Forecasting Stock Prices through Univariate *ARIMA* Modeling

Mustafa Afeef[1] Anjum Ihsan[2] Hassan Zada[3]

Abstract

Stock price forecasting is, and will always be, one of the most imperative financial conjectures investors are confronted with. There are plentiful ways of effectively forecasting a company's share price, most of which rely on various factors that have a bearing on the market price of shares. This paper, however, has employed a method of forecasting which is based on the previous values of the variable itself. This method, technically known as the ARIMA methodology, was developed by Box and Jenkins in 1970. The current paper employed this method on stock prices of one of the largest companies in Pakistan, i.e. Oil & Gas Development Company Limited (OGDCL). Daily adjusted closing stock prices of the company were taken from 2004 to 2018 covering almost 15 years with 3632 observations. Results showed that some of the ARIMA archetypes used in the study had a strong potential for prediction in the short run. It was, therefore, deduced that ARIMA modeling works pretty efficiently for short-term prediction. Investors in stocks may use the findings of the study to supplement their forecasting aptitude.

Keywords: ARIMA, Box-Jenkins Method, Stock Prices, Stationarity, Prediction

Introduction

Stock price prediction is one of the most talked-about phenomena in contemporary financial literature. Individual and institutional investors put all their effort to better envisage the future probable price of a given company's common stock. The million dollar question is; how to anticipate as closely as possible the future market price of a given stock. Most of the researchers have traditionally attempted to forecast stock prices through factors that have an effect, positive or negative, on given firms' value and/or profitability. In other words, the explained variable (prices) is to be predicted by regressing it on multiple explanatory variables. In this paper, an endeavor has been made to speculate our variable of interest by means of the lagged values of the variable itself, based on a popular notion of letting the data speak for themselves (Gould, 1981). Therefore, Autoregressive Integrated Moving Average, or commonly known as *ARIMA*, modeling has been employed to allow the previous values of the dependent variable and the error term to guess the most probable value of our variable of interest.

The prime rationale behind this work is to check whether or not the model employed in this study, i.e., the *ARIMA* works reasonably well in predicting future stock prices. Further, it is also intended to know if the model is more efficient in short-term forecasting or has it got more capability to anticipate stock prices in the longer run. Hence, the objectives of the study are two-fold --- to check for the applicability of *ARIMA* model in predicting the values of a variable, and to investigate which type of

[1] Islamia College, Peshawar E-mail: mustafaafeef@gmail.com
[2] Islamia College, Peshawar
[3] SZABIST, Islamabad

prediction, short-term or long-term, is best provided by the model. It is expected that the work undertaken will significantly help investors decide when to invest in a given company's stock. In other words, implications of the study are that it is expected to be useful for potential stock investors by helping them determine the correct time to invest or disinvest in a given stock.

There have been many studies in the developed part of the world that have used *ARIMA* technique for forecasting various time series variables, some of which have used the model for estimating stock prices as well. However, fewer studies have been conducted in Pakistan for anticipating stock prices engaging *ARIMA*. More specifically, no study, to the best of our acquaintance, has been conducted in the country using the *daily* stock prices data of any non-financial KSE 100 Index company. The current work seeks to fill this gap by taking daily stock prices data of an oil exploratory company in Pakistan popularly named as the OGDCL.

The Oil and Gas Development Company Limited (OGDCL) is the largest company of Pakistan in the E&P (Exploration & Production) sector of Pakistan Stock Exchange. The company was established by the Government of Pakistan in 1961 and later went public on October 23, 1997. The company is undoubtedly the largest of its kind in the country in terms of production, market capitalization, acreage, oil exploration, reserves and profitability. The company's shares are primarily listed in Pakistan Stock Exchange and have a secondary listing on London Stock Exchange since 2006 making it a Pakistani multinational. The government of Pakistan still owns 74% of shares in the company.

Review of Literature

A plentiful amount of research has been undertaken in numerous disciplines or subjects that involve *ARIMA* methodology for the purpose of forecasting the future value(s) of a given variable. To discuss a few, Gay (2016) made an effort to investigate the relationship of macroeconomic variables on stock returns of BRIC countries that include Brazil, Russia, India and China. He made use of the Box-Jenkins method to serve the purpose. The factors taken into account were the exchange rates and the oil prices. No statistically significant association was found to be there between the given macroeconomic factors and stock returns for any of the BRIC economies. Moreover, no significant link was identified of stock return with its lagged values for any of the four countries. Similarly, Manoj and Madhu (2014) used the Box-Jenkins approach to predict the production of sugarcane in India. They found that the model was able to predict future production of sugarcane for almost five years. The most suitable *ARIMA* configuration for sugarcane was found to be *ARIMA* (*2*, *1*, *0*). Hamjah (2014) also employed *ARIMA* for anticipating rice production in Bangladesh. He made a comparison between the actual data of rice production and the predicted series and found that the model had a very good forecasting ability in the short run.

