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.991	0.991	1008.2	0.000
		2	0.982	-0.009	1998.6	0.000
		3	0.972	-0.032	2970.3	0.000
		4	0.962	-0.002	3923.6	0.000
		5	0.953	0.035	4860.1	0.000
		6	0.945	0.062	5782.3	0.000
		7	0.939	0.053	6692.3	0.000
		8	0.933	0.028	7591.5	0.000
		9	0.928	0.055	8482.1	0.000
		10	0.923	0.002	9364.0	0.000
		11	0.917	-0.054	10235.	0.000
		12	0.911	0.016	11097.	0.000
		13	0.904	-0.063	11945.	0.000
		14	0.895	-0.066	12779.	0.000
		15	0.887	0.007	13598.	0.000
		16	0.878	-0.025	14402.	0.000
		17	0.869	-0.013	15191.	0.000
		18	0.861	0.017	15965.	0.000
		19	0.854	0.015	16727.	0.000
		20	0.847	0.000	17477.	0.000
		21	0.840	0.004	18216.	0.000
		22	0.833	0.011	18944.	0.000
		23	0.827	0.002	19662.	0.000
		24	0.820	0.007	20369.	0.000
		25	0.815	0.042	21067.	0.000
		26	0.809	-0.000	21756.	0.000
		27	0.802	-0.012	22434.	0.000
		28	0.795	-0.023	23101.	0.000
		29	0.789	0.001	23758.	0.000
		30	0.781	-0.051	24402.	0.000
		31	0.773	0.007	25035.	0.000
		32	0.766	0.021	25657.	0.000
		33	0.759	-0.052	26267.	0.000
		34	0.752	0.039	26867.	0.000
		35	0.745	-0.040	27456.	0.000
		36	0.738	-0.001	28035.	0.000

Fig -2: Correlogram Test to check stationarity of raw data

Finally, a formal test, i.e., Standard Unit Root Test or Augmented Dickey-Fuller Test was conducted on the series. After carrying out the test on the EViews software, the results are shown in Fig -3:. The test hypothesis was as follows, the null hypothesis, H0 was "Data has Unit Root" and the alternate hypothesis H1 was "The series does not have a unit root. The series is stationary."
The calculated p-value was found to be 0.2168 which is crossing the threshold significance level α (0.05), hence, the test fails to reject the null hypothesis.

Null Hypothesis: PLATINUM has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.172172	0.2168
Test critical values:	1% level	-3.436523	
	5% level	-2.864154	
	10% level	-2.568214	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PLATINUM)
Method: Least Squares
Date: 05/17/23 Time: 13:58
Sample (adjusted): 2 1024
Included observations: 1023 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PLATINUM(-1)	-0.009188	0.004230	-2.172172	0.0301
C	8.626789	3.998871	2.157306	0.0312

R-squared	0.004600	Mean dependent var	0.018573
Adjusted R-squared	0.003625	S.D. dependent var	17.13514
S.E. of regression	17.10406	Akaike info criterion	8.518462
Sum squared resid	298692.3	Schwarz criterion	8.528101
Log likelihood	-4355.193	Hannan-Quinn criter.	8.522121
F-statistic	4.718331	Durbin-Watson stat	1.981007
Prob(F-statistic)	0.030072		

Fig -3: Standard Unit Root Test or Augmented Dickey-Fuller Test on raw data

In our case, after testing the raw data for the stationarity using the tests described above, we concluded that our model is not true AR or true MA and hence, we would be working with the AR(I)MA (p, d, q) model. We, therefore, transformed the data by taking the difference between consecutive series values.

3.2.2 Estimation

We took the first Difference (d=1) and the transformed data after 1st difference was tested for stationarity using Augmented Dickey-Fuller Test as shown in Fig -4:. In this case, as the p-value was < 0.05, we could reject the null hypothesis. Hence, we would be working with the 1st Difference Stationary time series. With $d = 1$ of the ARIMA (p, d, q), we identified several possible models using the AC and PAC.

