Can Stock Returns in Pakistan be Reasonably Forestalled? Employing the Box-Jenkins Approach to Univariate Arima Analysis

Mustafa Afeef*, Fayaz Ali Shah†, Shahid Jan Kakakhel‡

Abstract

Will the stock market go up or down? Only few investors, if any, might be able to bequeath meticulous predictions. However, since most of the subject matter in Finance is future oriented, the capability of investors to forecast the *future* is of utmost importance. When it comes to predicting movements in a stock market, one way is to take account of all those factors, particularly macroeconomic in nature, that have a bearing on the upward and/or downward movement of stock prices as a whole. There is, nevertheless, another way of anticipating changes in a time series --- to foresee its future values by looking at its past values. Therefore, the autoregressive integrated moving average technique was employed for the study that accounted for the lagged values of the variable and its disturbance term. Weekly figures of KSE 100 Index were taken for 22 years from 1997 to 2019 leading to 1143 observations. Results revealed that the model was able to predict the index quite precisely in the short run. The findings of the study might prove to be helpful for investors who wish to invest at Pakistan Stock Exchange in deciding when to increase, or decrease, investment in their portfolios.

***Keywords**:* Box-Jenkins methodology, *ARIMA*, KSE 100 index, stationarity, prediction

Introduction

When it comes to investments, one of the most important concerns of an investor is predicting the future. One needs to critically evaluate all the alternative possibilities of investing a certain sum of money in terms of their probable future outcomes. Since the future is by and large, if not always, uncertain (especially with regard to stock prices), none of the possible alternatives to an investment is likely to warrant a definitive return. That said, many investors use different techniques to predict the future possible price of a given stock. There are broadly two ways of

* Assistant Professor, Department of Management Sciences,
Islamia College Peshawar, Pakistan
Email: mustafa@icp.edu.pk Cell: 0092-333-9534693
† Assistant Professor, Department of Management Sciences,
Islamia College Peshawar, Pakistan
fayaz@icp.edu.pk. Cell: 0092-341-9375701
‡ Associate Professor, Department of Management Sciences,
Islamia College Peshawar, Pakistan
Email: shahidjan@icp.edu.pk Cell: 0092-333-9316678

forecasting a stock's future price. One method is to predict a share's price keeping in view the various major elements that are usually responsible for affecting it, whether internal, i.e., related to the firm, or external, i.e., pertaining to the overall macro-environment in which the company operates. The other method is to let the lagged prices of a share for anticipating its future price. The study at hand implements the latter method of econometric forecasting, commonly referred to as the autoregressive integrated moving average (hereafter termed as the *ARIMA*) technique and popularly recognized as the Box-Jenkins methodology. The technique makes it possible to predict the future value of any given time series by considering its previous values and also the lagged values of its error term. "Let the data speak for themselves" as was once stated by Gould (1981), the *ARIMA* model is designed to work exactly under the same philosophy.

The reason for conducting the study is to explore whether *ARIMA* is equally useful in predicting returns at Pakistan Stock Exchange. If yes, how many previous values of the PSX 100 index can accurately, if at all, anticipate the future/coming value of the index? The objectives of this study, in a manner, are dual; i.e., initially to examine whether the Box-Jenkins method is applicable at all for predicting stock returns at Pakistan Stock Exchange, and secondarily to examine what number of the previous values of the index and of the disturbance term is required to adequately predict the index or the returns. The work is expected to help investors foretell where the stock market is expected to move in the future.

Review of Literature

Researchers seem to have used in the past the *ARIMA* model for determining the future values of their variables of concern. Studies have been conducted to forecast stock prices --- our variable of interest --- as well as other time series of different nature. The ensuing lines present some of the previous work relating to the current study:

The Box-Jenkins method was used by Gay (2016) for exploring the relationship of stock returns with exchange rate and oil prices for BRIC countries. According to him, the method could not effectively do what was expected of it.

Mondal, Shit and Goswami (2014) took 56 Indian firms and classified them based on the sectors they belonged to. They used *ARIMA* model to forecast their stock returns and found that the model was successful for 85% of the firms in its prediction.

