ENVISAGING KSE 100 INDEX USING THE BOX-JENKINS METHODOLOGY

Mustafa Afeef, Shahid Jan, Fayaz Ali Shah, Hamid Ullah, Raza Ullah,
Department of Management Sciences, Islamia College Peshwar, KP, Pakistan.
Email: mustafa@icp.edu.pk

Received November 18, 2019
Accepted December 09, 2019

Abstract. *Investment in stock portfolios has never been a risk-free course of action as countless factors impinge on the end result of such a venture. Although fairly rewarding, the element of uncertainty involved keeps many potential investors away as they fail to adequately forecast what moves the stock market is going to make in the near future. The enticement of receiving returns, however, is appealing enough for investors to have their money invested in the stock market. But the ability to forecast the market remains their major necessity. In operational terms, there are two ways of forecasting the current and future values of any time series including stock indices. One way is to regress stock returns over all those factors that have an effect on stock market performance. The other method is making predictions on the basis of the past performance of the stock market. The current paper has adopted the second method of forecasting and has made use of the autoregressive integrated moving average (ARIMA) technique. Monthly stock returns data of KSE 100 Index was collected from 1997 to 2019 which translated into 266 observations. It was realized that the technique used in the study helped in adequately predicting stock returns, although only in the short run. The outcomes of this study may be of help for prospective stock market investors, specifically short-term, in deciding when, and when not, to extend their investments at Pakistan Stock Exchange.*

Keywords: ARIMA, Box-Jenkins methodology, KSE-100 Index, prediction

Introduction

The ability to predict the future can never be underestimated when it comes to investments. Since the future will always remain uncertain, investors will almost always be wondering about finding the appropriate time to invest their excess money. Stock market index speculation is no different than prediction of other types of investment as there are many factors involved with some being very hard to predict. The movement of a stock market index represents the direction the economy is heading towards. Therefore, an increase in the Index connotes an increase in the share prices of companies of all, or most, of the sectors in an economy. It is probably for this reason that almost all type of investors do try to keep an eye over what is happening to the stock market index.

Primarily there are two distinct methods of forecasting any time series. One method is to anticipate the direction the variable, in our case the stock market, is expected to move keeping in view all the factors (taken as explanatory variables in the regression model) that potentially affect it. Although more rational, this method involves a huge amount of data to be collected for all those influential factors. Also, many other, rather invisible or unmeasurable, factors that may affect the dependent variable may be missed out leading to biased results. There is, however, another method as well of forecasting a given time series, and that is, to anticipate its forthcoming values on the basis of its past values. In time series econometrics, models facilitating such kind of forecasting are becoming increasingly popular thanks to their superior ability to predict a given variable. The current study also uses one such model commonly known as the autoregressive integrated moving average model or the *ARIMA* model. This model allows for a time series to be forecasted based on its values in the previous period and also based on the previous values of its error term.

Box and Jenkins in their famous book *'Time Series Analysis: Forecasting and Control'* formulated a method, commonly known as the Box-Jenkins methodology, which enables a researcher to identify how many lagged values of a given variable and that of the error term effectively predict the future value of that variable (Box & Jenkins, 1970). The method, in a sense, employs what Gould was later found as saying "*let the data speak for themselves*" (Gould, 1981, p.167).

The objectives of the study are two-fold --- to see whether *ARIMA* model is capable of helping a researcher predict the future value(s) of KSE 100 Index, and, if yes, to determine how many previous values of the index and that of the previous values of its error term are effective in forecasting the current, or future, value of the index. The assertions from this study will help potential investors in determining the suitable time they should go for investing their funds in the capital market of stocks.

Review of Literature

Over time, there have been some studies conducted to forecast various time series variables using the *ARIMA* modeling technique. The model has been proved to be successful for prediction of many time series if not for all. We have divided the review of literature section into two parts with the first part discussing studies that have used the *ARIMA* model for forecasting stock returns or stock market index and the second part revealing studies that have used the model for prediction of variables other than stocks.