Guha (2016) anticipated gold prices in India using *ARIMA* model in order to give insinuations to the investors about when, and when not, to buy gold. Jadhav, Reddy and Gaddi (2017) used *ARIMA* for predicting the prices of farms and then further used the same for major crops in the Karnataka state of India, namely the Paddi, Ragi and Maize.

They took the data from 2002 to 2016 and found that the model had a very strong power to estimate values for the future. On the basis of this, they also forecasted the 2020 prices of the crops.

Mondal, Shit and Goswami (2014) employed a sector-specific analysis of Indian stocks using *ARIMA* model. They conducted a study on the capability of the model using fifty six Indian stocks from various sectors. They found that the model correctly predicted stock prices to the extent of 85% for all sectors. Banerjee (2014) also used *ARIMA* to predict future Indian stock market index and found the model very accurate in short-term forecasting.

Some Pakistani researchers also made use of *ARIMA* technique to forecast different time series variables. For instance, Zakria and Muhammad (2009) used the model predict the future population of the country of Pakistan. They found out that if the country's population continues to grow at the same rate, there are expected to be 230.68 million people in the country by 2027. The different statistical bureaus of Pakistan, on the other hand, have estimated the country's population to reach 229 million by 2025. Hence, they concluded that the model worked well for predicting their variable of interest. In another study, Farooqi (2014) used ARIMA for predicting the imports and exports of Pakistan. They took the data from 1947 to 2013 and found that *ARIMA* (*2*, *2*, *2*) and *ARIMA* (*1*, *2*, *2*) worked better for predicting both imports and exports. Similarly, Saeed, Saeed, Zakria and Bajwa (2000) attempted to anticipate the production of wheat in Pakistan using the ARIMA model. They found in the diagnostic checking stage of their study that *ARIMA* (*2*, *2*, *1*) was most appropriate for the estimation of wheat production. They believed that the findings of the study would prove helpful for the concerned persons to foretell in advance the requirements of imports and exports of grain storage. In the same manner, Khan, Khan, Shaikh, Lodhi and Jilani (2015) also employed ARIMA model for predicting rice production in Pakistan. The data related to rice production was taken from 1993 to 2015. The diagnostic checking showed *ARIMA* (*2*, *1*, *1*) to be the most suitable *ARIMA* configuration for estimating rice production in Pakistan.

The previous work, therefore, shows that *ARIMA* has a good capacity of estimating various time series data including stock prices. Whether or not the model could successfully foretell the future values of our variable of interest (i.e., OGDCL's stock prices) will be analyzed in the later sections of the study. We now move on to discuss what *ARIMA* technique is and how it works.

The *ARIMA* Model

ARIMA model was introduced by statisticians George Box and Gwilym Jenkins in their book *'Time Series Analysis: Forecasting and Control'* (Box & Jenkins, 1970). This method is suitable for time series of medium to longer length. According to Chatfield (1996), there should be at least 50 observations for *ARIMA* model to work. Many of the other researchers argue that the number of observations should be larger than 100 for the model to give better results. The model predicts future values of a time series on the basis of its past values and on the basis of the past values of error term. The foremost difficulty in *ARIMA* modeling is to *identify* how many lagged values of a variable as well as the error term effectively forecast the current, and future, value of the variable. The developers of the model, Box and Jenkins, have emphasized on going along with the principle of *parsimony*, i.e., keeping the model as simple and condensed as possible. A lengthy model which includes a larger number of regressors would, of course, better forecast a given time series (since R^2 will increase) but at the cost of decreasing degrees of freedom. The two scientists proposed a three-stage model for predicting a given time series. Therefore, the model is also popularly known as the Box-Jenkins methodology; although the econometric term for this type of model prediction is called the *ARIMA* modeling. The four stages of the Box-Jenkins model are (a) identification of the model, (b) model estimation, (c) diagnostic checking, and (d) forecasting. In the first stage, i.e., identification, the researcher visually inspects plots of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) simultaneously to check for patterns such as spikes, exponential decay or damped sinewave etc. Through this process, the most suitable *ARIMA* configuration, including the number of autoregressive processes (i.e., the *AR*) influencing a given time series variable, the number of moving average processes (i.e., the *MA*) and the number of times the series should be differenced in order to render it stationary (i.e., the d), for the time series is identified (Asteriou, 2007).

In the second stage of the Box-Jenkins methodology, econometric estimations are made. All the tentative models are estimated so that their respective coefficients including the R^2 values, the Akaike Information Criterion (*AIC*) and the Schwarz Bayesian Criterion (*SBC*) coefficients etc. are compared. The third stage of the process involves diagnostic checking in which a *comparison* is made among the models on the basis of the just-mentioned criteria and the one which is the most parsimonious is selected. In the final stage of the methodology, i.e., the *forecasting*, the *next* or the future value of the time series is mathematically computed to see how close the forecasted value is with the actual value. This also gives calculation of the error term.

