American Journal of Operational Research 2020, 10(2): 23-29
DOI: 10.5923/j.ajor.20201002.01

Time Series Analysis of KSE-100 Using Box- Jenkin's Methodology

M. Haseeb Khan, Nadia Mushtaq*

Department of Statistics, Forman Christian College University, Lahore, Pakistan

Abstract The stock market is one of the biggest sectors for generating capital for both investor and company as well as for the government and it plays a vital role for the economic progress of any country. The most reliable way to calculate the stock market performance is to check its indices. We are taking KSE-100 Index of Pakistan Stock Exchange (PSX) as it is one of the best performing Index with companies having large capitalization to analyze and forecast its values based on the data of 250 days from 1st May 2017 – 30th April 2018 to the next 90 days by using Box-Jenkins methodology. We also compare the forecast values with actual values, and we can observe that the fits behave more like the actuals.

Keywords Box-Jenkin's Methodology, Forecast, KSE-100 Index

1. Introduction

The stock market is a platform which deals with the trading of securities through various physical and electronic means. It allows investors to buy shares of ownership in a company which means he is entitled in profit and loss of a company and because of this it runs on the principle of market economy. The main purpose for which the investors buy shares of the companies is to sell them at a higher price to gain profit on their investment but not everyone succeeds in gaining return on their investment because of the constant fluctuation in the prices of stocks. To keep themselves on safer side, the investors analyze the market on the basis of its indices. A stock index is said to be measurement of the section of stock market which measures the change of stock prices of its components. In this paper, we are going to talk about the KSE-100 Index of PSX. The index was launched in Nov 1991. It is taken as an Index having companies with large capitalization. It consists of 100 companies and is the most generally accepted measure with about 90% market capitalization of the whole PSX.

Everyone wants to predict the future in order to secure his investments. However, the future is always uncertain but the investor always wants to invest the spare money in order to obtain lucrative results. The fluctuation in the stock market is due to many reasons, some which are really hard to predict but whatever it maybe but it indicates the direction of the economy. Higher the stock market capitalization, higher the company's valuation and in return higher the country's economy.

There are mainly two methods for the time series forecast. One method is to take explanatory variables in regression analysis by taking factors that are responsible to fluctuate the stock market. It has a drawback. Many factors are not easy to understand or to measure and in some cases the required data is not available. However, there is also another method in which we predict the future value based on the past ones. One such model which we are using is and Autoregressive Integrated Moving Average (ARIMA) model in case of non-stationarity. Non-stationarity is defined by Gujarati & Porter (2004) as the one that hasn't a constant long-run mean and a time-invariant covariance. In ARIMA model, time series is forecasted on the basis of its previous values and previous residuals value.

We are following the Box-Jenkin's methodology to do this research. Box and Jenkins published this methodology in their renowned book "Time Series Analysis: Forecasting and Control". Box & Jenkins (1970) helps us to identify importance of lagged value and residuals of a variable in predicting its future value.

The objective of our study is to forecast the values of KSE-100 based on the data from 1st May 2017 – 30th April 2018 for the next 90 days and also compare the forecast values with actual values so that we can observe that either the fits behave more like the actuals.

2. Literature Review

Forecasting is an interesting area of research for the researchers and it will continue to fascinate to better the current predictive models. The reason is that everyone can make an investment in the light of decisions and ability to plan and develop effective strategy about their daily and

* Corresponding author:
nadiamushtaq@fccollege.edu.pk (Nadia Mushtaq)
Received: Jul. 21, 2020; Accepted: Aug. 27, 2020; Published: Sep. 15, 2020
Published online at http://journal.sapub.org/ajor

future financial desires. Pai, P. et al. (2005) said that stock price forecasting is one of most troublesome assignment to accomplish in financial forecasting due to complex nature of stock market.

Stock market predictors always try to develop successful techniques for forecasting index value of prices. The fundamental point is to acquire high profit utilizing best trading strategies and a successful stock market prediction to achieve best result and also to minimize inaccurate forecast stock price. The objective of investors is to make any forecasting strategy that could help them in easy benefitting and limiting investment risk from the stock market. Wang, J.J. et al. (2012) said that the prediction should be possible from two perspectives: statistical and artificial intelligence techniques.

Chi. S.C. et al. (1999) told us that there are two analytical models to anticipate the stock market, namely, fundamental analysis and technical analysis. Fundamental analysis deals with economic analysis, industry analysis and company analysis. Technical analysis ensures us with predicting future price based on the past market data. Time series model is a dynamic research field which has pulled in considerations of analysts group over last couple of decades.

Raicharoen, T. et al. (2003) said that the fundamental point of time arrangement demonstrating is to precisely gather and thoroughly ponder the past perceptions of a period arrangement to build up a proper model which depicts the inalienable structure of the arrangement. This model is then used to create future esteems for the arrangement, i.e. to make estimates.

Zhang, G.P. (2007) described that time series models determine along these lines can be named as the demonstration of foreseeing the future by comprehension the past. Because of the crucial significance of time series models is to determine in various fields, for example, business, financial aspects, economics, science and engineering, etc., high care ought to be taken to fit a satisfactory time series model to the required arrangement. Tong, H. (1983) describes that an effective time series analysis relies upon a proper model fitting. A huge amount of work has been done by analysts over various years for the advancement of capable models to enhance the forecasting accuracy. Accordingly, different important time series forecasting models have been made. Raicharoen, T. et al. (2003) elaborated that time series forecasting thus can be called as the demonstration of foreseeing the future by understanding the past.