Beginning with the first part of the review, Gay (2016) involved the *ARIMA* model to determine the association between two macroeconomic variables --- oil prices and exchange rates --- and stock returns for Brazil, Russia, India and China. It was found in their study that neither the macroeconomic variables nor the past values of stock prices were effective in forecasting stock returns for BRIC countries.

A big attempt was made by Mondal, Shit and Goswami (2014) who took 56 stocks of India from different sectors with an intention to forecast their future returns using the *ARIMA* model. Their study concluded that the model was successful in its prediction for around 85% of the cases studied by them.

An attempt was also made by Adebiyi, Adewumi and Ayo (2014) of using *ARIMA* for predicting stock returns of Zenith Bank and Nokia. They found the model to be a good predictor in the short run. Similar were the results obtained by Banerjee (2014) who also attempted to forecast Indian stock market index and found a short run prediction power of the model.

There also have been studies conducted to anticipate variables other than stock prices or stock index. For instance, Jarrett (2010) used the model for anticipating earnings of corporations and used estimated corporate earnings through conventional methods. He concluded that *ARIMA* model was no better than the factor-based models for prediction of earnings. Raymond (1997) endeavored to predict real estate prices through the Box-Jenkins methodology and was successful in observing trends in it. In the same manner, application of the model was also made by Meyler, Kenny and Quinn (1998) for predicting inflation in Ireland. The focused more on minimizing estimation errors rather than maximizing the goodness of fit.

ARIMA model was also successfully employed by Contreras et al (2003) for predicting electricity prices in Spain and California. Gilbert (2005) also used the model for multistage supply chain processes. He concluded that inventories, orders placed by customers, demands and lead times are all *ARIMA* processes and could be easily anticipated in the short run. Guha (2016) also employed the model for estimating gold prices in India and came up with a positive relationship of gold prices in the short run.

Some researchers have also used the model for forecasting crop production. Among them were Manoj and Madhu (2014) forecasted the production of

Sugarcane in some Indian regions using *ARIMA* and found that the model nicely predicted sugarcane production for as much as five years. Similarly, Hamjah (2014) estimated the production of rice in Bangladesh and he also found the model helpful in predicting the time series in the short run.

In an earlier study, Padhan (2012) assessed the productivity of 34 Indian crops. She found that the *tea* was highest predictable crop while the *papaya* was the lowest. Following her study, Jadhav, Reddy and Gaddi (2017) also attempted forecasting major Indian crops including, but not limited to, Ragi, Paddi and Maize in the Karnataka state of India. They found that the model was very accurate in its prediction of crop production overall. They used the model for predicting production of major Indian crops for 2020.

The Box-Jenkins Methodology

ARIMA modeling has long been used by researchers for time series forecasting. Researchers have been using different techniques for predicting their variables of interest using the precious *data* of that variable. In the regards, however, the most sought after technique to have ever been developed is the one known as the Box-Jenkins methodology devised by George Box and Gwilym Jenkins (Box and Jenkins, 1970). The method, of course, makes its predictions on the basis of the previous values of the variable concerned as well as the previous values of the error term. As a rule, the variable having the most number of observations available is more likely to be predicted finely than the one having a lesser number of previous observations. In this regards, Chatfield (1996) suggests at least as much as 50 observations of a variable for a decent forecast. There are some statisticians who argue the minimum number of observations should be 100 for a meaningful prediction.

The purpose of using the Box-Jenkins methodology is to be able to find the right number of previous values of a variable and its error term that are effectively relevant in determining its current or future value. The model involves three steps, namely, the *model identification*, *model estimation* and the *diagnostic checking*. In the first step, the researcher visually checks for the plots of correlation and partial correlation functions of a given variable in the variable's correlogram. The shape of the correlogram that involves spikes, since waves and decays assist the researcher in determining the influential number of lagged values of the variable and the error term. Hence, the first step helps in identifying the *right* model. The second step involves estimating the model identified by the first step. In order to confirm the model prescribed in the first step is better than any other model, a few other closer *ARIMA* configurations are also tested to ensure the superiority of the one prescribed by the Box-Jenkins method.