An effort of predicting stock returns using the *ARIMA* technique was also marked by Adebiyi, Adewumi and Ayo (2014) who used the model for forecasting share market values of Zenith Bank and that of Nokia. They saw that the model was good enough for short term forecasting. Similarly, Banerjee (2014) also anticipated the Bombay stock market index by employing the Box-Jenkins method and concluded that it accurately performed in the short run forecasting.

ARIMA technique has also been employed by studies for forecasting different time series variables other than stock prices. To discuss a few, for example, Jarrett (1990) used the model for predicting corporate earnings. He found no legitimate difference in results between the traditional models and *ARIMA* based model with respect to forecast errors. Raymond (1997) also employed the model for predicting Hong Kong's real estate prices and was able to detect patterns or tendencies in the prices of real estate properties. Likewise, the model was also used by Contreras, Espinola, Nogales and Conejo (2003) for anticipating the prices of electricity in California and Spain. Gilbert (2005), on the other hand, applied the technique on his multistage supply chain model. It was noticed by him that the orders made by customers, the inventories, customers' demands and the lead times are all processes requiring *ARIMA* or likewise treatments. The Box-Jenkins method was also utilized by Guha (2016) who anticipated the prices of gold in India for determining when to make purchase decision in the Indian gold market.

ARIMA modeling was also used by researchers for determining the production volume of agricultural crops. For instance, Manoj and Madhu (2014) employed the Box-Jenkins method for predicting the Indian Sugarcane production. The model appeared to predict the production of Sugarcane for approximately five years. In a similar study carried by Hamjah (2014), production of rice crop in Bangladesh was estimated. The results showed a good estimation although only in the short run. In India, the possible productivity of agricultural crops was also forecasted by Padhan (2012) who studied 34 heterogeneous Indian crops. The best estimation made in the analysis of her study was for the *tea* crop with the worst being for the *papaya* among those crops. A few more Indian researchers also attempted to predict prices of agricultural crops. Among them were Jadhav, Reddy and Gaddi (2017) who estimated the production of major crops in the Karnataka state of India including *Ragi*, *Maize* and *Paddi*. Their findings revealed good predictions for the crops studied by them --- estimations that were then employed for crop production estimations for future years in Karnataka.

Methodology

The study involved univariate time series *ARIMA* technique for forecast of the variable of interest. In its general arrangement, an *ARMA* process, as taken from Asteriou & Hall (2007), is mentioned below:

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

Where

Y_t denotes the predicted value of our variable, Y_{t-1}, Y_{t-2}, - - - , Y_{t-p} represent the previous or lagged values of that variable (commonly referred to as the autoregressive terms), ε_t symbolizes the error term, ε_{t-1}, ε_{t-2}, - - - , ε_{t-q} are the past figures of the error term (also called the moving average terms), φ_1, φ_2, - - - , φ_p correspond to the slopes parameters of autoregressive terms, and θ_1, θ_2, - - - , θ_p characterize the slopes parameters of the moving average terms.

The study used the methodology devised by Box and Jenkins (1970) for *ARIMA* modeling that includes three steps --- *identification*, *estimation* and *diagnostic checking*. For the purpose of analysis, weekly data of Karachi Stock Exchange 100 index was collected for 22 years, starting from July 1997 and ending with July 2019, which translated into 1143 observations, a sample sufficiently large for conducting *ARIMA* analysis.

Results

Prior to proceeding with the analysis, the data was checked for its stationarity since *ARIMA* modeling is only valid for trend-stationary data. A simple line plot of our variable was examined for detection of possible trend and, as can be seen in figure 1, the variable had visible trend overtime.

