Will the Stock Market Index Upsurge or Deflate? Making Calculated Predictions Using the Univariate Autoregressive Integrated Moving Average Technique

Mustafa Afeef[*] Nazim Ali[†] Adnan Khan[‡]

Abstract *Movements in a stock market index may safely be considered one of the most-watched out phenomena by investors in almost every economy. One method to forecast the index is to study all those external factors that directly affect it. Another way, however, is to base one's predictions on the past behavior of the variable of interest. This paper has employed the method described latter and has, therefore, made use of the ARIMA modeling. In this connection, the daily stock market index data of the Karachi Stock Exchange 100 index was taken for twenty years from 1997 to 2017 which translated into 4940 observations. The study revealed that the model was decently efficient in forecasting the KSE 100 Index, though only for the short-range. The upshot of this study may be utilized specifically by short term investors in deciding on when, and when not, to invest in the stock market.*

Introduction

The key to a lucrative investment lies in how accurately one can predicts the future. While dealing with decisions concerning capital budgeting, the most important aspect to consider in an investment is the uncertainty with which future benefits are expected to be realized from it. Given the very basic goal of financial management of maximizing stockholder's wealth, most of the subject seems to be future-oriented. Therefore, the ability to forecast well a given investment's future value is of utmost importance. The current study also deals with predicting the future value of stocks

---

[*] Assistant Professor, Department of Management Sciences, Islamia College Peshawar, KP, Pakistan.

[†] Assistant Professor, Department of Management Studies, University of Malakand, KP, Pakistan.
Email: nazimali100@yahoo.com

[‡] Lecturer, Department of Management Studies, University of Malakand, KP, Pakistan.

as a whole. Therefore, the stock market index which is the representative of all stocks in a given market is the focus of the study. Although individual stocks are, of course, always subject to unique fluctuations due to factors that are mostly firm-specific, the stock market index only increases (or decreases) when there is a clear overall increase (or decrease) in the prices of most stocks listed on the concerned exchange.

Factors that affect the prices of most stocks (and hence affect the stock market index) in a given market are mostly macroeconomic in nature. Hence, an attempt to predict the future direction of the stock market index would entail determining all those factors that influence the prices as a whole. This is, however, a tedious job and has its complications. For instance, it may also require whether those influencing macroeconomic factors are expected to increase (i.e., get better) or decrease (i.e., worsen) in the future. There is, however, a rather simpler method of estimating a variable's future values, and that is to anticipate its future measurements based on its previous measurements. Therefore, the method is also known in the academic text as the Box-Jenkins methodology. The method allows for the lagged values of a variable as well as its error term to forecast the most likely current (or future) value of the variable.

The rationale behind the study, thus, is to judge whether the methodology devised by Box and Jenkins (1970) in forecasting any given time series works well in predicting the stock market index. It is also aimed to get to know how many lagged values of the stock market index are effectively involved in the prediction of the current (or future) value of the index. We, therefore, have dual objectives of the study --- firstly, to examine if Box-Jenkins methodology would help at all in determining the direction the stock market index is expected to go, and secondly, to see how many previous values of the index and the error term is the current value of the index dependent upon. The study will give insights to potential investors in determining the appropriate time they should raise (or retrench) their investments in stock markets.

Literature Review

There has been a good deal of work undertaken on the use of the Box-Jenkins method for the prediction of time series variables. The method has worked efficiently well for many of the longitudinal data but not for all. The related literature has been, however, divided into two portions with the first part mentioning studies carried out previously by researchers to forecast stock prices or stock market index using the Box-Jenkins method. The second part gives examples of other research work that was aimed at finding the future probable values of various time series variables except for the stock prices or stock indices. The ensuing lines present those previously completed studies that have made use of the model:

Stock returns were also predicted by Mondal, Shit, and Goswami (2014) who made a big effort to take fifty-six (56) Indian stocks by classifying them into their respective sectors. They then endeavored to estimate their future stock returns by using the Box-Jenkins method and found that the model correctly forecasted returns for around 85% of the time for all sectors.