The third stage of the Box-Jenkins method evaluates all the models on the basis of the Akaike Information Criterion (*AIC*), Schwarz Bayesian Criterion (*SBC*), Hannan-Quin Criterion (HQC), and the adjusted R^2values. The authors of the model recommend selecting the model having the minimum information criterion values, the highest adjusted R^2 and the least number of insignificant parameters. It is often observed, however, that over-parameterized models tend to be healthier than the ones having fewer parameters. Nonetheless, the principle of parsimony has to be kept into consideration.

Research Methodology

The current study employs the time series data of a single variable, i.e., the stock market index. Hence, univariate *ARIMA* model has been used to forecast the current/future value of the index. An *ARMA* process in its general fashion as adapted from Asteriou & Hall (2007) is as follows:

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

Where,

Y_t represents the variable we are interested in trying to predict, Y_{t-1}, Y_{t-2}, Y_{t-p} are the previous or lagged values of that variable (also called the autoregressive terms), ε_t is the disturbance or error term, ε_{t-1}, ε_{t-2}, - - - , ε_{t-q} are the previousor lagged values of the error term(also known as the moving average terms), φ_1, φ_2, - - - , φ_p are the coefficients of autoregressive terms, and θ_1, θ_2, - - - , θ_p are the coefficients of the moving average regressors.

It is noteworthy that for an *ARMA* process to work, the variable must be stationary --- the one that has a constant long-run mean and a time-invariant covariance (Gujarati & Porter, 2004). This is seldom a case in time series data where the data are often highly integrated. Such was the case for our variable too, and the series was found to be integrated of order 1 meaning that it had to be differenced for once. Therefore, stock index returns were taken for the analyses which were computed by dividing the first difference of the index over its previous value for all observations. Thus the study employed *ARIMA* (*p*, *1*, *q*) model.

For analysis, monthly figures of Karachi Stock Exchange (now Pakistan Stock Exchange) were taken for 22 years from August 1997 to August 2019 which rendered 266 observations making the sample large enough to be considered for *ARIMA* analysis (Chatfield, 1996).

Results and Findings

Before subjecting the data for *ARMA* analysis, the variable KSE 100 Index was checked for any trends or non-stationary element. This was made possible through line graph, unit root test and the correlogram of the variable. The following table represents the line graph of the variable clearly showing trends in the data.

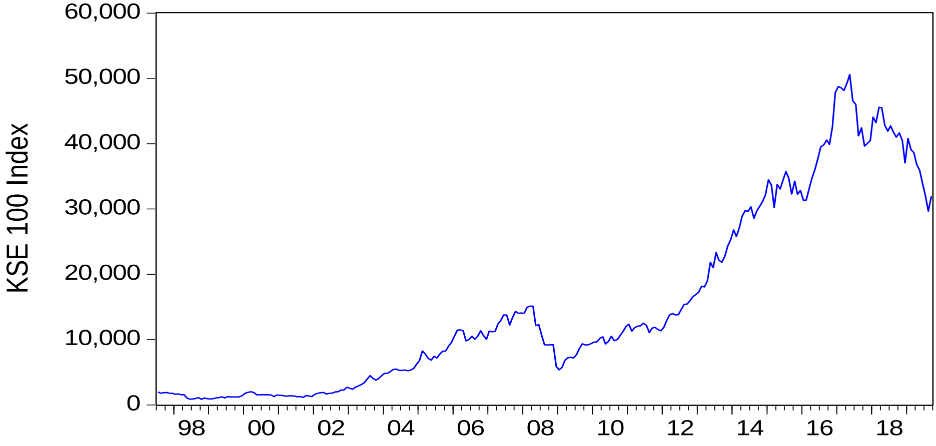

Figure 1: The Non-Stationary KSE 100 Index

The figure just presented speaks loudly of the presence of trends in the data. Of course such data is not suitable for *ARMA* calculations. In cases where a time series is integrated or trended, an integrated version of the *ARMA* process, known as the *ARIMA* process, is implemented which smoothens out any trends in the variable by taking the required number of differences. But before this is done, the (non-) stationarity of our variable is also verified through the augmented dickey fuller statistic which too advocates in favor of its trended nature. The table that follows presents the ADF test result being highly insignificant.