Null Hypothesis: D(PLATINUM) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=21)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-31.78207	0.0000
Test critical values:	1% level	-3.436530	
	5% level	-2.864157	
	10% level	-2.568215	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PLATINUM,2)
Method: Least Squares
Date: 05/17/23 Time: 14:44
Sample (adjusted): 3 1024
Included observations: 1022 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(PLATINUM(-1))	-0.995507	0.031323	-31.78207	0.0000
C	0.023470	0.536494	0.043747	0.9651

R-squared	0.497562	Mean dependent var	0.020548
Adjusted R-squared	0.497069	S.D. dependent var	24.18446
S.E. of regression	17.15104	Akaike info criterion	8.523950
Sum squared resid	300041.4	Schwarz criterion	8.533597
Log likelihood	-4353.738	Hannan-Quinn criter.	8.527612
F-statistic	1010.100	Durbin-Watson stat	1.999200
Prob(F-statistic)	0.000000		

Fig -4: Standard Unit Root Test or Augmented Dickey-Fuller Test For 1st Difference Data

From the Correlogram of the transformed stationary data as shown in Fig -5:, with AC, we can decide the MA component (q) and with PACF we will determine the AR component (p) and as we have selected 1st Difference as Stationary data, d is 1. Hence, we identified eight ARIMA Models as shown in **Error! Reference source not found.**

Sample (adjusted): 2 1024
Included observations: 1023 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.004	0.004	0.0207	0.886
		2	0.027	0.027	0.7773	0.678
		3	-0.002	-0.002	0.7809	0.854
		4	-0.039	-0.040	2.3573	0.670
		5	-0.067	-0.067	6.9845	0.222
		6	-0.061	-0.058	10.761	0.096
		7	-0.036	-0.032	12.064	0.098
		8	-0.059	-0.059	15.713	0.047
		9	-0.001	-0.006	15.715	0.073
		10	0.056	0.050	18.984	0.040
		11	-0.008	-0.018	19.044	0.060
		12	0.072	0.058	24.493	0.017
		13	0.071	0.062	29.737	0.005
		14	-0.005	-0.012	29.765	0.008
		15	0.021	0.021	30.228	0.011
		16	0.002	0.009	30.233	0.017
		17	-0.034	-0.021	31.435	0.018
		18	-0.038	-0.019	32.972	0.017
		19	-0.015	-0.004	33.204	0.023
		20	-0.020	-0.009	33.610	0.029

Fig -5: Correlogram for Stationary Data

Table -1: Possible ARIMA Models

P (AR)	Q (MA)	d
5	5	1
5	6	1
5	12	1
5	13	1
13	5	1
13	6	1
13	12	1
13	13	1

Now, that we have identified the eight potential ARIMA models, the next step was to estimate the model coefficients for all 8 models. The ARIMA procedure in the Eviews time series software estimates the coefficients of the (p, q, d) model equation using a maximum likelihood estimation algorithm. Fig -6: shows the ARIMA (13, 1, 12) model's equation estimation in Eviews which was selected as the best model as explained in the next paragraph.
For the best model selection, we compared the significance of the ARMA Components along with the Adjusted R-Squared Value (higher value is better) and Akaike, Schwartz and Hannan-Quinn Values (lower value is better) which is summarized in **Error! Reference source not found.**
AR and MA components are significant if the corresponding p-value is less than 0.05 and the model which has the lowest value for the Akaike, Schwartz and Hannan-Quinn criterion will be the best.

