

Tracing Stock Returns on Quarterly Basis: The Case of KSE-100 Index

Mustafa Afeef* Nazim Ali† Adnan Khan‡

Abstract

The stock market index can be forecasted in two ways --- either through taking those external factors that influence movements in the index or by basing one's predictions on the previous values of the index. The current study has used the method described later by employing the Box-Jenkins methodology --- a method famously used by most researchers while conducting ARIMA modeling--- by taking past figures of KSE 100 Index. Quarterly figures of the Index were, therefore, taken for 22 years from August 1995 to October 2017 that translated into 90 observations. Results revealed that the forecasting model used in the study did well in anticipating returns in the short-run. The findings of the study can be consumed by investors, particularly short-term, in deciding when, and when not, to risk their hard-earned funds at Pakistan Stock Exchange.

Key Words:

ARIMA, Box-Jenkins Methodology, PSX 100 Index, Prediction, Stationarity

Introduction

The capability to forecast the future can never be underrated when one talks about investments. Since, of course, the future is always not certain, investors often seem to be pondering about discovering the suitable time to invest their funds. Prediction of the stock market index is no different from forecasting other kinds of investment as many factors play their role in the returns. A stock market index movement tells us in which direction a given economy is heading towards. Owing to these reasons, investors keep looking at the index to observe what is going on in the stock market.

A time-series variable can be forecasted in two ways. One is to forestall the direction the variable is expected to go in keeping with all the factors that have a

* 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.

bearing on it. The other method is to predict its future values based on its lagged observations. In today's era, researchers are increasing their reliance on using the method mentioned later for forecasting their variables and this study has also employed the method of relying on the lagged values of our variable of interest. The current study has also employed the same methodology known as the Autoregressive Integrated Moving Average (hereinafter referred to as the *ARIMA*) technique of forecasting time series.

ARIMA technique has been long in practice but the method was more popularized by Box and Jenkins (1970) after they discovered a method for effectively using *ARIMA*. The model is now often employed using what is commonly known as the Box-Jenkins methodology.

The study serves two objectives, i.e., to investigate whether stock returns could be adequately forecasted every quarter using the *ARIMA* modeling, and secondly, to explore how many previous quarters of data will be engaged in efficiently forecasting the current (or future) value of the index or the returns. It is hoped that by using the results of the current study, short term investors may get some insights into how the stock market behaves in the short run.

Review of Literature

Forecasting of time series variables has always remained challenging. Researchers over time have put their efforts to efficiently forecast variables of their interest. *ARIMA* models also have a history of being employed by many academicians and professional investors. A discussion of some of the previous work follows:

Meyler, Kenny, and Quinn (1998) employed the *ARIMA* technique for anticipating inflation in Ireland and found that the model had a decent prediction capability. Jarrett (1990) used *ARIMA* for corporate earnings estimation but concluded that the model gave no better results than the conventional models. Another attempt was made by Raymond (1997) who predicted prices of real estate using *ARIMA*. He found the model helpful in envisaging the direction the prices of the real estate were going. The model was also used by Contreras et al (2003) for anticipating electricity prices in the country of Spain and the city of California who also found a good short-run prediction power of the model. Gilbert (2005), on the other hand, used the *ARIMA* model for processes related to the supply chain. He found that all those supply chain-related variables including inventories, demands, orders placed, and lead times could be easily forecasted using the model. Among the users of *ARIMA* was Guha (2016) too who endeavored to estimate the prices of gold in India and who found that the prices could be easily anticipated in the short run.

The production of different crops has also been forecasted by researchers using the *ARIMA* model. For instance, Padhan (2012) checked the productivity of some 34 crops in India and found that the *tea* crop was the most predictable whereas the

papaya crop was not predictable at all. Similarly, the production of Sugarcane was predicted by Manoj and Madhu (2014) in some parts of India employing the *ARIMA* model and discovered that a good forecast was made by the model for around five years. Another attempt was made by Hamjah (2014) who predicted the production of rice, using *ARIMA*, in Bangladesh and found the model adequately successful. The production of the major crops of an Indian state named Karnataka was estimated by Jadhav, Reddy, and Gaddi (2017) in a paper in which they were able to adequately predict the production of these crops for the next three years.