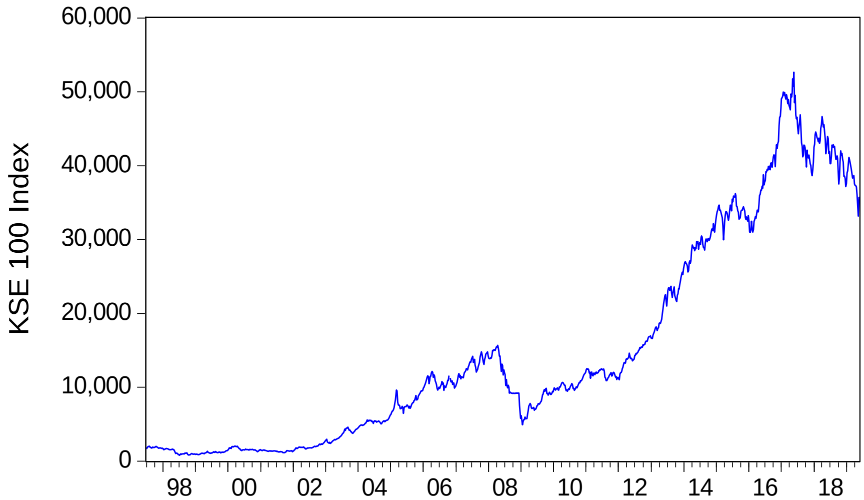

Figure 1: KSE 100 Index --- The Non-Stationary Trend

In order to make the variable under consideration stationary, and thus appropriate for *ARIMA* analysis, weekly *returns* were accounted for by dividing the first difference by the lagged value of the Index for any given week. Figure 2 represents the graph of the weekly returns of KSE 100 Index. As is very visible, taking returns of the index makes the variable absolutely stationary.

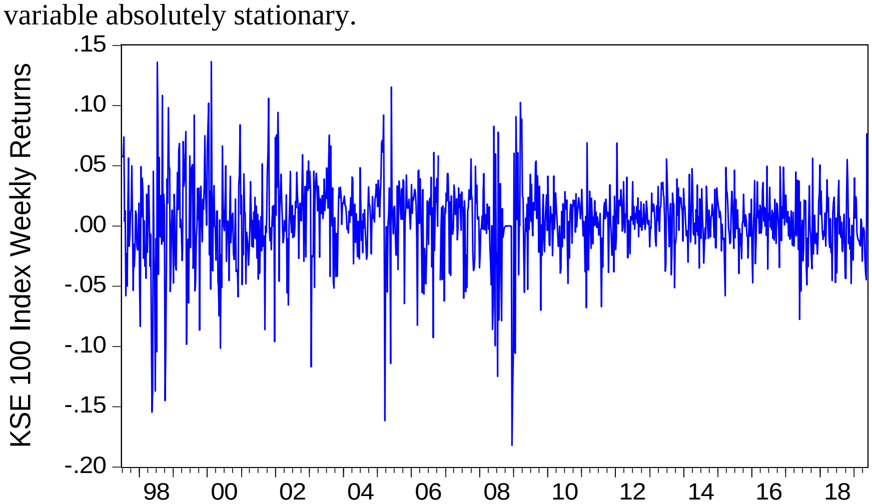

Figure 2: KSE 100 Index --- The Stationary Returns

Model Identification

Identification of the appropriate *ARIMA* configuration is the foremost step in Box-Jenkins method. It involves ascertaining the number of past values of a given variable and the previous error terms that may have a bearing on the current/future value of that variable. This is made possible by inspecting the Correlogram of the variable to see how many *AR* and *MA* terms have significant positive spikes. The correlogram is depicted in table 1.

Table 1: *ACF and PACF of Stationary Returns of KSE 100 Index*

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>|* |</td><td>|* |</td><td>1</td><td>.161</td><td>.161</td><td>29.741</td><td>.000</td></tr>
<tr><td>|* |</td><td>| |</td><td>2</td><td>.080</td><td>.055</td><td>37.040</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>3</td><td>.033</td><td>.012</td><td>38.277</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>4</td><td>.005</td><td>-.006</td><td>38.309</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>5</td><td>-.044</td><td>-.048</td><td>40.527</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>6</td><td>-.049</td><td>-.037</td><td>43.328</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>7</td><td>.019</td><td>.039</td><td>43.727</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>8</td><td>.020</td><td>.020</td><td>44.205</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>9</td><td>-.003</td><td>-.011</td><td>44.212</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>10</td><td>.037</td><td>.034</td><td>45.777</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>11</td><td>-.036</td><td>-.053</td><td>47.253</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>12</td><td>-.039</td><td>-.032</td><td>49.031</td><td>.000</td></tr>
<tr><td>| |</td><td>| |</td><td>13</td><td>.006</td><td>.026</td><td>49.073</td><td>.000</td></tr>
</table>