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

Augmented Dickey-Fuller test statistic		t-Statistic	Prob.
		-.445	.898
Test critical values:	1% level	-3.455	
	5% level	-2.872	
	10% level	-2.573	

Null Hypothesis: KSE 100 Index has a unit root
Exogenous: Constant; Lag Length: 0 (Automatic - based on SIC, maxlag=15)

Theoretically, the correlogram of a trended variable should not die down or fade away as the lag length increases. The correlogram of KSE 100 index, as shown in table 2, also shows the same pattern hinting towards the fact that the index is non-stationary.

Table 2 *Autocorrelation and Partial Autocorrelation Function of KSE 100 Index*

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|*******	.|*******	1	0.993	0.993	266.11	0.00
.|*******	.|. |	2	0.986	0.043	529.7	0.00
.|*******	*|. |	3	0.978	-0.11	789.94	0.00
.|*******	*|. |	4	0.969	-0.07	1046.4	0.00
.|*******	*|. |	5	0.959	-0.09	1298.6	0.00
.|*******	.|. |	6	0.949	0	1546.3	0.00
.|*******	*|. |	7	0.937	-0.1	1788.9	0.00
.|*******	.|. |	8	0.926	0.027	2026.6	0.00
.|*******	.|. |	9	0.914	-0.02	2259.1	0.00
.|*******	.|. |	10	0.903	0.055	2486.9	0.00
.|******|	*|. |	11	0.891	-0.07	2709.4	0.00
.|******|	.|. |	12	0.878	-0.05	2926.5	0.00
.|******|	.|. |	13	0.865	-0.01	3138	0.00
.|******|	.|. |	14	0.852	-0.03	3343.9	0.00

Following the requirements that are to be met for using an *ARMA* process, the monthly index has been transformed into monthly *returns* in order to induce stationarity in the series.

The following graph portrays the monthly returns behavior of KSE 100 index which has now become stationary. Since returns are computed by taking first differences, it can therefore be concluded that KSE 100 index is stationary at first differences.

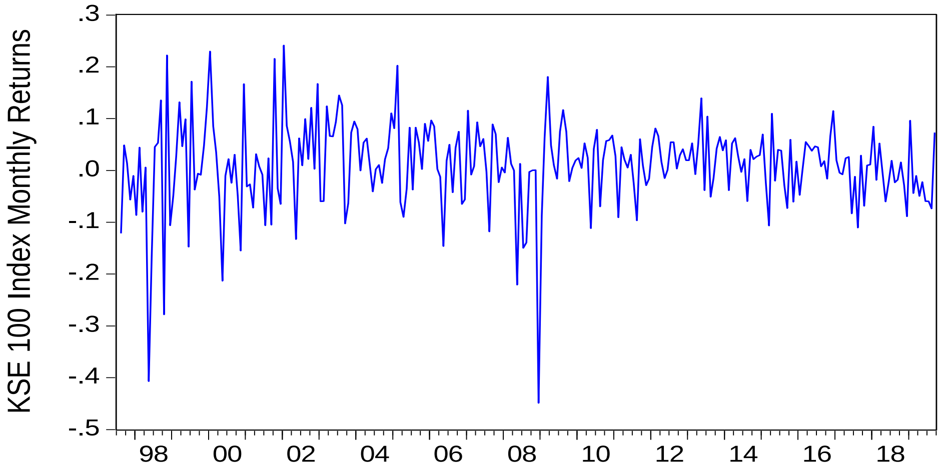

Figure 2: The Stationary KSE 100 Index Monthly Returns

To further ensure that the returns have become fully stationary and that the analysis can now be safely initiated, the augmented dickey fuller test has been run. The results indicate that the statistic is highly significant at 1% level leaving no doubt that KSE 100 index returns have no unit root (see table 3).

