3/30/2021

Using ARIMA to Forecast Copper Future Prices

Applied Econometrics

TYBSc Finance A: Group 7
ISHAN PENDSE: A018
JANIT SINGH: A019
JANVI MEHTA: A020
PRATIK DOSHI: A030
REYAAN WANCHOO: A033

Contents

About the Data 2
Testing for Autocorrelation 2
Test for Stationarity 3
ACF and PACF on First Difference 6
Research Objective 7
ARIMA 7
ACF & PACF on Residuals 9
Conclusion and Interpretation 10

About the Data

The prices of futures contract on copper traded on MCX (Multi Commodity Exchange) from December 2011 to March 2021 have been used as the dataset in this report. The prices are expressed in INR. The data is sourced from Investing.com.

Testing for Autocorrelation

Autocorrelation refers to the correlation between a time series variable with its lagged self. Autocorrelation in residuals is a challenge when dealing with time series based regression models. In order to conduct an autocorrelation test, a simple time series regression has been estimated using the natural logarithm of the copper future prices as the dependent variable and a serially incrementing variable (starting from 1) as the independent variable to represent time. We then use the BG Test to check for autocorrelation in the residuals of the regression expressed above. The BG statistic follows a Chi-squared distribution with the number of lags as the degrees of freedom. Here, we use the critical value approach to check for autocorrelation up to 16 lags. A summary of the regression fit and result of BG Test is given below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.30076519
R Square	0.090459699
Adjusted R Square	0.090094275
Standard Error	0.137255541
Observations	2491

ANOVA

	df	SS	MS	F	Significance F
Regression	1	4.663563153	4.663563	247.5472	3.00492E-53
Residual	2489	46.89047894	0.018839		
Total	2490	51.55404209			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	5.955371157	0.005501787	1082.443	0	5.944582606	5.966159708	5.944582606	5.966159708
Time	6.01712E-05	3.82437E-06	15.73363	3E-53	5.2672E-05	6.76705E-05	5.2672E-05	6.76705E-05

Figure 1: Time Series Regression Fit

Lag	BG STAT	BG Crit	Significant at 5%
1	2485.2573	3.8415	TRUE
2	2484.2597	5.9915	TRUE
3	2483.2634	7.8147	TRUE
4	2482.2662	9.4877	TRUE
5	2481.2682	11.0705	TRUE
6	2480.2701	12.5916	TRUE
7	2479.2720	14.0671	TRUE
8	2478.2764	15.5073	TRUE
9	2477.2784	16.9190	TRUE
10	2476.2805	18.3070	TRUE
11	2475.2850	19.6751	TRUE
12	2474.2869	21.0261	TRUE
13	2473.2889	22.3620	TRUE
14	2472.2916	23.6848	TRUE
15	2471.2935	24.9958	TRUE
16	2470.2960	26.2962	TRUE

Figure 2: Result of BG Test using Multiple Lags

Test for Stationarity

A data generating process is said to be stationary when its statistical properties do not change with time. In other words, a stationary series is one whose mean, variance or autocorrelation structure does not vary with time. Graphically, it has a randomly fluctuating and horizontal plot (refer to figure 4). In order to check for stationarity, the Augmented Dickey-Fuller (ADF) test is used. Here, we run an ADF Test on the log of copper future prices to check if the series is stationary. The series is found to be non-stationary in nature. We then take a first difference of the series and recheck for stationarity. The first difference of the log of copper prices is found to be stationary in nature. Since we have arrived at a stationary series by taking a difference once, the series is said to be integrated of order 1 or I(1). Below is the result of the ADF Tests along with a visual representation of the series and its first difference.

ADF Test on Log of Copper Future Prices

Criteria	Schwert
Drift	yes
Trend	yes
Lag	27
Alpha	0.05
Tau-stat	-1.0197
Tau-crit	-3.41211
Stationary	no
AIC	-6.07668
BIC	-6.06733
Lags	1
Coefficient	-0.00173
p-value	> .1

ADF Test on First Difference

Criteria	Schwert
Drift	no
Trend	no
Lag	27
Alpha	0.05
Tau-stat	-51.2616
Tau-crit	-1.94111
Stationary	yes
AIC	-6.07756
BIC	-6.07522
Lags	0
Coefficient	-1.02739
p-value	< .01