By inspecting the table, it is found that the partial correlation function has a single spike at lag 1 and then it fades away soon. On the other hand, the autocorrelation function has significant spikes up to lag 2, after which it diminishes. This points out towards usage of *ARIMA* (*1*, *d*, *2*) model. There are, however, some other more popularly used *ARIMA* configurations like *ARIMA* (*1*, *d*, *0*) and *ARIMA* (*2*, *d*, *1*) which are more commonly used in financial economic data. We will, therefore, employ all these and a few other models to be able to select the one that forecasts well and yet has adequate level of parsimony.

Model Estimation

In the model identification stage, we found *ARIMA* (*1*, *d*, *2*) to be the one more suitable for predicting KSE 100 Index returns. We will, however, *try* a few other models as well to find whether there can be a better alternative to the one we found using the Box-Jenkins approach.

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

Dependent Variable: KSE 100 Index Returns
Method: Least Squares
Included observations: 1145 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.002	.001	1.889	.059
AR(1)	.508	.216	2.349	.019
MA(1)	-.355	.218	-1.625	.105
MA(2)	-.006	.050	-.123	.902

R-sqd.	.032	Mean dep. Var	.003
Adj. R-sqd.	.030	S.D. dep. Var	.034
Std. Err. of reg	.034	Akaike info cri	-3.927
Sum of sqd. resid	1.311	Schwarz cri	-3.910
Log likelihood	2252.34	Hannan-Quinn cri	-3.921
F Statistic	12.603	Durbin-Watson	1.998
Prob (F-statistic)	.000		

The regression results for *ARIMA* (*1*, *d*, *2*) are presented in table 2. Surprisingly, only the *AR* (*1*) term has a coefficient that is significant whereas both *MA* (*1*) and *MA* (*2*) terms are highly insignificant. We will now be trying *ARIMA* (*1*, *d*, *1*).

Table 3: *Regression Results using ARIMA (1, d, 0) Model*

Dependent Variable: KSE 100 Index Returns
Method: Least Squares
Included observations: 1145 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.002	.001	2.108	.035
AR(1)	.166	.029	5.703	.000

R-sqd.	.027	Mean dep. Var	.002
Adj. R-sqd.	.027	S.D. dep. Var	.034
Std. Err. of reg	.034	Akaike info cri	-3.926
Sum of sqd. resid	1.317	Schwarz cri	-3.917
Log likelihood	2249.743	Hannan-Quinn cri	-3.923
F Statistic	32.529	Durbin-Watson	2.019

Prob (F-statistic)	.000

Table 3 presents *ARIMA* (*1*, *d*, *0*) model. Of course, the model has only one parameter --- the *AR* (*1*) --- which is significant. However, as per the Box-Jenkins requirements, the two models are to be compared based on the adjusted R^2, *AIC*, *HQC*, *SBC* and the insignificant parameters. *ARIMA* (*1*, *d*, *2*) has a larger adjusted R^2 value than the *ARIMA* (*1*, *d*, *0*) making the former a better model. Nonetheless, the three information criterion values tell a different story. The *AIC* for *ARIMA* (*1*, *d*, *2*) is slightly smaller than that for *ARIMA* (*1*, *d*, *0*). But *ARIMA* (*1*, *d*, *0*) has got smaller values for both the *SBC* and the *HQC* making it preferable than the former model since *SBC* is always preferred over *AIC*. Also, *ARIMA* (*1*, *d*, *0*) does not have any insignificant parameters in contrast with *ARIMA* (*1*, *d*, *2*) which has two insignificant ones.

It will be wiser to check and compare for the results the *ARIMA* (*1*, *d*, *1*) presents.