Dependent Variable: D(PLATINUM)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 05/18/23 Time: 09:32
Sample: 2 1024
Included observations: 1023
Convergence achieved after 12 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.010491	0.651182	0.016111	0.9871
AR(13)	0.070933	0.025733	2.756550	0.0059
MA(12)	0.078891	0.029627	2.662814	0.0079
SIGMASQ	290.1389	7.631580	38.01819	0.0000

R-squared	0.010866	Mean dependent var	0.018573
Adjusted R-squared	0.007954	S.D. dependent var	17.13514
S.E. of regression	17.06686	Akaike info criterion	8.516194
Sum squared resid	296812.1	Schwarz criterion	8.535473
Log likelihood	-4352.033	Hannan-Quinn criter.	8.523513
F-statistic	3.731371	Durbin-Watson stat	1.998584
Prob(F-statistic)	0.011000		

Inverted AR Roots	.82	.72-.38i	.72+.38i	.46+.67i
	.46-.67i	.10-.81i	.10+.81i	-.29-.76i
	-.29+.76i	-.61+.54i	-.61-.54i	-.79-.20i
	-.79+.20i			
Inverted MA Roots	.78-.21i	.78+.21i	.57-.57i	.57+.57i
	.21-.78i	.21+.78i	-.21+.78i	-.21-.78i
	-.57-.57i	-.57-.57i	-.78-.21i	-.78+.21i

Fig -6: ARIMA (13, 1, 12) Model

As we can see from **Error! Reference source not found.**ARIMA Model with $p = 13$ and $q = 12$ has the lowest values for the Akaike, Schwartz and Hannan-Quinn criterion. Also, the Adj R-Sq value for this model is the highest compared to the rest of the possible models. Hence, we identified ARIMA (13, 1, 12) as our best model. The next step would be to diagnose this model for best fit.

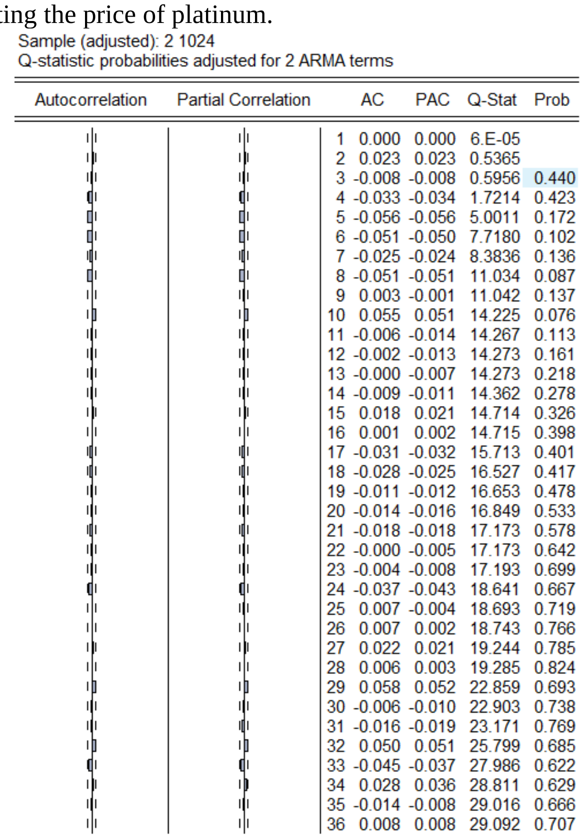
Table -2: Model Selection Criteria

d	AR (p)	MA (q)	Lower is Better			Higher is Better	p-Value < 0.05	
			AC	SC	HQ	Adj R-Sq	AR(p)	MA(q)
1	5	5	8.521	8.540	8.529	0.003	0.260	0.366
1	5	6	8.519	8.538	8.526	0.005	0.007	0.016
1	5	12	8.517	8.537	8.525	0.007	0.014	0.013
1	5	13	8.518	8.538	8.526	0.006	0.016	0.011
1	13	5	8.519	8.538	8.526	0.006	0.008	0.029
1	13	6	8.519	8.538	8.526	0.005	0.006	0.025
1	13	12	8.516	8.535	8.524	0.008	0.006	0.008
1	13	13	8.522	8.541	8.529	0.002	0.443	0.579