Another attempt of estimating stock return using the Box-Jenkins method was made by Adebiyi, Adewumi, and Ayo (2014) who made use of the model for predicting stock prices of Nokia and Zenith Bank. They also observed that the model was a good predictor of returns in the short run. Similarly, Banerjee (2014) also estimated the Indian stock market index using the ARIMA model and found that the results were very precise in the short run.

The Box-Jenkins method has also been used by researchers for estimating various time series data other than stock returns or stock index and has more often got very encouraging results. For instance, one of the initial users of the model was Jarrett (1990) who employed the model for estimating corporate earnings. He observed no obvious difference between the conventional pre-specified models and the one based on *ARIMA* modeling in terms of prediction errors. The model was also used by Raymond (1998) for anticipating real estate prices in Hong Kong. He observed trends in property prices in the country. Similarly, Meyler, Kenny, and Quinn (1998) also utilized the method to forecast inflation in Ireland. They concluded that the more important thing in a model is the minimization of prediction errors rather than maximization of the goodness of fit. Contreras, Espinola, Nogales, and Conejo (2003) also employed *ARIMA* for estimating the prices of electricity in California &Spain and were successful. The model was also engaged by Gilbert (2005) while dealing with a multistage supply chain model. He found that the customers' orders and the inventories are *ARIMA* processes in the same way as the customers' demands and the lead times. Among the users of the Box-Jenkins method was also Guha (2016) who attempted to determine the prices of gold in India and who claimed that the results could have been used for deciding when to invest in the gold market.

The Box-Jenkins method also seems to be popularly used by researchers who aimed at determining the production of crops. For example, Manoj and Madhu (2014) appeared to employ the method for anticipating the Indian production of Sugarcane. It was known that the model could forecast Sugarcane production for around five years. A very similar study was conducted by Hamjah (2014) who wanted to estimate rice production in Bangladesh. He also found that the method devised by Box and Jenkins for time series forecasting worked very well but in the short run. The productivity of crops in India was also assessed by Padhan (2012) who included 34 Indian crops in her study. Her study gave the best prediction for the *tea* crop which had the lowest mean absolute percent error. The worst predicted crop was found to be that of *papaya*. To discuss more, Jadhav, Reddy, and Gaddi (2017) also endeavored to forecast prices of major crops in Karnataka, India

including Maize, Ragi, and Paddi crops. The results showed very accurate predictions based on which they estimated the production of these crops for the next three years.

The Box-Jenkins Method

The Box-Jenkins method developed by statisticians George Box and Gwilym Jenkins entails a way of forecasting a time series data (Box and Jenkins, 1970). The method is, of course, based on the autoregressive integrated moving average (or *ARIMA*) modeling in which a variable's future values are estimated based on its lagged values as well as its lagged error terms. As a matter of fact, the larger the set of values available for a given time series, the better is the model's ability to predict future values. Chatfield (1996) recommends a minimum of 50observations of a time series to be available for a reliable forecast of its future values. Some researchers, however, have the understanding that the observations should preferably exceed 100 for a good prediction.

The Box-Jenkins methodology is designed to guide the researcher in determining the number of lagged values of a variable and lagged values of its disturbance term that particularly *influence* or *decide* its current or forthcoming value. This is not an easy task, however.But one of the basic principles of the method is to devise a model as parsimonious as possible.Although a large and an over-parameterized model will tend to maximize the model's R^2, introducing parsimony in a model entails that minimum regressors are kept. An advantage of this would be that we would not lose excessive degrees of freedom utilizing keeping as low the number of regressors in the model as possible.

Research Methodology

The current study has a single time series data of the stock market index. Therefore, univariate *ARIMA* modeling has been employed to predict the most probable future value of our variable of interest. The general form an *ARMA* process as adapted from Asteriou & Hall (2007) is:

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

Where,

Y_t is the anticipated value of our variable of interest, Y_{t-1}, Y_{t-2}, - - -, Y_{t-p} are the previous values of that variable (also known as the autoregressive terms), ε_t is the disturbance term, ε_{t-1}, ε_{t-2}, - - -, ε_{t-q} are the previous values of the disturbance term(also called the moving average terms), φ_1, φ_2, - - -, φ_p are the slopes or coefficients of autoregressive regressors, and θ_1, θ_2, - - -, θ_p are the slopes or coefficients of the moving average regressors.