Figure 3: Visual Representation of Natural Log of Copper Futures

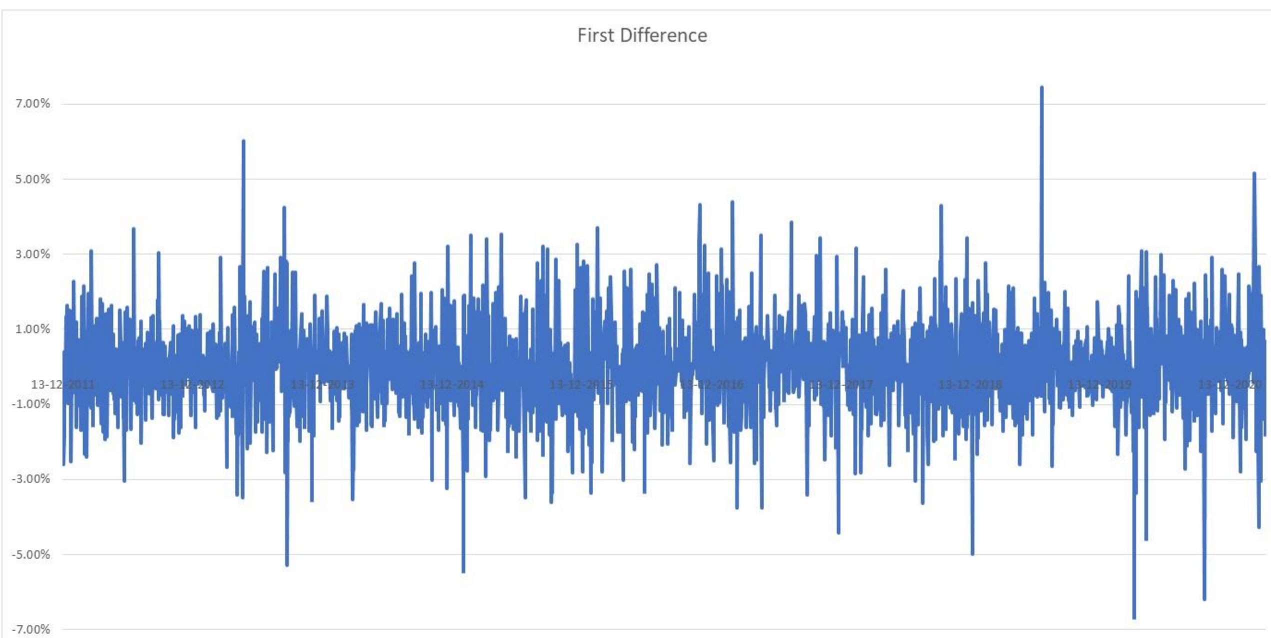


Figure 4: Visual Representation of First Difference of Series in Figure 3

ACF and PACF on First Difference

A correlogram is plotted to identify the appropriate lags till which a significant autocorrelation is observed in the stationary first difference. The appropriate lag in both the plots will help in determination of the appropriate ARIMA Model. In case of both the ACF and PACF, a significant negative correlation is found at lag 16. This implies that 16 is the appropriate value of both ‘p’ and ‘q’, which form 2 of the parameters of the ARIMA(p, d, q) model.

When we extend the x-axis on the correlogram to the maximum possible limit of 60 imposed by Real Stats, we observe significant auto correlation occurring at lags > 30. This significant auto correlation at lags above 30 may be because of similarity in price pattern occurring on account of similar time to maturity of the futures contract. For example, the price movement on 15th Jan (for the contract expiring in January) and the movement on 15th Feb (for the contract expiring in February) may be similar because both the contracts have roughly two weeks to expiry.