3.2.3 Diagnostic

Now that we have our potential best-fit model, we checked if this model satisfies the requirements for a stable univariate process. For this, we checked if the residuals of the model are white Noise using the Ljung-Box Q Statistic. As can be seen in **Error! Reference source not found.**, Autocorrelation and Partial Correlation are within the standard error lines and the p-values are bigger than 0.05 which implies that we cannot reject the null hypothesis which is "Residuals are White noise" for the Ljung-Box Q Statistic test. Next, we check if the estimated ARMA process is stationary. For this, we checked if the AR roots lie inside the unit circle. As can be seen in **Error! Reference source not found.**, the AR Roots lie within the Unit circle hence the ARMA process is stationary.

The last thing that we checked was if the ARMA process is invertible, for this the MA roots should lie within Unit Circle which is also true as seen in **Error! Reference source not found.**. Now that these conditions are satisfied, we can use this model for forecasting the price of platinum.



Sample (adjusted): 2 1024
Q-statistic probabilities adjusted for 2 ARMA terms

	AC	PAC	Q-Stat	Prob
1	0.000	0.000	6.E-05	
2	0.023	0.023	0.5365	
3	-0.008	-0.008	0.5956	0.440
4	-0.033	-0.034	1.7214	0.423
5	-0.056	-0.056	5.0011	0.172
6	-0.051	-0.050	7.7180	0.102
7	-0.025	-0.024	8.3836	0.136
8	-0.051	-0.051	11.034	0.087
9	0.003	-0.001	11.042	0.137
10	0.055	0.051	14.225	0.076
11	-0.006	-0.014	14.267	0.113
12	-0.002	-0.013	14.273	0.161
13	-0.000	-0.007	14.273	0.218
14	-0.009	-0.011	14.362	0.278
15	0.018	0.021	14.714	0.326
16	0.001	0.002	14.715	0.398
17	-0.031	-0.032	15.713	0.401
18	-0.028	-0.025	16.527	0.417
19	-0.011	-0.012	16.653	0.478
20	-0.014	-0.016	16.849	0.533
21	-0.018	-0.018	17.173	0.578
22	-0.000	-0.005	17.173	0.642
23	-0.004	-0.008	17.193	0.699
24	-0.037	-0.043	18.641	0.667
25	0.007	-0.004	18.693	0.719
26	0.007	0.002	18.743	0.766
27	0.022	0.021	19.244	0.785
28	0.006	0.003	19.285	0.824
29	0.058	0.052	22.859	0.693
30	-0.006	-0.010	22.903	0.738
31	-0.016	-0.019	23.171	0.769
32	0.050	0.051	25.799	0.685
33	-0.045	-0.037	27.986	0.622
34	0.028	0.036	28.811	0.629
35	-0.014	-0.008	29.016	0.666
36	0.008	0.008	29.092	0.707

Fig -7: Ljung-Box Q Statistic Test for White Noise

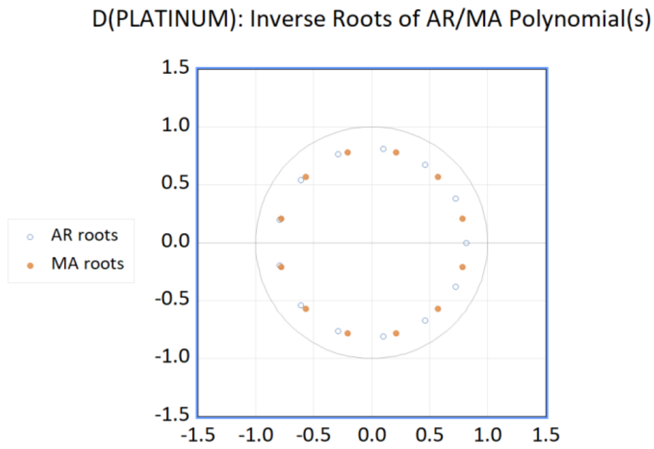


Fig -8: ARMA Structure - AR Roots and MA Roots

3.2.3 Forecasting

Now that we have found our ARIMA model, we forecasted the platinum prices from day 1024 to 1624 (600 days) as shown in **Error! Reference source not found.**. We calculated accuracy measures for our model like Root Mean Squared Error, RMSE = 16.03020, Mean Absolute Error = 16.03020, Mean Absolute Percent Error and MAPE = 1.712628.