An *ARMA* process can only be run over a stationary time series. A stationary time series has a constant mean and variance over time and a time-invariant covariance (Gujarati & Porter, 2004). If, however, a series is non-stationary, it will have to be differenced for enough times to render it completely stationary. Since most of the time series are non-stationary and such was the case for our variable too, an *ARIMA* process was employed which is used for integrated or non-stationary data and which allows the differenced values of the variable to be included in the model.

The study used daily stock market index data of the Karachi Stock Exchange 100 index for 20 years from December 1997 to December 2017. This translated into 4940 observations, a size large enough for carrying the *ARIMA* analysis safely (Chatfield, 1996).

Analysis and Findings

The daily stock market index data of the Karachi Stock Exchange was collected for 20 years covering the time from December 1, 1997, to December 3, 2017. Before subjecting a time series data for ARIMA forecasting, it is necessary to assure that the series does not have trends and is stationary. In order to check whether the same was the case with our time series, a simple line plot was inspected for the variable which depicted trends in the data.
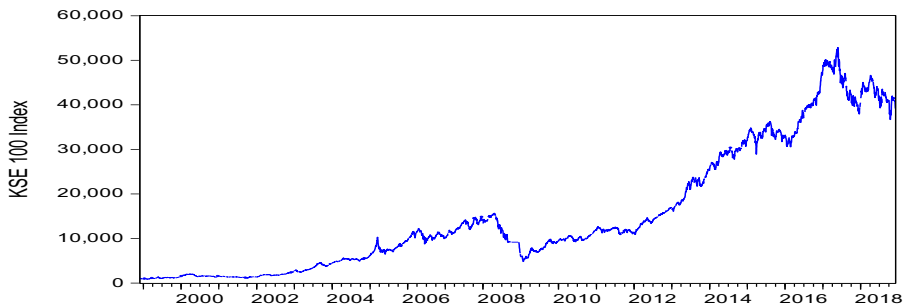

Figure 1: *The Non-Stationary KSE 100 Index*

Figure 1 represents the line graph of KSE 100 index evidently portraying the non-stationary or trended nature of the series. The result was validated by using the Augmented Dickey Fuller (ADF) test to check whether or not the variable had a unit root. The ADF test, as can be seen in table 1, was highly insignificant again proving that the time series had a unit root and therefore was non-stationary.

We proceed by making the correlogram of our time series variable. Theoretically, the correlogram of any stationary time series should fade away for higher lag

Table 1: *Augmented Dickey Fuller Test for KSE 100 Index*

Null Hypothesis: KSE 100 Index has a unit root
Exogenous: Constant
Lag Length: 7 (Automatic - based on SIC, maxlag=31)

		t-Statistic	Prob.
Augmented Dickey-Fuller test statistic		.849	.995
Test critical values:	1% level	-3.432	
	5% level	-2.862	
	10% level	-2.567	

lengths. But the correlogram of our KSE 100 Index variable does not fade away at all which again hints towards nothing else but the non-stationary dimension of the variable. The correlogram of our variable of interest is shown in table 2.

Table 2: *ACF and PACF of KSE 100 Index*

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
|*******	|*******	1	.965	.965	4602.3	.000
|*******	|***	2	.957	.376	9130.3	.000
|*******	|**	3	.952	.223	13614	.000
|*******	|*	4	.951	.191	18085	.000
|*******	|**	5	.954	.234	22588	.000
|*******	|	6	.949	.063	27047	.000
|*******	|	7	.947	.072	31488	.000
|*******	|*	8	.946	.077	35918	.000
|*******	|*	9	.946	.089	40351	.000
|*******	|*	10	.947	.090	44795	.000
|*******	|	11	.945	.038	49219	.000
|*******	|	12	.944	.056	53639	.000
|*******	|	13	.944	.055	58056	.000
|*******	|	14	.945	.063	62480	.000
|*******	|	15	.944	.029	66896	.000
|*******	|	16	.942	.027	71301	.000
|*******	|	17	.941	.012	75693	.000