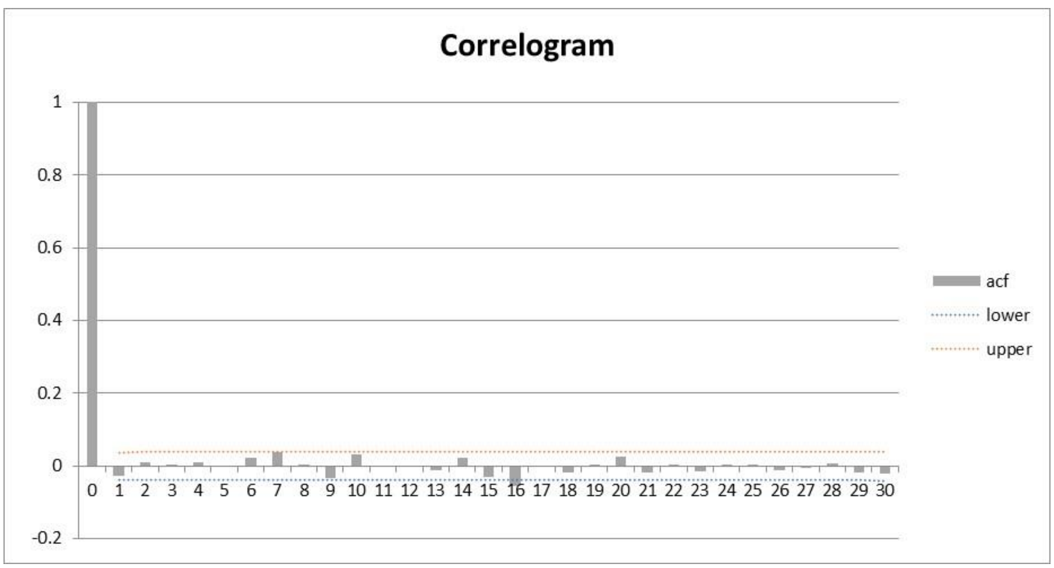


Figure 5: ACF of First Difference

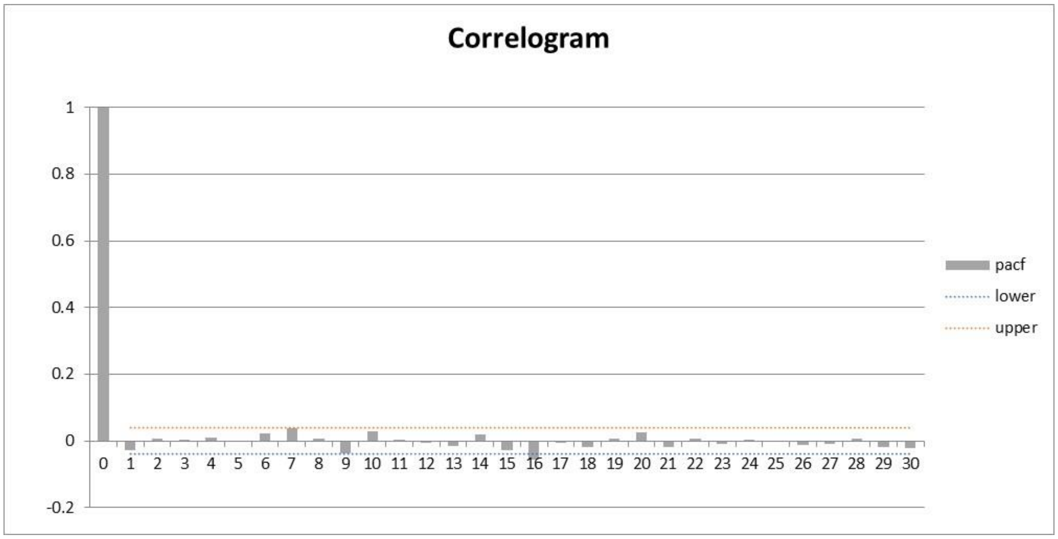


Figure 6: PACF of First Difference

Research Objective

The objective of this research is to check if the relative changes in copper future prices are dependent on the lagged (up to a 16 period lag) relative changes in copper future prices and the resulting lagged (up to a 16 period lag) errors of overall model.

ARIMA

ARIMA models are a combination of Autoregressive (AR) models and Moving Average (MA) models. AR models are fit by regressing a time series variable with its lagged self. MA models are fit by regressing a time series variable with the lagged errors of that model. The lag parameter for the AR model is denoted by ‘p’ and that of the MA model is denoted by ‘q’. The ‘I’ in ARIMA stands for ‘Integrated’ which refers to the degree of differentiation of the target variable to make it stationary. After conducting the above analysis, we arrive at the appropriate order of differencing (1), the appropriate lag value for the AR component of ARIMA (16) and the appropriate lag term for the MA component of ARIMA (16). We therefore use the ARIMA(16, 1, 16) model. The comprehensive model equation is as follows.

$$Y_t = c + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_{16} Y_{t-16} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_{16} \varepsilon_{t-16} + \varepsilon_t$$

Where Y_t represents the first difference of natural logarithm of Copper Future Prices. The fit of the above model is described below.