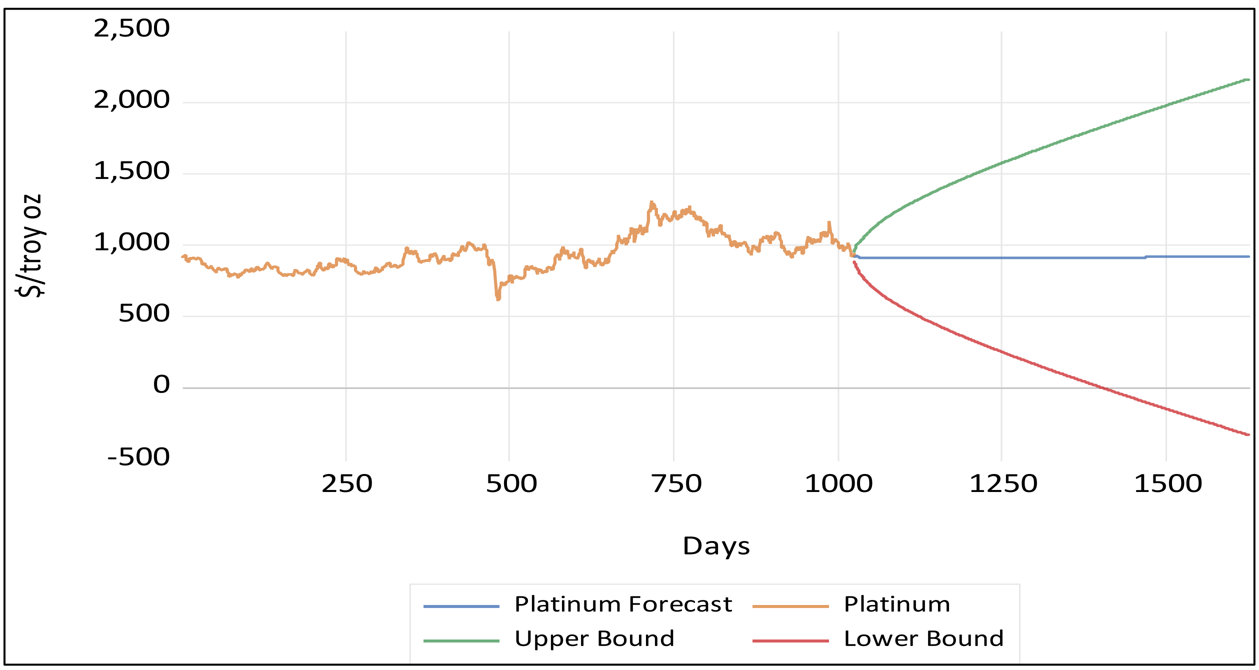


Fig -9: Forecast for next 600 Days

4. CONCLUSIONS

In this paper, we have attempted to use the Box–Jenkins procedure for ARIMA modelling to forecast the prices of platinum metal. The Box–Jenkins procedure for the ARIMA model has been discussed in detail. First, the four-year, past data is collected and checked if the data is stationary. Since the data was nonstationary, data is transformed to 1st Difference and found that the 1st Difference data is stationary. Using the transformed data, all the possible ARIMA models are identified, and the best model is selected by comparing the Akaike, Schwartz and Hannan-Quinn criteria

(a smaller one is better). After selecting the best potential model for forecasting, the selected model is first diagnosed to ensure it satisfies the requirements of a stable univariate process. To do this, we checked the residuals of the model are white noise (Ljung-Box Q statistics), and then we checked if the estimated ARIMA process is stationary and invertible. After the above criteria are satisfied for the selected model, the model has been used to forecast the prices for the next 600 Days.
In our future work, we shall be attempting to develop an ANN model for price forecasting and a Hybrid model combining ARIMA & ANN.