For the induction of stationary in the variable, daily *returns* have been calculated instead of simply taking the first differences. The returns have been computed by dividing the difference in stock index values of any two consecutive days by the previous day. Of course, this practice also involves taking differences. Hence it is expected that transforming the index values into *returns* would effectively make the variable stationary and remove trends.
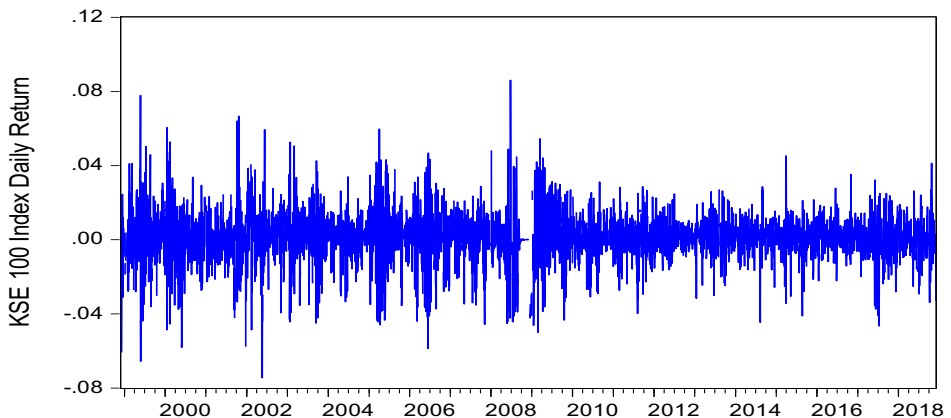
Figure 2: *The Stationary KSE 100 Index Returns*

As can be seen in figure 2, the returns of the KSE 100 Index are completely stationary. This is also evident from the ADF test which is highly significant (*p*-value =.000) showing that the variable has no unit root (see table 3).

Table 3: *Augmented Dickey-Fuller Test for KSE 100 Index Returns*

Null Hypothesis: KSE 100 Index Returns has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=31)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-60.658	.000
Test critical values:	1% level	-3.432	
	5% level	-2.862	
	10% level	-2.567	

Model Identification

After achieving stationarity in our variable of interest by taking its daily returns in percentages, we proceed by applying the Box-Jenkins methodology step-by-step. The first stage of the methodology is identifying the model. This refers to correctly identifying the number of lagged values of the variable and that of the error term that effectively influences the variable. Therefore, in this first step, a Correlogram is again made to inspect the number of *AR* and *MA* terms that are expected to affect the current value of the variable.

Table 4 gives the correlogram of the daily returns of the KSE 100 Index for

Table 4: *ACF and PACF of KSE 100 Index Returns*

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
⏐* ⏐	⏐* ⏐	1	.104	.104	50.984	.000
⏐ ⏐	⏐ ⏐	2	.043	.033	59.837	.000
⏐ ⏐	⏐ ⏐	3	.045	.037	69.339	.000
⏐ ⏐	⏐ ⏐	4	.038	.029	76.348	.000
⏐ ⏐	⏐ ⏐	5	.013	.003	77.135	.000
⏐ ⏐	⏐ ⏐	6	.021	.016	79.227	.000
⏐ ⏐	⏐ ⏐	7	.008	.002	79.547	.000
⏐ ⏐	⏐ ⏐	8	.010	.006	80.002	.000
⏐ ⏐	⏐ ⏐	9	.035	.032	85.923	.000
⏐ ⏐	⏐ ⏐	10	.052	.044	98.623	.000
⏐ ⏐	⏐ ⏐	11	.001	-.012	98.625	.000
⏐ ⏐	⏐ ⏐	12	.028	.024	102.49	.000
⏐ ⏐	⏐ ⏐	13	-.001	-.012	102.50	.000

the period from December 1997 to December 2017.

The Box-Jenkins methodology requires the researcher to set an upper limit for *p* and *q* (i.e., p_{max} and q_{max}) and then estimate all probable models whose *p* and *q* values are between 0 and the maximum value set for each of them.

Looking at table 4, it is evident that both the autocorrelation function and the partial autocorrelation function have a single positive spike at Lag 1 and then both

tend to die down immediately. This infers an *ARIMA* (*1, d, 1*) model to be exercised. However, one of the most common *ARIMA* configurations often used in business and economic data is that of *ARIMA* (*1, d, 0*) and *ARIMA* (*2, d, 1*). We will, however, have to estimate all these probable models and decide the one that is the most parsimonious.