There have been a few studies conducted to anticipate stock prices or stock market index engaging the *ARIMA* model. For instance, Mondal, Shit, and Goswami (2014) did a big attempt by including shares of as many as 56 companies of India for their future price anticipation using the *ARIMA* model. They found that for around 85% of shares included in their study the prediction carried out through the Box-Jenkins method was very accurate. Similar studies were conducted by Adebisi, Adewumi, and Ayo (2014) and Banerjee (2014) who estimated stock returns using *ARIMA* and found the model to be decently capable. We now give a brief description of how the Box-Jenkins methodology works in practice.

The Box-Jenkins Method

The Box-Jenkins method offers a way of using *ARIMA* modeling for time series variables. Developed by Box and Jenkins (1970), the method helps us in identifying the number of the previous values of our variable of interest as well as the number of lagged values of the error term that our variable depends upon. The method works better when there is a larger number of observations for a given time series. However, 50 observations are considered to be the minimum acceptable number for a given variable below which the model is not likely to give meaningful results (Chatfield, 1996).

The Box-Jenkins methodology consists of three steps. In the first step known as the model identification step, the researcher inspects the autocorrelation and partial correlation functions to explore the number of lagged values of the variable and that of the error term that significantly affects the variable. The second step known as model estimation involves estimating the model *identified* in the first step. A few other models that could give better results are also estimated to have something for comparison. The third and final step includes the diagnostic testing in which the models estimated are compared based on the information criterion values, adjusted R^2 value, and the number of insignificant parameters. The Box-Jenkins method requires the selection of the model that has the lowest information criterion values, the highest adjusted R^2 value, and the least number of insignificant parameters.

Research Methodology

The study at hand uses the time series data of only one variable ---the stock market index. Therefore, the univariate *ARIMA* technique has been employed to predict the variable's future values. Needless to mention that a stationary time series variable (the one which is not integrated) needs an *ARMA* process. In other words, it does not need to be made stationary since it is already in a stationary position. In its standard form, an *ARMA* process as taken from Asteriou& Hall (2007) is as follows:

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

In the equation above, Y_t depicts the explained variable to be predicted, Y_{t-1} to Y_{t-p} are the lagged or autoregressive terms of Y_t , ε_t is the error term, ε_{t-1} to ε_{t-q} are the lagged or moving average terms, φ_1 to φ_p are the autoregressive coefficients, and θ_1 to θ_p are the moving average coefficients.

In most cases, time series variables happen to be non-stationary in which case we need to differentiate them for enough time to make them stationary. A variable achieves stationarity only when its long-run mean becomes constant and its covariance becomes time-invariant (Gujarati & Porter, 2004). Since in our case the dependent variable, i.e., the stock index was also non-stationary, we took quarterly returns by taking its first difference and then dividing it over the lagged value of the variable. Therefore, an *ARIMA* model, i.e., the one that allows for the variable to be integrated, instead of an *ARMA* model was employed in the study.

For analysis, quarterly figures of Pakistan Stock Exchange were engaged for 22 years from August 1995 to October 2017 which produced 90 observations making the sample large enough to be considered for *ARIMA* analysis (Chatfield, 1996).

Results and Findings

The graphical analysis, as well as the unit root test of the variable stock market index, showed obvious trends in the data. Therefore, quarterly returns were computed, and then the data was checked for stationarity again. As can be seen in figure 1, the returns of the KSE 100 index were stationary.

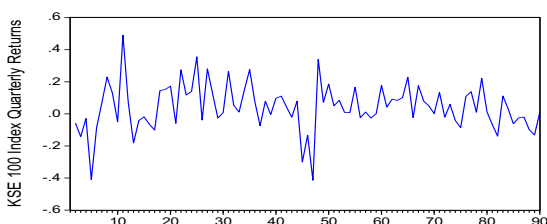


Figure 1: The Stationary KSE 100 Index Quarterly Returns

The unit root test of KSE 100 index returns also gave a t-statistic of -8.367 which was highly significant (p -value = .000) and which left no doubt that the returns were made stationary (see table 1).

Table 1. ADF Test for KSE 100 Index Quarterly Returns

Null Hypothesis: KSE 100 Index Quarterly Returns Has A Unit Root				
			t-Statistic	p-value
Augmented Dickey-Fuller test statistic			-8.367	.000
Test critical values:	1% level		-3.506	
	5% level		-2.895	
	10% level		-2.585	

Model Identification

After stationarity has been induced in the variable, we proceed by employing the Box-Jenkins methodology. In the first stage, we try to identify the most appropriate model by looking at the correlogram to check for the number of autoregressive and moving average terms. The correlogram of quarterly returns of the KSE 100 Index is given in table 2.