5. REFERENCES

[1] Bozarth CB and Handfield RB. Introduction to operations and supply chain management, 4th ed. Raleigh: North Carolina State University, 2016
[2] Wisner JD, Tan KC, and Leong GK. Principles of supply chain management: a balanced approach. Mason: Thomson South-Western, 2011
[3] Gaafar, L. K., and Choueiki, M. H. (2000). A neural network model for solving the lot-sizing problem. Omega, 28(2), 175–184.
[4] Sustrova, T. (2016). An Artificial Neural Network Model for a Wholesale Company's Order-Cycle Management. International Journal of Engineering Business Management, 8, 2.
[5] Prybutok VR, Yi J, and Mitchell D. (2000). Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. Eur J Oper Res; 122(1): 31–40.
[6] Ho, S., Xie, M., and Goh, T. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. Computers and Industrial Engineering, 42(2-4), 371–375.
[7] Kandananond, K. (2012). A Comparison of Various Forecasting Methods for Autocorrelated Time Series. International Journal of Engineering Business Management, 4, 4.
[8] Aburto, L., and Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. Applied Soft Computing, 7(1), 136–144.
[9] Mitrea, C. A., Lee, C. K. M., and Wu, Z. (2009). A Comparison between Neural Networks and Traditional Forecasting Methods: A Case Study. International Journal of Engineering Business Management, 1, 11.
[10] Catalão, J. P. S., Mariano, S. J. P. S., Mendes, V. M. F., and Ferreira, L. A. F. M. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. Electric Power Systems Research, 77(10), 1297–1304.
[11] Contreras, J., Espinola, R., Nogales, F. J., and Conejo, A. J. (2002). ARIMA Models to Predict Next-Day Electricity Prices. IEEE Power Engineering Review, 22(9), 57–57.
[12] Conejo, A. J., Plazas, M. A., Espinola, R., and Molina, A. B. (2005). Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models. IEEE Transactions on Power Systems, 20(2), 1035–1042.
[13] Bahi, Y. F. E., Ezzine, L., El Moussami, H., and Aman, Z. (2018). Modeling and Forecasting of Fuel Selling Price Using Time Series Approach: Case Study. 2018 5th International Conference on Control, Decision and Information Technologies (CoDIT).
[14] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., and Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10, 184797901880867.
[15] Song, H., and Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. Tourism Management, 29(2), 203–220.
[16] Willemain, T. R., Smart, C. N., and Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. International Journal of Forecasting, 20(3), 375–387.
[17] Taylor, J. W., de Menezes, L. M., and McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. International Journal of Forecasting, 22(1), 1–16.
[18] Matsumoto, M., and Ikeda, A. (2015). Examination of demand forecasting by time series analysis for auto parts remanufacturing. Journal of Remanufacturing, 5(1).
[19] Qingjie Liu, Huachun Wang. (2015). Research on the Forecast and Development of China's Public Fiscal Revenue Based on ARIMA Model. Theoretical Economics Letters > Vol.5 No.4
[20] Shichang Shen, Yue Shen. (2016). ARIMA Model in the Application of Shanghai and Shenzhen Stock Index. Applied Mathematics 07(03):171-176
[21] Kurawarwala, A. A., and Matsuo, H. (1998). Product Growth Models for Medium-Term Forecasting of Short Life Cycle Products. Technological Forecasting and Social Change, 57(3), 169–196.
[22] Miller, D. M., and Williams, D. (2003). Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. International Journal of Forecasting, 19(4), 669–684.
[23] Kendall, M. G., Box, G. E. P., and Jenkins, G. M. (1971). Time Series Analysis, Forecasting and Control. Journal of the Royal Statistical Society. Series A (General), 134(3), 450.