Model Estimation

The Box-Jenkins method has prescribed *ARIMA* (*1, d, 1*) to be the most suitable one for forecasting the returns of our KSE 100 Index. In this segment, however, we will also estimate a few other *probable* models close to *ARIMA* (*1, d, 1*) to see whether any other *ARIMA* configuration can, or cannot, better predict the current (or future) value of our variable.

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

Dependent Variable: KSE 100 Index Returns
Method: Least Squares
Included observations: 4939 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.001	.000	3.327	.001
AR(1)	.803	.049	16.353	.000
MA(1)	-.738	.056	-13.257	.000
R-squared	.015	Mean dependent var		.001
Adj. R-squared	.014	S.D. dependent var		.014
S.E. of regression	.013	Akaike info criterion		-5.767
Sum squared resid	.903	Schwarz criterion		-5.763
Log likelihood	14245.67	Hannan-Quinn cri.		-5.766
F-statistic	36.88717	Durbin-Watson stat		1.944
Prob(F-statistic)	.000			

Table 5 presents the regression estimation for *ARIMA* (*1, d, 1*) model. Both the *AR* and *MA* terms have strong coefficients that are highly significant. However, in order to compare the R^2, *AIC,* and *SBC* of the model with other possible models, we estimate those other probable *ARIMA* configurations as well.

Table 6: *Regression Results using ARIMA (1, d, 0) Model*

Dependent Variable: KSE 100 Index Returns
Method: Least Squares
Included observations: 4939 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.001	.000	3.826	.000
AR(1)	.101	.014	7.102	.000
R-squared	.010	Mean dependent var		.001
Adj. R-squared	.009	S.D. dependent var		.014
S.E. of regression	.014	Akaike info criterion		-5.763
Sum squared resid	.908	Schwarz criterion		-5.761
Log likelihood	14234.14	Hannan-Quinn cri.		-5.762
F-statistic	50.432	Durbin-Watson stat		2.003
Prob(F-statistic)	.000			

The model *ARIMA* (*1, d, 0*) is given in table 6. A comparison between the two models articulates that *ARIMA* (*1, d, 1*) has a larger R^2 value than that of *ARIMA* (*1, d, 0*). Also, all of the three information criteria values, like the Akaike information criterion, the Schwarz Bayesian criterion, and the Hannan-Quinn criterion, have slightly lower values for *ARIMA* (*1, d, 0*) when compared with those

for the *ARIMA* (*1*, *d*, *0*) model. Hence, *ARIMA* (*1*, *d*, *1*) seems to be better than *ARIMA* (*1*, *d*, *0*) in terms of its ability to forecast. Let us also check *ARIMA* (*2*, *d*, *1*) model for comparison with *ARIMA* (*1*, *d*, *1*).

Table 7: *Regression Results using ARIMA (2, d, 1) Model*

Dependent Variable: KSE 100 Index Returns
Method: Least Squares
Included observations: 4938 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.001	.000	3.458	.000
AR(1)	.844	.085	9.971	.000
AR(2)	-.036	.019	-1.813	.069
MA(1)	-.752	.083	-9.034	.000
R-squared	.016	Mean dependent var		.001
Adj. R-squared	.015	S.D. dependent var		.014
S.E. of regression	.014	Akaike info criterion		-5.769
Sum squared resid	.901	Schwarz criterion		-5.764
Log likelihood	14247.88	Hannan-Quinn cri.		-5.767
F-statistic	26.863	Durbin-Watson stat		1.996
Prob(F-statistic)	.000			