Research Methodology

The study deals with analysis of a univariate time series. When dealing with time series econometric framework, it is often better to extract information about a variable that can be gathered from the variable itself (Asteriou & Hall, 2007). Therefore, as mentioned before, the autoregressive integrated moving average (*ARIMA*) model, also popularly known as the Box-Jenkins methodology, which has been discussed in the previous section, has been employed in the study. The general configuration for an *ARMA* process as taken up from Asteriou and Hall (2007) is:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \text{- - -} + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \text{- - -} + \theta_q \varepsilon_{t-q}$$

Where,

Y_t is the predicted value of the variable, Y_{t-1}, Y_{t-2}, - - - , Y_{t-p} are the lagged values of the dependent variable or the autoregressive terms, ε_t is the error term, ε_{t-1}, ε_{t-2}, - - - , ε_{t-q} are the lagged values of the error or the moving average terms, and φ and θ are the coefficients or slopes of the regressors.

But this process assumes that the time series variable under consideration is weakly stationary. The term stationarity here indicates that the mean and the variance of the series are roughly constant overtime and that the covariance of the series is one that is time-invariant (Gujarati & Porter, 2004). However, our familiarity with time series data depicts that most, if not all, of the time series are integrated and therefore clearly non-stationary. Using *ARMA* process over a non-stationary data will, of course, give no results. Hence, the more appropriate and customized *ARIMA* procedure was employed so that the series, if integrated, is *differenced enough* to render it sufficiently stationary.

The data used for the study were OGDCL's daily stock returns for around fifteen years computed through the closing prices of the company. To reiterate, the current study uses a *univariate* time series analysis. This is the reason why only a single company was taken up for the study. Of course the analysis can be extended to any other single company as well. However, taking multiple companies at the same time will require *multivariate* time series analysis, a procedure not addressed in the current study.

In order to discover the most fitting *ARIMA* configuration for the stock prices of the company studied, the following criteria, as prescribed by Box and Jenkins (1970), was followed for model selection:

- The model with the least insignificant parameters
- The model with the highest adjusted R^2
- The model with the lowest Akaike Information Criterion and Schwarz Bayesian Criterion values
- If all else the same, then the model which is the most parsimonious

Results and Findings

The stock price data for OGDCL contains daily closing prices from Jan 23, 2004 to Nov 19, 2018 covering nearly 15 years. This translated into 3682 observations. Before using the *ARIMA* technique, it is necessary to ensure that the variable obeys the convention of Stationarity. A variable is said to be stationary if it has a time-invariant mean, variance and covariance (Gujarati & Porter, 2004). The Stationarity of stock prices of the company under study was checked through line graph for visible inspection and was further validated through the Augmented Dickey Fuller test.

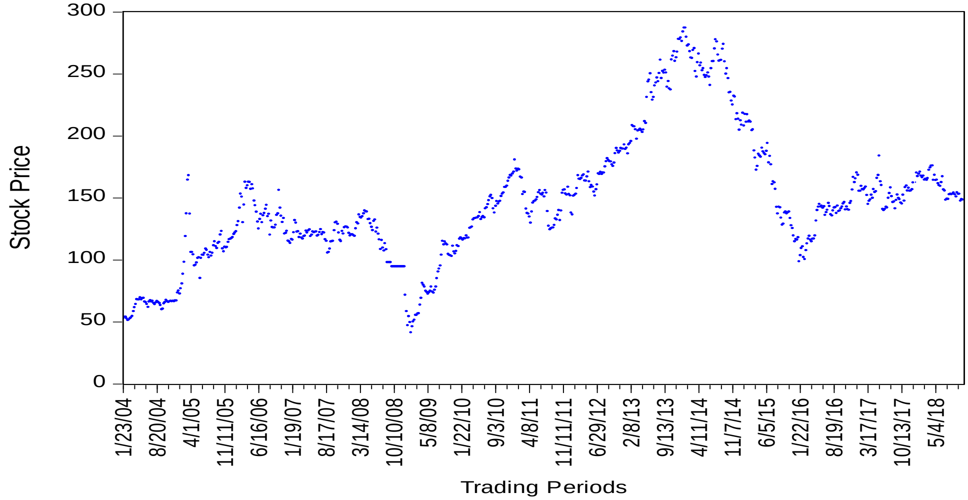

Figure 1: *Non-Stationary Share Prices of OGDCL*

Figure 1 shows the non-stationary behavior of share prices as expected. The result has been attested by the ADF test statistic which is insignificant at 5% level.