Table 1: Estimates Table

Parameter	***Index***	***Coefficient***	***t-stat***	***t-crit***	***p-value***	***Significant***
c		0.0014	0.8311	1.9609	0.4060	No
phi	1	-0.3730	-4.7737	-1.9609	0.0000	Yes
phi	2	-0.3332	-4.2101	-1.9609	0.0000	Yes
phi	3	-0.0702	-1.2165	-1.9609	0.2239	No
phi	4	-0.4350	-7.6835	-1.9609	0.0000	Yes
phi	5	-0.4592	-6.0900	-1.9609	0.0000	Yes
phi	6	-0.7873	-11.8078	-1.9609	0.0000	Yes
phi	7	-0.1573	-3.1096	-1.9609	0.0019	Yes
phi	8	-0.4944	-9.5901	-1.9609	0.0000	Yes
phi	9	-0.1531	-2.8649	-1.9609	0.0042	Yes
phi	10	-0.7463	-15.2568	-1.9609	0.0000	Yes
phi	11	-0.4894	-8.5887	-1.9609	0.0000	Yes
phi	12	-0.3189	-5.0092	-1.9609	0.0000	Yes
phi	13	-0.0096	-0.1877	-1.9609	0.8511	No
phi	14	-0.3695	-7.8055	-1.9609	0.0000	Yes
phi	15	-0.2466	-3.8481	-1.9609	0.0001	Yes
phi	16	-0.7491	-12.1745	-1.9609	0.0000	Yes
theta	1	0.3439	4.1811	1.9609	0.0000	Yes
theta	2	0.3248	3.9255	1.9609	0.0001	Yes
theta	3	0.0692	1.1867	1.9609	0.2354	No
theta	4	0.4358	7.5683	1.9609	0.0000	Yes
theta	5	0.4596	5.9568	1.9609	0.0000	Yes
theta	6	0.8058	12.1899	1.9609	0.0000	Yes
theta	7	0.1893	3.8081	1.9609	0.0001	Yes
theta	8	0.5291	10.0066	1.9609	0.0000	Yes
theta	9	0.1229	2.2659	1.9609	0.0235	Yes
theta	10	0.7780	15.9530	1.9609	0.0000	Yes
theta	11	0.5183	8.6842	1.9609	0.0000	Yes
theta	12	0.3445	4.9246	1.9609	0.0000	Yes
theta	13	0.0243	0.4445	1.9609	0.6568	No
theta	14	0.4090	8.1957	1.9609	0.0000	Yes
theta	15	0.2377	3.3308	1.9609	0.0009	Yes
theta	16	0.7646	10.9819	1.9609	0.0000	Yes