Table 4: *Regression Results using ARIMA (1, d, 1) Model*

Dependent Variable: KSE 100 Index Returns
Method: Least Squares
Included observations: 1145 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.002	.001	1.894	.056
AR(1)	.492	.130	3.798	.000
MA(1)	-.340	.140	-2.421	.016

R-sqd.	.032	Mean dep. Var	.002
Adj. R-sqd.	.030	S.D. dep. Var	.034
Std. Err. of reg	.034	Akaike info cri	-3.929
Sum of sqd. resid	1.311	Schwarz cri	-3.916
Log likelihood	2252.330	Hannan-Quinn cri	-3.924
F Statistic	18.910	Durbin-Watson	1.996
Prob (F-statistic)	.000		

ARIMA (*1*, *d*, *1*) as depicted in table 4 has all significant parameters. The adjusted R^2 is equal to that of *ARIMA* (*1*, *d*, *2*) model or .03. But the information criterion values are all less than that for *ARIMA* (*1*, *d*, *2*) making it undoubtedly a better model. When compared with *ARIMA* (*1*, *d*, *0*), *ARIMA* (*1*, *d*, *1*) has a larger adjusted R^2 and very close information criterion values. However, based on the R^2 value, *ARIMA* (*1*, *d*, *1*) takes the lead.

Diagnostic Checking

So, what did we conclude finally? Obviously looking at the models offered so far, there seems to be no conclusive model that outclasses its rivals. Let us now summarize all those, and a few other, models into a table for a diagnostic check.

Table 5: *A Comparison of ARIMA models: The row with bold figures indicates the most suitable model*

ARIMA Model	Adjusted R^2	AIC	SBC	HQC	Insignificant lags
(*1*, *d*, *0*)	.027	-3.926	-3.917	-3.923	None
(*1*, *d*, *1*)	**.030**	**-3.929**	**-3.916**	**-3.924**	**None**
(*2*, *d*, *1*)	.028	-3.927	-3.910	-3.921	Three
(*3*, *d*, *1*)	.026	-3.927	-3.905	-3.919	Two
(*1*, *d*, *2*)	.030	-3.927	-3.910	-3.921	Two
(*1*, *d*, *3*)	.029	-3.926	-3.903	-3.917	Four
(*2*, *d*, *2*)	.032	-3.930	-3.908	-3.921	None
(*3*, *d*, *3*)	.035	-3.934	-3.903	-3.923	One

The combined estimates of all models, presented in table 5, give interesting but perplexing findings. If adjusted R^2 is to be considered the factor based on which a given model will be accepted, then *ARIMA* (*3*, *d*, *3*) is the best model followed by *ARIMA* (*2*, *d*, *2*). On the basis of the *AIC* values again, *ARIMA* (*3*, *d*, *3*), having the minimum *AIC*, takes the lead over the rest. However, *SBC* which is more often preferred over the other information criteria signals at *ARIMA* (*1*, *d*, *0*) followed by *ARIMA* (*1*, *d*, *1*) with the former having only .001 lesser value than the latter. Nonetheless, *ARIMA* (*1*, *d*, *1*) should be considered better than *ARIMA* (*1*, *d*, *0*) for multiple reasons in that it has a clearly larger adjusted R^2 value and lower *AIC* and *HQC* values than *ARIMA* (*1*, *d*, *0*). The whole debate can, however, be concluded by giving an edge to *ARIMA* (*1*, *d*, *1*) model over all others by jointly considering all the factors that lead to selection of a given model.

Discussion

ARIMA method of forecasting seems to be a good option to be used if prediction is to be made for the short run. However, what is surprising is that the most appropriate model as per the identification step of the Box-Jenkins methodology appeared to be *ARIMA* (*1*, *d*, *2*). This was identified by inspecting the correlogram of autocorrelation and partial correlation functions. However, when it came to using the diagnostic checking for models, *ARIMA* (*1*, *d*, *1*) took a clear edge over *ARIMA* (*1*, *d*, *2*) in terms of its supremacy. Hence it was concluded that *ARIMA* (*1*, *d*, *1*) should be employed to better predict weekly stock returns at Karachi Stock Exchange.