Table 1: *Augmented Dickey Fuller Test for OGDCL Stock Price*

Null Hypothesis: STOCKPRICE has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, max lag = 19)

		t-Statistic	Prob.
Augmented Dickey-Fuller test statistic		-2.224	.198
Test critical values:	1% level	-3.439	
	5% level	-2.865	
	10% level	-2.569	

Table 2 gives the correlogram of the time series under study. As per the theory, the correlogram of a stationary process should fade away as the lag length increases. However, as is visible in the given figure, the autocorrelation function (ACF) of OGDCL's stock prices does not vanish at all portraying the non-stationary nature of the series.

Table 2: *Autocorrelation and Partial Autocorrelation Function of OGDCL Stock*

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|*******	.|*******	1	.991	.991	734.62	.000
.|*******	*| |	2	.981	-.079	1455.0	.000
.|*******	.| |	3	.971	.017	2161.6	.000
.|*******	.| |	4	.961	.003	2855.0	.000
.|*******	.| |	5	.951	-.015	3534.9	.000
.|*******	.| |	6	.941	.021	4202.1	.000
.|*******	.| |	7	.933	.063	4858.6	.000
.|*******	.| |	8	.924	-.028	5503.9	.000
.|*******	.| |	9	.917	.033	6139.1	.000
.|*******	.| |	10	.909	-.003	6764.2	.000
.|******|	.| |	11	.901	.016	7379.9	.000
.|******|	.| |	12	.894	.008	7986.5	.000
.|******|	.| |	13	.887	.054	8585.3	.000
.|******|	.| |	14	.881	-.026	9176.1	.000
.|******|	.| |	15	.874	-.011	9758.5	.000
.|******|	.| |	16	.867	.013	10333.	.000
.|******|	.| |	17	.860	-.065	10898.	.000

In order to induce stationarity, therefore, natural logarithms have been employed for stock prices and then the change in the log of stock prices, i.e., the first difference, has been taken into account. The result is a stationary series as evident from figure 2.

The evidence of the stationarity of the series is also blatant from the ADF test that has been employed again on the logged differenced series of stock prices. This time the test statistic is highly significant at 1% level (see table 3).

Table 3: *Augmented Dickey-Fuller Test for Logged Differenced Stock Prices*

Null Hypothesis: DLSTOCKPRICE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=19)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-23.174	.000
Test critical values:	1% level	-3.439	
	5% level	-2.865	
	10% level	-2.569	

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

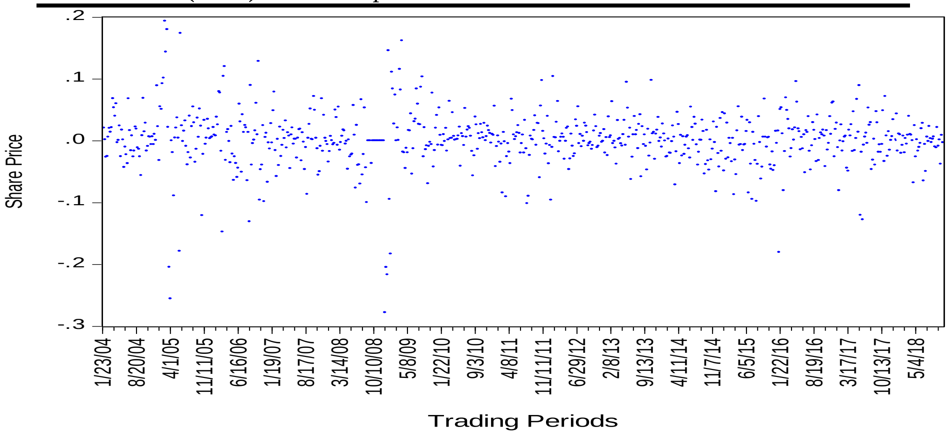

Figure 2: Stationarity-Induced Share Prices of OGDCL

Model Identification

After stationarity in the given time series has been achieved through logged differencing, we move on to apply the Box-Jenkins methodology. The first step is the identification of the appropriate model. Hence, a Correlogram is again made to explore the number of *AR* and *MA* terms that the stock price of OGDCL *depends* on.