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

		t-Statistic	Prob.
Augmented Dickey-Fuller test statistic		-15.841	.000
Test critical values:	1% level	-3.455	
	5% level	-2.872	
	10% level	-2.573	

Null Hypothesis: KSE 100 Index has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=15)

Model Identification

After inducing stationarity in the variable, the Box-Jenkins methodology is applied. Stage 1 of the methodology is meant to identify the most suitable model. This involves finding the number of lagged values of the variable and that of the error term sufficient to explain the variable. Hence we start by making a correlogram of the variable, i.e., the stock index returns, to check for the number of positive spikes in the correlation and partial correlation columns which will help us in identifying the right number of the autoregressive and moving average terms necessary for predicting the variable.

Table 4 presents a correlogram of the monthly returns of KSE 100 Index covering the period from August 1997 to August 2019.

Table 4: *Autocorrelation and Partial Autocorrelation Functions of KSE 100 Index Returns*

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|. |	.|. |	1	.027	.027	0.197	0.66
.|. |	.|. |	2	.037	.036	0.563	0.76
.|. |	.|. |	3	-.049	-.051	1.222	0.75
.|. |	.|. |	4	.038	.040	1.620	0.81
.|* |	.|* |	5	.097	.099	4.164	0.53
.|. |	.|. |	6	.065	.055	5.306	0.51
.|. |	.|. |	7	-.032	-.040	5.595	0.59
.|. |	.|* |	8	.073	.080	7.050	0.53
.|. |	.|. |	9	.045	.045	7.614	0.57
.|. |	.|. |	10	-.006	-.033	7.625	0.67
.|. |	.|. |	11	-.001	-.005	7.625	0.75
.|. |	.|. |	12	-.020	-.015	7.740	0.81
.|. |	.|. |	13	-.007	-.022	7.752	0.86
.|. |	.|. |	14	-.013	-.031	7.803	0.90

At first glance, the aforementioned table seems to give no clue of how many autoregressive and moving average terms to retain. Up until the first four terms of both autocorrelation and partial correlation columns, no significant spikes can be seen on either direction. However, on the fifth term, there are small but positive and visible spikes on both columns. This signals towards a strange and apparently over-parameterized *ARIMA* (*5*, *d*, *5*) model. In the next stage, however, we will also check for models that are simpler (having lesser parameters) than the one formulated through the Box-Jenkins approach in the hope that we may explore a model that is more parsimonious.

Model Estimation

In this stage, we will again be trying to *identify* the most appropriate *ARIMA* configuration by estimating several probable models along with the one prescribed by the Box-Jenkins methodology. We start with *ARIMA* (*5*, *d*, *5*).

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

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.014	0.006	2.39	0.02
AR(1)	0.242	0.079	3.064	0.00
AR(2)	0.14	0.055	2.533	0.01
AR(3)	-0.27	0.042	-6.461	0.00
AR(4)	-0.248	0.057	-4.34	0.00

AR(5)	0.77	0.069	11.232	0.00
MA(1)	-0.195	0.084	-2.316	0.02
MA(2)	-0.138	0.058	-2.361	0.02
MA(3)	0.253	0.042	6.087	0.00
MA(4)	0.37	0.058	6.352	0.00
MA(5)	-0.868	0.082	-10.533	0.00

Dependent Variable: KSE 100 Index Monthly Returns
Method: Least Squares; Included observations: 261 after adjustments

The results given in table 5 of *ARIMA* (*5*, *d*, *5*) look impressive as all of the 10 coefficients are highly significant. The adjusted R^2 is 13.7% and the *SBC* is -2.06. We, however, cannot be sufficiently certain about whether there exists a model any better than *ARIMA* (*5*, *d*, *5*) unless we check for these other possibilities.

Since *ARIMA* (*5*, *d*, *5*) seems to be a bit over-parameterized, we will try making use of models that are more parsimonious. Let's attempt using *ARIMA* (*3*, *d*, *3*).