Table 2. Correlogram of KSE 100 Index Quarterly Returns

Autocorr.	Partial Corr.		AC	PAC	Q-Stat	Prob
. *	. *	1	.105	.105	1.014	.314
. *	. *	2	.129	.119	2.567	.277
.	.	3	.030	.005	2.650	.449
.	.	4	.037	.019	2.782	.595
. *	. *	5	-.126	-.139	4.321	.504
. *	. *	6	-.172	-.162	7.222	.301
.	. *	7	.073	.141	7.742	.356
.	. *	8	.049	.085	7.978	.436
.	.	9	-.002	-.024	7.978	.536
.	.	10	-.010	-.039	7.988	.630
.	.	11	.073	.023	8.544	.664
.	.	12	.032	.026	8.654	.732
.	. *	13	.058	.106	9.012	.772

Looking at the correlogram in the aforementioned table, it is visible that the *decay* is beginning at lag 2 for both the autocorrelation and the partial autocorrelation functions. This hints towards an *ARIMA* (2, *d*, 2) model. We will, nonetheless, also check a few other models to see whether *ARIMA* (2, *d*, 2) is the best solution to our case or could there be another possible alternative to it.

Model Estimation

The model identification stage of the Box-Jenkins methodology has prescribed *ARIMA* (2, *d*, 2) to be employed. So we start by estimating *ARIMA* (2, *d*, 2). We will later compare its results with other possible configurations of *ARIMA*.

Table 3. OLS Estimation using *ARIMA* (2, *d*, 2), Model

Explained Variable: KSE 100 Index Quarterly Returns				
Estimation Method: Ordinary Least Squares				
Observations included: 88 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	p-Value
C	.052	.019	2.70	.008
AR(1)	.023	.426	.053	.958
AR(2)	.480	.387	1.239	.219
MA(1)	.017	.445	.039	.969
MA(2)	-.426	.398	-1.069	.288
R-squared	.057	Akaike info. criterion		-.984
Adj. R-squared	.011	Schwarz Bay. criterion		-.842
		Hannan-Quin criterion		-.927

The results of *ARIMA* (2, *d*, 2) are presented in table 3. Surprisingly, this model has all its coefficients highly insignificant. The model also has a worryingly low adjusted R^2 value. Prescribed by the Box-Jenkins methodology, this model couldn't at all be the best possible solution for our problem, however. The only way to find a better model is to try other possibilities one by one and compare them based on their adjusted R^2 , information criterion values, and the number of insignificant parameters that they have. We now try the much common *ARIMA* (1, *d*, 1) model to see whether it outperforms the one suggested by the Box-Jenkins method.

Table 4. OLS Estimation using *ARIMA* (1, *d*, 1), Model

Explained Variable: KSE 100 Index Quarterly Returns				
Estimation Method: Ordinary Least Squares				
Observations included: 88 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	p-Value
C	.052	.019	2.784	.006
AR(1)	.735	.227	3.236	.002
MA(1)	-.693	.252	-2.748	.007
R-squared	.055	Akaike info. criterion		-1.02
Adj. R-squared	.033	Schwarz Bay. criterion		-.936
		Hannan-Quin criterion		-.986

Table 4 returns the output of *ARIMA (1, d, 1)*. Astonishingly, this model is better in every respect than the Box-Jenkins prescribed *ARIMA (2, d, 2)*. For one thing, all of the coefficients of *ARIMA (1, d, 1)* are significant in contrast with *ARIMA (2, d, 2)* in which none of the coefficients were statistically significant. Also, all of the information criteria for *ARIMA (1, d, 1)* have much lower values than those for *ARIMA (2, d, 2)*. Finally, the adjusted R^2 value is also larger for *ARIMA (1, d, 1)* than the model proposed by the Box-Jenkins method. This inevitably makes *ARIMA (1, d, 1)* much superior to *ARIMA (2, d, 2)*. However, before we give our final word about the best possible *ARIMA* configuration for our dependent variable, i.e., the quarterly returns of the KSE 100 index, we need to check all other possible models as well.

The next and the final stage of the Box-Jenkins methodology involves a comparison of the various alternatives to seek the most practical and parsimonious solution to our problem.

Diagnostic Checking

Analysis in the former segment indicated *ARIMA (2, d, 2)* to be the appropriate model for our variable. It was, however, shown that *ARIMA (1, d, 1)* is better than *ARIMA (2, d, 2)* in all respects. We will want to know whether we can have an even better solution to our problem. We will, therefore, estimate other models as well for this purpose. The following table presents the comparative results of a few other models.