Table 4: *Autocorrelation and Partial Autocorrelation Function of Logged Differenced OGDCL Stock Prices*

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|* |	.|* |	1	.160	.160	19.057	.000
.| |	.| |	2	.014	-.012	19.203	.000
.| |	.| |	3	.025	.026	19.684	.000
.| |	.| |	4	.041	.034	20.937	.000
.| |	.| |	5	-.037	-.050	21.969	.001
*| |	*| |	6	-.098	-.087	29.260	.000
.| |	.| |	7	-.010	.018	29.343	.000
.| |	.| |	8	-.044	-.046	30.792	.000
.| |	.| |	9	-.035	-.014	31.702	.000
.| |	.| |	10	-.027	-.013	32.245	.000
.| |	.| |	11	-.044	-.045	33.689	.000

.| |	.| |	12	-.055	-.047	35.965	.000
.| |	.| |	13	-.022	-.006	36.349	.001
.| |	.| |	14	.016	.014	36.539	.001
.| |	.| |	15	-.015	-.021	36.712	.001

The general methodology of *ARIMA* modeling as prescribed by Box and Jenkins involves:

a. Selecting an upper limit for p, say $p_{\max}$, and q, say $q_{\max}$;

b. Estimating all models with $0 \leq p \leq p_{\max}$ and $0 \leq q \leq q_{\max}$; and

c. Using information criteria like *AIC, SBC* and R^2 to make a distinction among the contending models.

Studying table 4, it is observed that there is only one positive spike in the autocorrelation function and one in the partial autocorrelation function and then both die down immediately. This signals, evidently, towards an *ARIMA* (*1*, *d*, *1*) model. However, researchers normally attempt to fit an *ARIMA* (*1*, *d*, *0*) model and an *ARIMA* (*2*, *d*, *1*) as a starting point. Using the principal of parsimony, nonetheless, the model which is the most *thrifty* and *tightfisted* will be preferred.

Model Estimation

In the current segment, a few of the most *probable* models of *ARIMA* including the one prescribed by the Box-Jenkins method are estimated and their results are compared in order to determine the one which fits the most and yet is the most parsimonious.

Table 5: *Regression Results using ARIMA (1, d, 1) Model*

Dependent Variable: Logged Differenced Stock Price
Method: Least Squares
Included observations: 3682 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.000	.000	.679	.497
AR(1)	.652	.079	8.271	.000
MA(1)	-.551	.087	-6.345	.000

R-squared	.018	Mean dependent var	.000
Adjusted R-squared	.017	S.D. dependent var	.018
S.E. of regression	.018	Akaike info criterion	-5.184
Sum squared resid	1.206	Schwarz criterion	-5.179
Log likelihood	9547.090	Hannan-Quinn criter.	-5.183
F-statistic	33.009	Durbin-Watson stat	1.982
Prob (F-statistic)	.000		

Table 5 presents the results of *ARIMA* (*1*, *d*, *1*) model, the one which has been selected using the typical Box-Jenkins method. Both the parameters of the model are significant at 5% level. The model has an adjusted R^2 value of .018. However, in order to ensure this model is the most fitting one, a few other probable models are also run so that a comparative analysis amongst them is made.

Table 6: *Regression Results using ARIMA (2, d, 1) Model*

Dependent Variable: Logged Differenced Stock Price
Method: Least Squares
Included observations: 3681 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.000	.000	.641	.521
AR(1)	.862	.108	7.967	.000
AR(2)	-.045	.026	-1.699	.089
MA(1)	-.753	.106	-7.074	.000

R-squared	.018	Mean dependent var	.000
Adjusted R-squared	.017	S.D. dependent var	.018
S.E. of regression	.018	Akaike info criterion	-5.185
Sum squared resid	1.204	Schwarz criterion	-5.178
Log likelihood	9546.159	Hannan-Quinn criter.	-5.182
F-statistic	22.464	Durbin-Watson stat	2.000
Prob (F-statistic)	.000		

Table 6 represents the results of *ARIMA* (*2*, *d*, *1*) model, the one which is often used in many *ARIMA* analyses by researchers. This model seems to be a bit inferior to *ARIMA* (*1*, *d*, *1*) model, or the one using Box-Jenkins approach, as one of the three parameters of the model is insignificant at 5% level. The adjusted R^2 value of the two models is, however, equal. As far as the information criteria are concerned, *ARIMA* (*2*, *d*, *1*) has a lesser Akaike information criterion (AIC) value and higher Schwarz Bayesian criterion (SBC), and the Hannan-Quinn criterion (HQC) values than *ARIMA* (*1*, *d*, *1*). Since the lower the information criterion values, the better the model is, therefore *ARIMA* (*1*, *d*, *1*) can be considered better than *ARIMA* (*2*, *d*, *1*).