The aforementioned table gives the estimation results of the *ARIMA* (*2*, *d*, *1*) model. Surprised as one may feel, this model seems to be better than the Box-Jenkins approved *ARIMA* (*1*, *d*, *1*) configuration. The R^2 value for *ARIMA* (*2*, *d*, *1*) is .0161 compared to .0148 which belongs to *ARIMA* (*1*, *d*, *1*). In the same manner, all the information criteria values including *AIC*, *SBC*, and *HQC* are lower for *ARIMA* (*2*, *d*, *1*) than for *ARIMA* (*1*, *d*, *1*). To compare, the values of *AIC*, *SBC*, and *HQC* are -5.769088, -5.763819, and -5.767240 respectively for *ARIMA* (*2*, *d*, *1*) and -5.767432, -5.763481and -5.766047 respectively for *ARIMA* (*1*, *d*, *1*).This hints the supremacy of *ARIMA* (*2*, *d*, *1*) over *ARIMA* (*1*, *d*, *1*). However, the Box-Jenkins methodology also gives weight to the model having the least number of insignificant parameters. By looking at the results again we notice that both the parameters of *ARIMA* (*1*, *d*, *1*) are highly significant. On the other hand, *ARIMA* (*2*, *d*, *1*) has one insignificant term --- the *AR* (2) term, which is insignificant at 5% level but significant at 10% (*p*-value = .0699). Also, the coefficient of *AR* (2) is very small depicting that it may have the least influence in forecasting the current value of the variable. However, the inclusion of *AR* (2) term has slightly increased the parameters of both the *AR* (1) and the *MA* (1) terms in the model. This may be deemed as a positive overall impact of the insertion of *AR* (2) term on the model. Nonetheless, let us also experience an over-parameterized model, the *ARIMA* (*3*, *d*, *3*) to see whether it comes up with an even better prediction of our variable.

Table 8: *Regression Results using ARIMA (3, d, 3) Model*

Dependent Variable: KSE 100 Index Returns
Method: Least Squares
Included observations: 4937 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	.001	.000	3.411	.000
AR(1)	-.623	.059	-10.472	.000
AR(2)	.037	.082	.450	.653
AR(3)	.743	.059	12.621	.000
MA(1)	.698	.066	10.532	.000
MA(2)	.069	.091	.754	.451
MA(3)	-.675	.066	-10.196	.000
R-squared	.018	Mean dependent var		.000
Adj. R-squared	.017	S.D. dependent var		.014
S.E. of regression	.013	Akaike info criterion		-5.774
Sum squared resid	.896	Schwarz criterion		-5.765
Log likelihood	14259.66	Hannan-Quinn cri.		-5.771
F-statistic	15.106	Durbin-Watson stat		1.958
Prob(F-statistic)	.000			

The results of *ARIMA* (*3, d, 3*) are presented in table 8. The R^2 has increased further, although very marginally. Also, the information criterion values are all slightly lower than that for *ARIMA* (*1, d, 1*). However, this model has now two highly insignificant terms --- the *AR* (2) and the *MA* (2). Based on the parsimony, this model will be rejected instantaneously. But there is one important thing to note in the results of this model and that is the highly insignificant *AR* (2) term. This term was also found to be insignificant in the *ARIMA* (*2, d, 1*), the model which had otherwise better results than the Box-Jenkins recommended *ARIMA* (*1, d, 1*). Hence, this also signals towards relative *supremacy* of *ARIMA* (*1, d, 1*) over *ARIMA* (*2, d, 1*).

Diagnostic Checking

For a thorough comparison of all the possible *ARIMA* configurations, the R^2
  values, the three information criterion values, and the number of insignificant terms for each *ARIMA* model have been summarized in table 9.

Table 9: *Comparing ARIMA models: The row with bold figures indicates the most fitting model*

ARIMA Model	Adj. R^2	AIC	SBC	HQC	Ins. Lags
ARIMA (*1, d, 0*)	.009	-5.763	-5.761	-5.762	None
ARIMA (1, d, 1)	**.014**	**-5.767**	**-5.763**	**-5.766**	**None**
ARIMA (*2, d, 1*)	.015	-5.769	-5.764	-5.767	One
ARIMA (*3, d, 1*)	.013	-5.770	-5.763	-5.768	One
ARIMA (*1, d, 2*)	.016	-5.769	-5.763	-5.767	None
ARIMA (*1, d, 3*)	.016	-5.768	-5.762	-5.766	One
ARIMA (*2, d, 2*)	.015	-5.769	-5.762	-5.766	Two
ARIMA (*3, d, 3*)	.017	-5.774	-5.765	-5.771	Two

The above table compares eight various *ARIMA* models based on the criteria formulated by the Box-Jenkins method. Of these, *ARIMA* (*1, d, 1*) seems to be the most appropriate (the one highlighted in bold letters). However, *ARIMA* (*2, d, 1*) is also a feasible option since it has a higher R^2 and lower information criterion values than *ARIMA* (*1, d, 1*). However, the *AR* (2) term in *ARIMA* (*2, d, 1*) is insignificant at a 5% level which again increases the weightage for *ARIMA* (*1, d, 1*) specification.