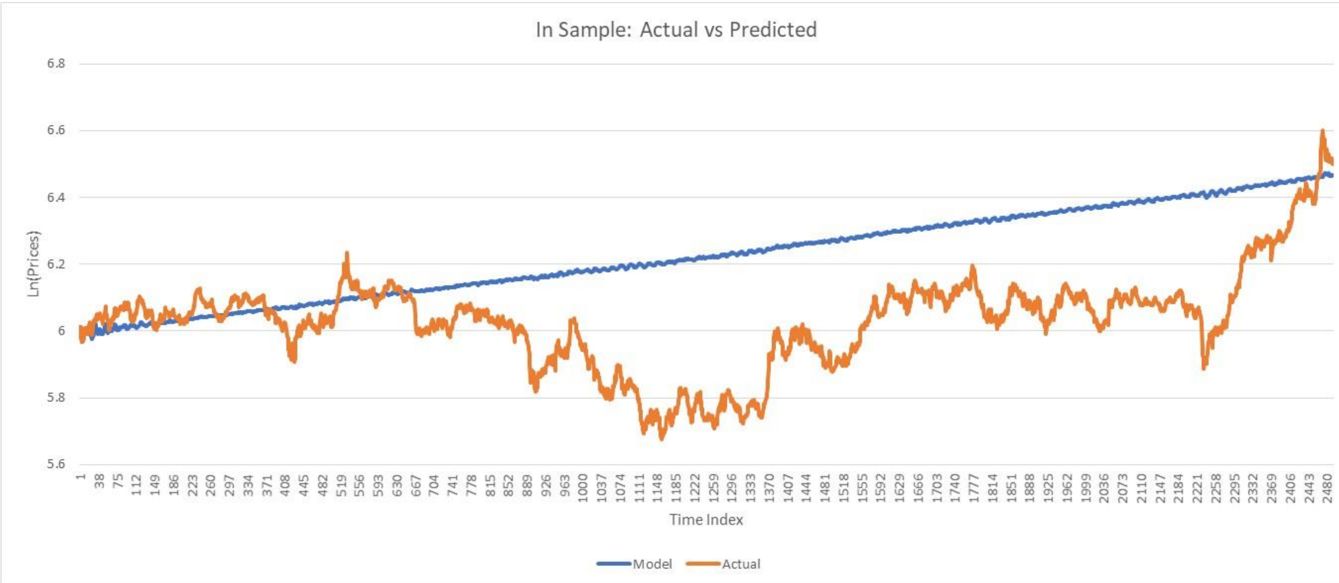


Figure 7: In Sample Comparison: Actual vs Model

ACF & PACF on Residuals

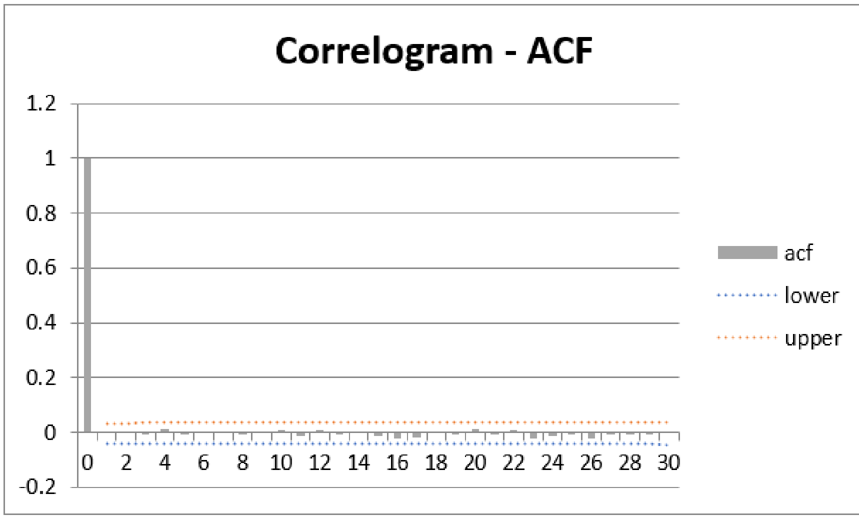
An ACF and PACF plot on residuals bears promising results. Both the plots are insignificant for all lags greater than 0. Moreover, the residuals are found to be stationary in nature.

ADF Test

criteria	schwert
drift	no
trend	no
lag	27
alpha	0.05

tau-stat	-49.8386
tau-crit	-1.94111
stationary	yes
aic	-6.1087
bic	-6.10636
lags	0
coeff	-0.99925
p-value	< .01

Figure 8: Stationarity of Residuals



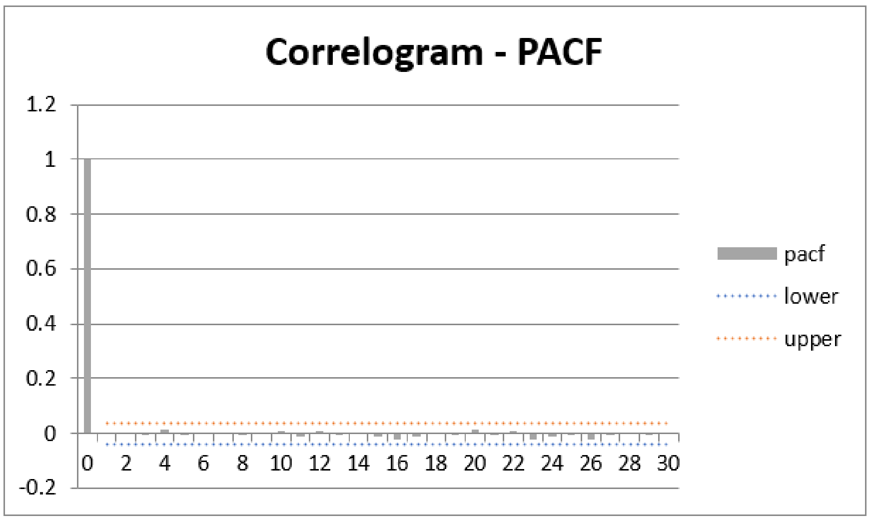


Figure 9: ACF and PACF plot for Residuals

Conclusion and Interpretation

Referring to 'Table 1', we infer that the coefficients are significant all the way till lag 16 for both the AR and MA components of the model. In both the components, lags 3 and 13 are not significant. Furthermore, the constant is not significant.