Table 7: *Regression Results using ARIMA (1, d, 0) Model*

Dependent Variable: Logged Differenced Stock Price
Method: Least Squares
Included observations: 3682 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.000	.000	.798	.424
AR(1)	.117	.016	7.191	.000

R-squared	.013	Mean dependent var	.000
Adjusted R-squared	.013	S.D. dependent var	.018
S.E. of regression	.018	Akaike info criterion	-5.181
Sum squared resid	1.211	Schwarz criterion	-5.177
Log likelihood	9540.037	Hannan-Quinn criter.	-5.180
F-statistic	51.712	Durbin-Watson stat	2.011
Prob(F-statistic)	.000		

In an attempt to further trim down or abridge our prescribed model, the simplest autoregressive process of order 1, i.e., the *ARIMA* (*1*, *d*, *0*) process or the *AR* (*1*) process, was also examined. It was, however, found that the adjusted R^2 value further decreased, although by a very small fraction. All the information criterion values including *AIC*, *SBC* and *HQC* also were slightly higher than those for *ARIMA* (*1*, *d*, *1*). This indicated that *ARIMA* (*1*, *d*, *1*) was probably the simplest possible configuration and that any further simplification introduced in the model would come at the cost of a reduced forecasting capacity. The results of *ARIMA* (*1*, *d*, *0*) model are summarized in table 7.

Table 8: *Regression Results using ARIMA (2, d, 2) Model*

Dependent Variable: Logged Differenced Stock Price
Method: Least Squares
Included observations: 3681 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.000	.000	.683	.494
AR(1)	-.319	.102	-3.118	.002
AR(2)	.580	.089	6.468	.000
MA(1)	.422	.108	3.886	.000
MA(2)	-.477	.094	-5.024	.000
R-squared	.018	Mean dependent var		.000
Adjusted R-squared	.017	S.D. dependent var		.018
S.E. of regression	.018	Akaike info criterion		-5.184
Sum squared resid	1.205	Schwarz criterion		-5.176
Log likelihood	9546.447	Hannan-Quinn criter.		-5.181
F-statistic	16.989	Durbin-Watson stat		1.985
Prob(F-statistic)	.000			

In search of getting the most appropriate model, *ARIMA* (*2*, *d*, *2*) was also checked for its prediction capacity. As can be seen in table 8, all the parameters of the model were highly significant. The adjusted R^2 of the model was also equal to the Box-Jenkins' specified *ARIMA* (*1*, *d*, *1*). However, all the information criterion values of the model were slightly higher than those for *ARIMA* (*1*, *d*, *1*) which, again, made the model less practicable than *ARIMA* (*1*, *d*, *1*).

In the final endeavor, an over-parameterized model was run to check whether it came up with an even better predictive ability or not. Hence, an *ARIMA* (*3*, *d*, *3*) model was also estimated. As can be examined in table 9, it too worked out to be less powerful than *ARIMA* (*1*, *d*, *1*).

Table 9: *Regression Results using ARIMA (3, d, 3) Model*

Dependent Variable: Logged Differenced Stock Price
Method: Least Squares
Included observations: 3680 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.000	.000	.643	.520
AR(1)	.064	.517	.125	.901
AR(2)	.784	.153	5.114	.000
AR(3)	-.132	.347	-.380	.704
MA(1)	.045	.519	.088	.930
MA(2)	-.734	.196	-3.750	.000
MA(3)	.072	.303	.237	.812

R-squared	.019	Mean dependent var	.000
Adjusted R-squared	.017	S.D. dependent var	.018
S.E. of regression	.018	Akaike info criterion	-5.183
Sum squared resid	1.204	Schwarz criterion	-5.172
Log likelihood	9544.920	Hannan-Quinn criter.	-5.179
F-statistic	11.848	Durbin-Watson stat	1.999
Prob (F-statistic)	.000		

The aforementioned table represents *ARIMA* (*3*, *d*, *3*) configuration, the model with the maximum number of parameters run so far. The model has the same adjusted R^2 value as that of *ARIMA* (*1*, *d*, *1*) but more information criterion values. The weakest part of this model is the presence of four insignificant parameters, i.e., those representing *AR*(*1*), *AR*(3), *MA*(*1*), and *MA*(3) lags.

Diagnostic Checking

Following is a summary table representing the adjusted R^2 values, *AIC*, *SBC*, *HQC*, and the number of insignificant lags or parameters of different combinations of *ARIMA* used in this study for a quick comparison.

Table 10: *Comparison of probable ARIMA models with the bold row representing the most appropriate model*

ARIMA Model	Adjusted R^2	AIC	SBC	HQC	Insignificant lags
ARIMA (*1*, *d*, *0*)	.013	-5.181	-5.178	-5.180	None
ARIMA (*1*, *d*, *1*)	.017	-5.184	-5.179	-5.182	None
ARIMA (*2*, *d*, *1*)	.017	-5.184	-5.178	-5.182	One
ARIMA (*1*, *d*, *2*)	.017	-5.184	-5.177	-5.182	One
ARIMA (*1*, *d*, *3*)	.017	-5.184	-5.175	-5.181	Two
ARIMA (*2*, *d*, *2*)	.017	-5.184	-5.176	-5.181	None
ARIMA (*3*, *d*, *3*)	.017	-5.183	-5.172	-5.179	Four

The summary of *ARIMA* models given in table 10 gives a clearer and an undisputed picture of the best model to be used for predicting OGDCL's stock prices. As can be judged, this model is *ARIMA* (*1*, *d*, *1*) which is marked by bold letters in the table. This model has an adjusted value of R^2 which is at par with many other models. What makes the model better than the rest of the choices is that it has the lowest value of *SBC* --- an information criterion always preferred over the remaining criteria when selecting among competing models.