Table 6: *Regression Results Using ARIMA (3, d, 3) Model*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.014	0.006	2.471	0.014
AR(1)	0.728	0.084	8.692	0.00
AR(2)	-0.753	0.03	-24.716	0.00
AR(3)	0.819	0.077	10.603	0.00
MA(1)	-0.684	0.088	-7.814	0.00
MA(2)	0.792	0.02	39.494	0.00
MA(3)	-0.886	0.086	-10.251	0.00
R-squared	0.094	Mean dependent var		0.011
Adjusted R-squared	0.073	S.D. dependent var		0.084
S.E. of regression	0.081	Akaike info criterion		-2.159
Sum squared resid	1.686	Schwarz criterion		-2.063
Log likelihood	290.846	Hannan-Quinn criter.		-2.12
F-statistic	4.428	Durbin-Watson stat		2.05
Prob (F-statistic)	0			

Dependent Variable: KSE 100 Index Monthly Returns
Method: Least Squares
Included observations: 263 after adjustments

The model *ARIMA* (*3*, *d*, *3*), presented in table 6, also has all parameters significant. However, at 7.3%, it has a lower adjusted R^2 than that for *ARIMA* (*5*, *d*, *5*). The other information criteria values are closely comparable for the two models. Nonetheless, a visible difference in the adjusted R^2 values for the two models suggests an upper edge for *ARIMA* (*5*, *d*, *5*).

In time series forecasting, many authors have found that models with fewer parameters tend to forecast better. The models *ARIMA* (*1*, *d*, *1*) and *ARIMA* (*1*, *d*, *0*) are of course one of the simplest ones used very frequently by the academic community. We need to check whether these models can come up with better *solution* to our problem than the previous models checked by us so far. So we now examine *ARIMA* (*1*, *d*, *1*).

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

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011	0.005	2.061	0.040
AR(1)	-0.116	0.593	-0.196	0.845
MA(1)	0.138	0.594	0.233	0.816
R-squared	0.001	Mean dependent var		0.011
Adjusted R-squared	-0.007	S.D. dependent var		0.084
S.E. of regression	0.084	Akaike info criterion		-2.098
Sum squared resid	1.861	Schwarz criterion		-2.058
Log likelihood	281.013	Hannan-Quinn criter.		-2.082
F-statistic	0.134	Durbin-Watson stat		1.984
Prob (F-statistic)	0.874			

Dependent Variable: KSE 100 Index Monthly Returns
Method: Least Squares; Included observations: 263 after adjustments

Surprisingly, the model *ARIMA* (*1*, *d*, *1*) presented in table 7 has very a week position with all its parameters being insignificant. Also, the model has a very low, in fact an unrealistic, adjusted R^2 value which is negative meaning that this model is not an option at all for our case.

Perhaps we have gone a bit too simple in our estimation this time, whereas our variable, on the other hand, demands a bit more parameters to be forecasted well. So this time, we increase slightly the number of parameters to find that lowest point at which our results are legitimately acceptable.

Table 8: *Regression Results Using ARIMA (2, d, 1) Model*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011	0.005	1.976	0.049
AR(1)	-0.754	0.332	-2.271	0.024
AR(2)	0.061	0.063	0.958	0.339
MA(1)	0.786	0.330	2.383	0.018
R-squared	0.007	Mean dependent var		0.011
Adjusted R-squared	-0.004	S.D. dependent var		0.084
S.E. of regression	0.084	Akaike info criterion		-2.094
Sum squared resid	1.848	Schwarz criterion		-2.039

Log likelihood	280.352	Hannan-Quinn criter.	-2.072
F-statistic	0.615	Durbin-Watson stat	1.992
Prob (F-statistic)	0.606		

Dependent Variable: KSE 100 Index Monthly Returns
Method: Least Squares; Included observations: 264 after adjustments

The aforementioned table presents *ARIMA* (*2*, *d*, *1*).Again the model is a complete failure in many respects with its most important deficiency being the unrealistic value of the adjusted R^2. There is also one insignificant parameter in the model.