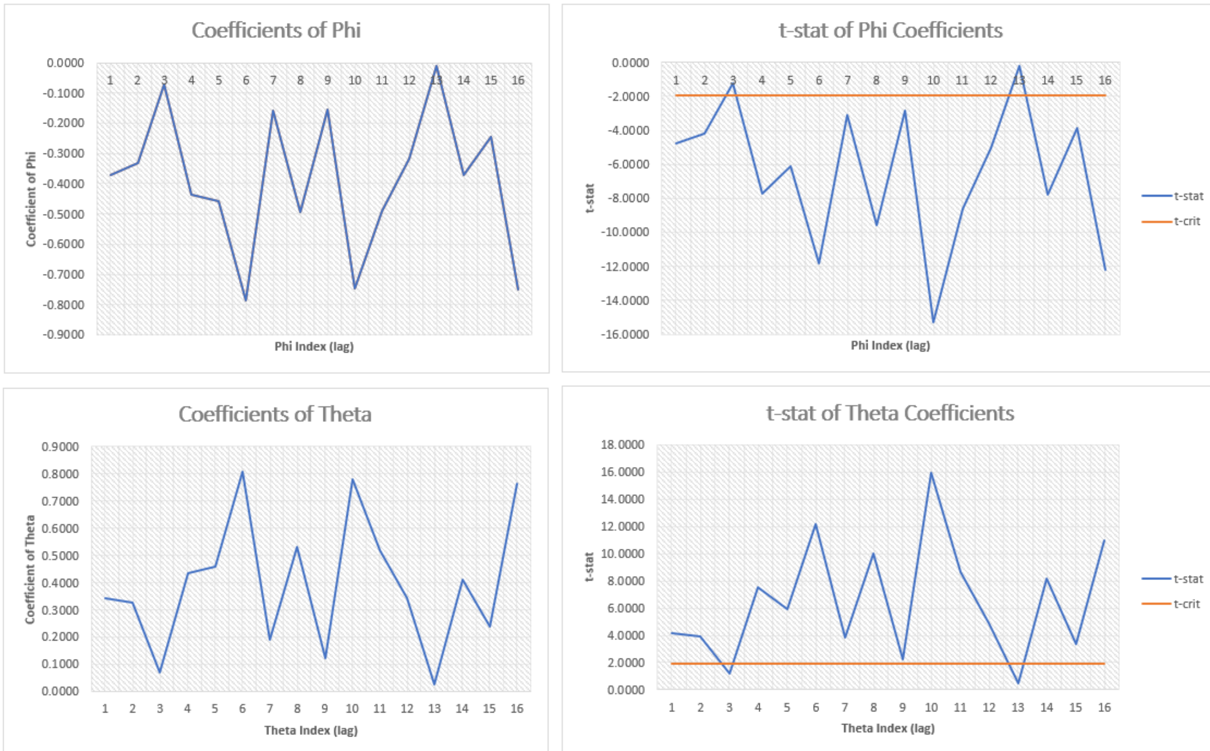


Figure 10: Analysis of Coefficients and Statistics

The charts above shed light on the following aspects:

1. The absolute value of coefficients of Phi is less than 1, thus indicating that Y_t is stationary in nature.
2. Coefficients of Phi are negative. This implies a negative relation between the present (relative) change in copper futures and the lagged (relative) change in copper futures. This implies that each lagged value of change in copper prices leads to an opposite change in the present copper future price to the extent of its weight denoted by Phi.
3. Coefficients of Theta are all positive. This implies that there is a positive relation between previous period errors and current change in copper prices.
4. Across both the MA and AR components: The 6^{th}, 10^{th} and 16^{th} lags have the highest absolute coefficient and are all significant. Among them the 10^{th} lag emerges as the most significant as indicated by the peak at 10 in the t-stat of Phi and Theta coefficients.
5. Since the 10^{th} lag emerges the most significant in both the AR and MA terms, there seems to be some 10 day cycle in the data generating process.

6. The results of this model motivates further research on Copper Futures using ARIMA with larger number of lags. Moreover, this report uses non seasonal ARIMA. Commodities and metal tend to exhibit seasonality. A seasonal ARIMA is worth exploring this area.