Forecasting

In the aforementioned section, it was found that the most suitable *ARIMA* configuration for forecasting OGDCL's prices is *ARIMA* (*1*, *d*, *1*). This implies that daily stock prices of OGDCL can be predicted by taking into account one-day previous stock prices and one-day previous error term. Hence, mathematically, the following *ARIMA* model is to be employed:

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

However, since we had taken logged differenced stock prices rather than their absolute values, therefore, assuming that Y^*_t represents stock prices after taking natural logarithms and then first differences, the equation just mentioned can be re-written as:

$$Y^*_t = \mu + \varphi_1 Y^*_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Since, of course, we want to predict stock prices rather than the change in stock prices, one method is to reverse the transformation we made in the form of taking first differences thereby making the series *integrated* again. Thus, in order to get the forecasting estimates of our stock prices rather than the first differences of stock prices, we re-formulate the model as:

$$Y_t - Y_{t-1} = \mu + \varphi_1(Y_{t-1} - Y_{t-2}) + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Re-arranging we get,

$$Y_t = \mu + \varphi_1(Y_{t-1} - Y_{t-2}) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + Y_{t-1}$$

The aforementioned formula represents the actual value of Y_t and therefore includes the error term. Since, we will be estimating the value of Y_t, our expression, i.e., $\hat{Y}_t$ will, thus not include the error coefficient (Gujarati & Porter, 2004) and will be expressed as:

$$\hat{Y}_t = \mu + \varphi_1(Y_{t-1} - Y_{t-2}) + \theta_1 \varepsilon_{t-1} + Y_{t-1}$$

In the context of the current study, we have taken daily stock price values of our variable of interest with effect from January 19, 2004 to December 14, 2018. Therefore, in order to forecast stock prices of OGDCL for December 15, 2018, we will use the following formula:

$$\hat{Y}_{15Dec2018} = \mu + \varphi_1(Y_{14Dec2018} - Y_{13Dec2018}) + \theta_1 \varepsilon_{14Dec2018} + Y_{14Dec2018}$$

Putting the values of the *constant* of the model and the *beta* coefficients, the formula can be re-written as:

$$\hat{Y}_{15Dec2018} = .000262 + (.652)(Y_{14Dec2018} - Y_{13Dec2018}) + (-.551)\varepsilon_{14Dec2018} + Y_{14Dec2018}$$

Finally, in order to compute the estimated value of $\hat{Y}_t$ for December 15, 2018, we incorporate in the equation all the other values including the current and the lagged values of logged stock prices and the lagged value of the error term.

$$\hat{Y}_{15Dec2018} = .000262 + (.652)\ (4.935265 - 4.935409) + (-.551)\ (.001624) + (4.935)$$

$$\hat{Y}_{15Dec2018} = .000262 - .000094 - .000895 + 4.935$$

$$\hat{Y}_{15Dec2018} = 4.934 \text{ (approx.)}$$

The value of 4.934 represents the logged stock price. By taking the anti-log of 4.934, we get 138.98. Hence, our forecasted value of the stock price of OGDCL for the following day, i.e., December 15, 2018 was Rs. 138.98. To reveal, no trading took place on December 15, 2018. However, the actual price of OGDCL's shares on the next trading day, i.e., on December 17, 2018 was Rs. 138.35. Hence, the forecast error was an overestimate of Rs. 0.63.

Analysis and Discussion

The findings of the study given in the above section have shown that stock prices can be plausibly forecasted using the Box-Jenkins methodology. And as has been just mentioned, the most fitting *ARIMA* model for the study at hand is the *ARIMA* (*1*, *d*, *1*) configuration. This result is quite reasonably in line with the findings of the previous studies undertaken with a view to test *ARIMA* model for its accuracy in forecasting a given time series. For instance, Manoj and Madhu (2014) also employed the model to predict the production of Sugarcane in India and found out that the model was very efficient in short-term forecasting. However, the most suitable *ARIMA* layout for them was *ARIMA* (*2*, *1*, *0*) in contrast with *ARIMA* (*1*, *d*, *1*) model that was appropriated for OGDCL's prices in the current study. Similarly, Hamjah (2014) also used ARIMA method to predict rice production in Bangladesh and concluded that the model was very good in short term forecasting.