In search of a simpler model than the Box-Jenkins' identified *ARIMA* (*5*, *d*, *5*), let us also attempt to estimate *ARIMA* (*2*, *d*, *2*).

Table 9: *Regression Results using ARIMA (2, d, 2) Model*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011	0.005	2.017	0.045
AR(1)	0.008	0.046	0.163	0.871
AR(2)	-0.914	0.041	-22.159	0.000
MA(1)	0.045	0.036	1.246	0.214
MA(2)	0.945	0.033	28.483	0.000
R-squared	0.043	Mean dependent var		0.011
Adjusted R-squared	0.028	S.D. dependent var		0.084
S.E. of regression	0.083	Akaike info criterion		-2.123
Sum squared resid	1.781	Schwarz criterion		-2.055
Log likelihood	285.256	Hannan-Quinn criter.		-2.096
F-statistic	2.928	Durbin-Watson stat		2.041
Prob (F-statistic)	0.021			

Dependent Variable: KSE 100 Index Monthly Returns
Method: Least Squares; Included observations: 264 after adjustments

Table 9 presents *ARIMA* (*2*, *d*, *2*). Both the autoregressive and the moving average terms are insignificant for the first order and significant for the second order. Overall there are two insignificant terms in the model. The adjusted R^2 is merely 2.8% --- much lower than what it was for *ARIMA* (*5*, *d*, *5*) and *ARIMA* (*3*, *d*, *3*). In the next stage of the Box-Jenkins methodology, however, we will extend our search of finding the most parsimonious, yet the most practical, model by checking for the performance of a few more models along with their comparison.

Diagnostic Checking

Discussion in the previous section alluded that *ARIMA* (*5*, *d*, *5*) is the probably the most suitable model for our variable. We, however, will estimate a few more models in the diagnostic checking stage to ensure all the possible models that could forecast the stock returns are checked out. The following table compares, therefore, 10 separate *ARIMA* configurations on the basis of their adjusted R^2, *AIC*, *SBC*, *HQC* values and the number of insignificant parameters.

Table 10: *Comparing ARIMA models: The rows with bold figures indicate the most apt models*

ARIMA Model	Adjusted R^2	AIC	SBC	HQC	Insignificant lags
ARIMA (1, d, 0)	-0.003	-2.105	-2.078	-2.095	One
ARIMA (1, d, 1)	-0.007	-2.098	-2.058	-2.082	Two
ARIMA (2, d, 1)	-0.004	-2.094	-2.039	-2.072	One
ARIMA (3, d, 1)	-0.005	-2.085	-2.018	-2.058	Two
ARIMA (1, d, 2)	-0.007	-2.094	-2.04	-2.073	Three
ARIMA (1, d, 3)	-0.008	-2.089	-2.021	-2.062	Four
ARIMA (2, d, 2)	0.028	-2.123	-2.055	-2.096	Two
ARIMA (3, d, 3)	0.073	-2.159	-2.063	-2.12	None
ARIMA (4, d, 4)	0.072	-2.149	-2.027	-2.1	Two
ARIMA (5, d, 5)	0.137	-2.21	-2.06	-2.15	None

Of the 10 models presented in table 10, *ARIMA* (*5*, *d*, *5*) undoubtedly takes the lead in many respects. For one thing, *ARIMA* (*5*, *d*, *5*) has the highest value of adjusted R^2. The model also minimizes all the information criterion values and has no insignificant parameters.

In comparison with *ARIMA* (*5*, *d*, *5*), a rather simpler *ARIMA* (*3*, *d*, *3*) also has a better performance than the rest of the possibilities except for *ARIMA* (*5*, *d*, *5*). So *ARIMA* (*3*, *d*, *3*) stands second in ranking after *ARIMA* (*5*, *d*, *5*) in terms of its forecast ability. All the other models in the table are a complete no-choice owing to their very poor performance in forecasting our variable of interest. Hence it has been established that the Box-Jenkins' prescribed *ARIMA* (*5*, *d*, *5*) works best for forecasting monthly returns of KSE 100 Index.