Table 5. Contrasting *ARIMA* models

<i>ARIMA</i> model	Adjusted R^2	AIC	SBC	HQC	Insignificant lags
<i>ARIMA (1, d, 1)</i>	.033	-1.020	-.936	-.986	None
<i>ARIMA (2, d, 1)</i>	.012	-.996	-.882	-.950	One
<i>ARIMA (1, d, 2)</i>	.021	-.998	-.885	-.952	One
<i>ARIMA (1, d, 3)</i>	.016	-.982	-.841	-.924	Two
<i>ARIMA (2, d, 2)</i>	.011	-.984	-.842	-.927	All
<i>ARIMA (3, d, 3)</i>	.077	-1.022	-.822	-.942	Two
<i>ARIMA (3, d, 4)</i>	.104	-1.042	-.813	-.949	Three
<i>ARIMA (4, d, 4)</i>	.083	-1.119	-.861	-1.015	Four

A comparison of the different models is given in table 5. In terms of the number of insignificant parameters, one can consider *ARIMA (2, d, 2)* to be the worst choice having all lags insignificant. As per the adjusted R^2 , however, *ARIMA (3, d, 4)* takes the lead with a value of 10.4%. The model also has the minimum Akaike Information Criterion (AIC) value of -1.042. The Hannan-Quin Criterion (HQC) value is lowest for *ARIMA (4, d, 4)* and the Schwarz-Bayesian Criterion (SBC)

value, which plays the most important role in selecting the correct model, is the lowest for *ARIMA* (1, *d*, 1).

If adjusted R^2 is to be considered the decisive factor for model selection, *ARIMA* (3, *d*, 4) is to take the lead followed by *ARIMA* (4, *d*, 4). However, the former has three insignificant parameters whereas the latter has four. If, however, the three information criteria are given priority in the choosing the right model, *ARIMA* (1, *d*, 1) gets the edge as the model has the lowest *SBC*, the second-lowest *HQC* value after the over-parameterized *ARIMA* (4, *d*, 4) and the fourth-lowest *AIC* value. The important point, nonetheless, is that all those models having lower *HQC* and *AIC* values than *ARIMA* (1, *d*, 1) have insignificant parameters in the first instance, and also have much more lags involved violating the very principle of parsimony advocated by the Box-Jenkins methodology. It is therefore held that *ARIMA* (1, *d*, 1) is the most appropriate model for our variable for the following reasons; a) the model has the lowest *SBC* than any other model, b) the model has *all* its parameters highly significant, c) the model has the second-lowest *HQC* value, and d) the model is substantially parsimonious than the ones offering larger adjusted R^2 and lower *AIC* or *HQC* values.

Discussion

We had found in the analysis section that *ARIMA* (1, *d*, 1) is the most fitting model for our variable. This very *ARIMA* configuration has remained very popular in the literature with many studies in the past finding the same model appropriate for their variables. There have been studies that found bigger models (the ones with more lags) to be more helpful in forecasting their variables. In general, however, *ARIMA* models have been considered to be reasonably successful in predicting a given variable's future values. For instance, some of the studies meant for predicting stock prices using *ARIMA* modeling are the ones conducted by Mondal, Shit, and Goswami (2014), and Adebisi, Adewumi, and Ayo (2014).

A few researchers have also used the *ARIMA* technique for anticipating crop production. These studies include the ones conducted by Padhan (2012), Manoj and Madhu (2014), Hamjah (2014), and Jadhav, Reddy, and Gaddi (2017). In nutshell, many attempts have been made by researchers using *ARIMA* for forecasting their variables and our study also, like previous studies, finds this method of prediction very helpful.

Conclusion

Around the world, security markets particularly stock markets are deemed to be the emblems of a given economy's financial prosperity. They indicate investment prospects available in an area. A stock market that portrays a constant or a continual bullish behavior gains investors' confidence, in turn, toss their

investments without much hesitation. However, in a struggling or an uncertain market, investors take steps very carefully concerning throwing their money. Of course, they need to be able to predict what is to happen in the next few weeks and/or months. *ARIMA* modeling offers a way for investors to make their predictions efficient in the short run. We used the model for forecasting KSE 100 index quarterly and it was found that the index could be very effectively anticipated based on a one-quarter previous value of the index and one-quarter previous value of the error term. Investors can use the findings of this study to track future changes in the stock market so that they can catch the most appropriate time to invest.

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