The study of Mondal, Shit and Goswami (2014) could be considered the most relevant to this work since it was also used for predicting stock prices. They conducted the study using fifty six Indian companies' stocks to check whether the model effectively gave any significant clue regarding the future prices of a given stock. They explored that *ARIMA* was very successful in predicting stock prices with around 85% of its predictions being correct. Adebiyi, Adewumi and Ayo (2014) also used the Box-Jenkins method to predict stock prices and came up with the same conclusion, i.e., an excellent short-term anticipation power of the model.

A very few studies, however, failed to significantly predict their time series using the model under consideration. One of them, for instance, was conducted by Gay (2016) who took the BRIC countries' data to explore the relationship of macroeconomic factors with stock returns but could not establish any. They then endeavored to check the relationship of stock return with its very own lagged values, thereby running the *ARIMA* method, but again ended up with no significant association of stock returns with its previous values.

Nonetheless, the current study is in line with the results of most of the research work undertaken with a view to check whether *ARIMA* is a better predictor of time series data or not. As can be seen, many studies, including this, have figured out that the model works adequately in the short run.

Conclusion

Prediction has always been a challenge for scientists in most of the disciplines. When it comes to financial theory, the process of speculation becomes even more susceptible as there are normally too many aspects or factors that need to be considered for a realistic conjecture. Some theorists, therefore, prefer to forecast the current (and/or future) price of a time series on the basis of its past values as well as the past values of the disturbance term. This concept, popularly known as the Box-Jenkins method and technically as the *ARIMA* method, was also employed in the current study. From the findings of the study, it has been construed that *ARIMA* has a very good capacity to forecast future values in the short run. Of course the long-term prediction using lagged values of a variable will make only little sense, however. It was also observed that almost all *ARIMA* configurations experimented in this study had vigorous prediction powers which testifies the idea of *letting the data speak about themselves*. As a policy implication, the autoregressive integrated moving average model can be potentially used by stock market investors as a clue to anticipate whether the stocks that they have invested in are likely to make an upward move, or vice versa, in the near future.

References

Asteriou, D., & Hall S. (2007). *Applied Econometrics*: A modern approach using eviews and microfit revised edition: Palgrave Macmillan, New York, USA.

Banerjee, D. (2014). *Forecasting of Indian stock market using time-series ARIMA model*. Paper presented at the 2nd International Conference on Business & Information Management (pp. 131-135). Durgapur, India. IEEE.

Box, G., & Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, California, USA, 553p.

Chatfield, C. (1996). *The Analysis of Time Series*, 5th ed., Chapman & Hall, New York.

Farooqi, Ahmad. (2014). ARIMA Model Building and Forecasting on Imports and Exports of Pakistan. *Pakistan Journal of Statistics and Operations Research*, *10*(2), 157-168.

Gay, R. (2016). Effect of Macroeconomic Variables on Stock Market Returns for Four Emerging Economies: Brazil, Russia, India, and China. *International Business & Economics Research Journal*, *15*(3), 119-126.

Gould, P. (1981). Letting the Data Speak for Themselves. *Annals of the Association of American Geographers, 71*(2), 166-176.

Guha, B., & Bandyopadhyay, G. (2016). Gold Price Forecasting using ARIMA Model. *Journal of Advanced Management Science*, *4*(2), 117-121.

Gujarati, D., & Porter, D. (2004). *Basic Econometrics*, 4th Edition, McGraw-Hill, New York, USA.

Hamjah, M. (2014). Rice Production Forecasting in Bangladesh: An Application of Box-Jenkins ARIMA Model. *Mathematical Theory and Modeling*, *4*(4), 1-11.

Jadhav, V., Reddy, B., & Gaddi, G. (2017). Application of ARIMA Model for Forecasting Agricultural Prices. *Journal of Agricultural Science and Technology, 19*(5), 981-992.

Khan, K., Khan, Gulawar., Shaikh, S. A., Lodhi, A. S., & Jilani, G. (2015). ARIMA Modeling for Forecasting of Rice Production: A Case Study of Pakistan. *Lasbela University Journal of Sciences and Technology*, *4*(1), 117-120.

Manoj, K., & Madhu, A. (2014). An Application of Time Series ARIMA Forecasting Model for Predicting Sugarcane Production in India. *Studies in Business and Economics, 9*(1), 81-94.

Mondal, P., Shit, L., & Goswami, S. (2014). Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications*, *4*(2), 13-29.

Saeed, N., Saeed, A., Zakria, M., & Bajwa, T. M. (2000). Forecasting of Wheat Production in Pakistan Using ARIMA Models. *International Journal of Agriculture & Biology*, *2*(4), 352-353.

Zakria, M., & Muhammad, F. (2009). Forecasting the Population of Pakistan using ARIMA Models. *Pakistan Journal of Agricultural Sciences*, *46*(3), 214-223.