Discussion

The third stage of the Box-Jenkins methodology has shown *ARIMA* (*5*, *d*, *5*) to be the most appropriate model for predicting stock market index. Although many previous studies have suggested much simpler models (with lesser parameters) for forecasting their variables of interest, the fact that *ARIMA* (*5*, *d*, *5*), in our case, has the highest adjusted R^2 value and the least *AIC*, *SBC* and *HQC* values makes it irresistible to be deemed as the best model.

Looking at the previous literature with regard to time series forecasting, it is evident that many attempts have been made making use of the *ARIMA* technique with a good success ratio. It should, however, also be noted that most, if not all, of the previous studies have concluded by suggesting models for their time series that were as simple or parsimonious as *ARIMA* (*1*, *d*, *0*) or *ARIMA* (*1*, *d*, *1*). The current study, on the other hand, proposes a rather unusually over-parameterized model for prediction of its variable of interest. However, this should not be a matter of some serious concern as some time series need a longer run to be predicted better.

Speaking in a broader context, however, results of the study are in line with previous works which have also successfully attempted to forecast stock returns through *ARIMA* modeling. For instance, Mondal, Shit and Goswami (2014) used *ARIMA* model to predict stock prices of as many as 56 Indian companies and found that 85% of the firms they selected in their study were forecasted precisely. In the same manner, Adebiyi, Adewumi and Ayo (2014) also attempted to predict stock prices through Box-Jenkins method and found that the method was superior to the conventional methods of forecasting.

There also have been few studies conducted to estimate variables other than stock returns using *ARIMA*. Manoj and Madhu (2014), for example, employed the model for anticipating Sugarcane production in India. Their findings revealed the model quite helpful. *ARIMA* (*2*, *d*, *1*) was the most suitable configuration for their study. Similarly, Hamjah (2014) also made use of the model for predicting rice production in Bangladesh and concluded that the model had a decent short-term prediction ability.

To summarize, the current study is in line with previous literature in that *ARIMA* model is very efficient in predicting various time series in the short run. However, the current study has ended up with the selection of a very over-parameterized model that engages five previous values of the variable along with five lagged values of the error term. This makes the current study somewhat atypical in its results in the sense that no previous study has so far conceived an *ARIMA* model with as many parameters as the one this study has suggested.

Conclusion

Stock markets all over the world are considered to be indicators of the economy's financial health. They indicate how much of an investment opportunity is there in a given region. When the stock market index grows, it develops people's confidence over the market and they throw more money for investment. Conversely, an ever-decreasing index agitates the investment community making them reluctant to bet their hard-earned money in the stock market. But the stock market more often experiences a haphazard behavior and there is no established trend, either upward or downward, in the short run. Although most stock market indices do grow in the long run, predicting the index in the short run is what makes the matter somewhat complicated.

If investors somehow find the way to efficiently forecasting the stock market index, they would gain much confidence required to invest in the risky venture. This paper has attempted to help investors forecast stock market index returns in the short run. The study used the *ARIMA* technique executed through the Box-Jenkins methodology in which predictions about the future value of a given variable are made on the basis of the past values of that variable as well as on the past values of the error term. The study took monthly data of stock market index for 22 years and found that *ARIMA* model was reasonably effective in speculating the returns expected from Karachi Stock Exchange 100 Index. It is positively expected that the current work will guide prospective short-term stock market investors in determining when, and when not, to devote their excess reserves in their stock portfolios.

References

Adebiyi, A., Adewumi, A., & Ayo, C. (2014). *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*. 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.

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 next day 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* (4th Ed.). 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.

Jarrett, J. E. (2010). Efficient markets hypothesis and daily variation in small Pacific-basin stock markets. *Management Research Review, 33*(12), 1128-1139.

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 of Finance, 8*(2), 152-163.