Looking at the previous studies, one finds that *ARIMA* (*1*, *d*, *1*) has been one of the most widely used *ARIMA* configurations for anticipating many time series. This makes the current work more in line with the previous work. However, irrespective of which specific *ARIMA* configuration is used more, the *ARIMA* technique in general has been found to have helped many researchers speculate their time series variables in the past. To give examples of a few, Manoj and Madhu (2014) used the model for estimating the production of sugarcane in India and found *ARIMA* (*2*, *d*, *1*) to be more appropriate model. In the similar manner, the model was also found to be very helpful for prediction by Hamjah (2014) who used it for forecasting rice production in Bangladesh.

Some studies that were conducted specifically to forecast stock prices also used *ARIMA* modeling. Among them were Mondal, Shit and Goswami (2014), for instance, who incorporated the model for predicting 56 Indian stocks, and Adebiyi, Adewumi and Ayo (2014) who concluded that *ARIMA* method clearly outclassed the traditional methods of time series forecasting, especially when it came to predicting stock prices.

Some studies, however, found *ARIMA* model not good enough to adequately forecast time series variables in the short run. Gay (2016), for instance, tried to predict prices of stocks in BRIC countries but could not get satisfactory results. However, considered overall, *ARIMA* modeling offers a good starting point to predict most time series variables.

Conclusion

One way of making a smart guess about how an economy is going overall is by looking at how its stock market is performing. A regular and steady increase in the stock Index over time is indicative of a sustainably growing economy. Most of the stock markets globally experience an upward moving trend in the long run owing to an increase in demand for products and services, among many other things, over time. From the perspective of investment in stock market, what is more important for investors is their ability to predict in the *short run* the direction the index is expected to move to. Hence, investors need to use different ways of forecasting stock returns in the short run. The *ARIMA* method of forecasting is but one of such ways to achieve the objective. Although traditional ways of predicting a variable based on *factors* that affect it are still available, they are not efficient enough in the short run though they may, of course, be more relevant for the long term.

The study at hand also used, therefore, the *ARIMA* modeling by using the famous Box-Jenkins methodology for forecasting stock returns at Karachi Stock Exchange. It has been concluded that the model is able to effectively forecast our variable of interest in the short run. Operationally speaking, the identification step of the Box-Jenkins method surfaced *ARIMA* (*1, d*, *2*) to be the most fitting model for forecasting. However, the diagnostic checking argued for the relevance of *ARIMA* (*1*, *d*, *1*), a configuration widely accepted by most of the researchers as the one predicting most time series variables. Put it the other way, the one-period lagged value of returns and the one-period lagged value of the error term were most instrumental in determining the future value of returns. As a matter of policy, the study might serve as a guiding tool for prospective investors wishing to invest in Karachi Stock Exchange (now Pakistan Stock Exchange).

References

Adebiyi, A., Adewumi, A., & Ayo, C. (2014, March). *Stock Price Prediction Using the ARIMA Model*. Paper presented at the 2014 UKSim-AMSS 16th International Conference on Computer Modeling and Simulation, Cambridge University, United Kingdom. Retrieved from http://ijssst.info/Vol-15/No-4/data/4923a105.pdf.

Asteriou, D., & Hall S. (2007). *Applied Econometrics*, 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.

Contreras, J., Espinola, R., Nogales, F., & Conejo, A. (2003). ARIMA Models to Predict Nextday Electricity Prices. *IEEE Transactions on Power Systems, 18*(3), 1014- 1020.

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.

Gilbert, K. (2005). An ARIMA Supply Chain Model, *Management Science, 51*(2), 305-310.

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*, Fourth 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.

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.

Padhan, P. (2012). Application of ARIMA Model for Forecasting Agricultural Productivity in India. *Journal of Agriculture & Social Sciences*, *8*(2), 50–56.

Raymond Y. (1997). An Application of the ARIMA Model to Real-Estate Prices in Hong Kong. *Journal of Property Finance, 8*(2), 152-163.