Discussion

The findings of the study discussed categorically in the previous section illustrate that the *ARIMA* technique of forecasting a time series is a worthwhile option. The current work has used this method for predicting the stock market index and it has been found that the model works reasonably well in the short run. As has also been discussed in the previous section, the two configurations, i.e., *ARIMA* (*1, d, 1*) and

ARIMA (*2*, *d*, *1*), were found to be very helpful in forecasting stock market index returns. However, of the two, *ARIMA* (*1*, *d*, *1*) was found to be a bit better and this is exactly the model which was also figured out by the Box-Jenkins approach of model selection. In a simpler language, it was observed that the current returns of the stock market index could be estimated by monitoring a one-year previous value of the returns as well as a one-year previous value of the error term.

The results of the study are much in line with what previous studies have broadly found out. To discuss a few, for instance, the model was also employed by Manoj and Madhu (2014) for estimating Indian sugarcane production. They discovered the model helpful and found *ARIMA* (*2*, *d*, *1*) to be the most fitting configuration for predicting the production of the crop. The same was the findings of Hamjah (2014) who used the model for estimating the production of rice in Bangladesh and found that the model was helpful in short term prediction.

Studies conducted to estimate future stock prices also found the Box-Jenkins model very efficient in the short run. Mondal, Shit, and Goswami (2014), for example, employed the model for stock price prediction using the data of 56 Indian firms. They found the model correctly predicting the future returns of the stocks understudy for around 85% of the time. Another study to predict stock prices using the said model was conducted by Adebiyi, Adewumi, and Ayo (2014). They also concluded that the model had a decent prediction power and worked better than conventional models specifically in forecasting stock prices.

There were, however, a few studies that did not find the Box-Jenkins method helpful in predicting future values. For instance, Gay (2016) wanted to find the impact of macroeconomic factors on stock prices using the data of BRIC countries. He also used the *ARIMA* model to check whether stock prices in these countries could have been forecasted using their lagged values. He, however, noticed that although stock prices were significantly associated with those macroeconomic factors, there was no significant link between the current and the previous stock prices for any of these countries.

In a nutshell, the model is a good predictor of time series data in the short run as evidenced by most of the empirical works undertaken in various disciplines.

Conclusion

The stock market index is normally believed to be the indicator of *prosperity* or *disaster* in an economy. Investors often procure symptoms of the health of an economy by reviewing the upward or downward movements of stock prices in stock exchanges before making any significant investment decision. An ability to better predict in which direction the stock market will go would, of course, give the relevant investors an incentive to make decisions about when, and when not, to invest. Hence, in a way, it can be said that the stock market index state of affairs

at any given time is a type of market information that is relevant to all types of investors in a given economy.

The current study estimated movements in the stock market index by using the Box-Jenkins method which forecasts the current value of a time series based on its lagged values as well as its lagged error terms. This method has proved to be rather more effective in the past for estimating many time series variables than the conventional method of allowing a variable to vary following all the factors that theoretically influence it. The current study, therefore, also found the Box-Jenkins method quite effective in forecasting the stock market index in the short run. It was however distinctively found that the one-period previous returns and the one-period previous error term were most operatives in ascertaining the current returns of the index. As a guide for investment, the findings of the study may be used by prospective investors in foreseeing how much return to expect *subsequent* form Karachi Stock Exchange 100 Index.

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 the 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.

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

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.

Meyler, A., Kenny, G., & Quinn, T. (1998). Forecasting Irish inflation using ARIMA models. *Central Bank and Financial Services Authority of Ireland Technical Paper Series, 1998*(3), 1-48.

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.