Bagnall, A.J. et al. (2004) emphasized that forecasting of stock prices has attracted attention from the research community. Time series analysis covers an extensive number of forecasting methods. Researchers have built up various modifications to the basic ARIMA model and found impressive achievements in these methods. The modifications incorporate clustering time series from ARMA models with clipped data.

Ali, A. et al. (2011), used ARIMA models to forecast stock market performance of the Oil and Gas Companies in Pakistan.

Mondal, P. et al. (2014) applied ARIMA model on 56 stocks of India from different sectors in order to forecast their future returns. Their research concluded with being 85% successful forecasts.

Banerjee, D. (2014) utilized ARIMA model to forecast Indian stock market index. He concluded with the result that the short run prediction power is most suitable for the model in order to take best results.

Time series analysis is the approach of prediction that focuses on the previous behaviour of dependent variable. Time series models give another approach to analyze and forecast future developments based on previous behaviour of the objective.

Fang, J. et al. (2003) concluded that time-series forecasting has been usually performed using statistical-based methods, for example, the linear autoregressive (AR) models because of their flexibility to model many stationary processes. These include the ARMA (autoregressive moving average) model. Methods such as the linear autoregressive integrated moving average (ARIMA) is based on the evolution of the increments are used at times to remove or reduce first order non-stationarity. Hipel, K.W. et al. (1994) concluded that the concept of stationarity can be seen as a form of statistical equilibrium. The properties of mean and variance of a stationary process do not rely on time. There is a greater possibility that the time series will not be stationary if there is a greater historical observations data whereas for short time span we usually use differencing or transformation in our model to make it stationary.

3. Box-Jenkins Methodology

The model consists of three main approaches:

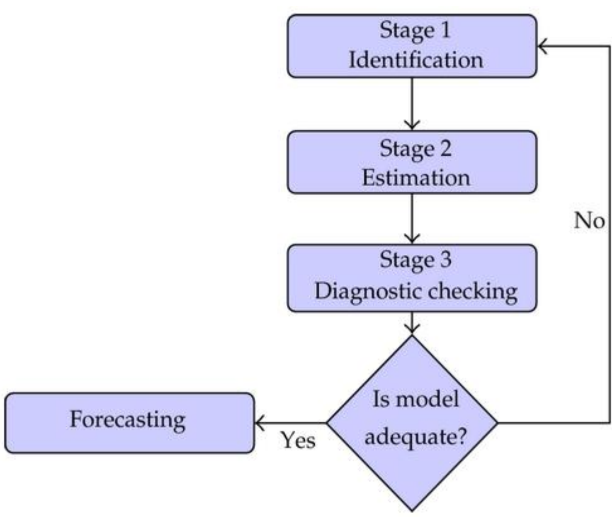

Figure 1. Stages of Box-Jenkins Methodology

I. Model Identification

In this approach, we check the stationarity of our variables and identify the seasonality in the dependent variable. If it is present then we remove it by differencing. We use plots of the autocorrelation (ACF) and partial autocorrelation (PACF) functions of the dependent time series to decide the usage of autoregressive or moving average component in our model.

II. Parameter Estimation

This approach deals with fitting our ARIMA model by choosing the best coefficients (p,d,q). The most common methods uses maximum likelihood estimation and checking AIC and BIC. The table below shows that how we choose our model on the basis of p,d,q values.

Table 1. Shapes and Indicated Model

Shape	Indicated Model
Exponential series decaying to 0	Auto Regressive (AR) model. pacf() function to be used to identify the order of the model
Alternative positive and negative spikes, decaying to 0	Auto Regressive (AR) model. pacf() function to be used to identify the order of the model
One or more spikes in series, rest all are 0	Moving Average(MA) model, identify order where plot becomes 0
After a few lags overall a decaying series	Mixed AR & MA model
Total series is 0 or nearly 0	Data is random
Half values at fixed intervals	We need to include seasonal AR term
Visible spikes, no decay to 0	Series is not stationary

III. Diagnostic Checking

This approach involves with checking the specified model is adequate or not. We can use different methods to check this like plotting histogram or Normal QQ plot, applying Ljung-Box test on residuals.

If the model is inadequate then we return to step one to build a new model.

4. Research Methodology

In the beginning, we will apply ARMA model if we get the stationary data. Stationarity is defined by Gujarati & Porter (2004) as the one that has a constant long-run mean and a time-invariant covariance. If the data is non-stationary then we will apply ARIMA model.

Autoregressive Moving Average (ARMA) model gives the explanation of a stationary process in terms of two polynomials, one for the autoregressive (p) and the second for the moving average (q). We express the ARMA(p,q) in mathematical form given as:

$$\begin{aligned} y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} \\ + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \end{aligned} \quad (1)$$

where $\varepsilon_t \sim N(0, \sigma^2)$

y_t represents the variable we are trying to predict

y_{t-1}, y_{t-2},…y_{t-p} are the previous or lagged values of that variable that are also called the autoregressive terms

ε_t is the disturbance or error term, ε_{t-1}, ε_{t-2},…, ε_{t-q} are the previous or lagged values of the error term that are also known as the moving averages

φ_1, φ_2 ,…, φ_p are the coefficients of autoregressive terms

θ_1, θ_2, … , θ_p are the coefficients of the moving average regressors

Autoregressive Integrated Moving Average (ARIMA) model is a conception of an autoregressive moving average (ARMA) model. Both of these models are applied to time series data for the better understanding of the data or to forecast future values in the series forecasting. These are mostly applied where the data is of non-stationary to make it stationary. It is composed of two parts

- an integrated (I) component (d) that represents the amount of differencing to be applied on the time series to make it stationary
- the autoregressive moving average (ARMA) model

We express the ARIMA (p,d,q) in mathematical form given as:

$$y_t^{'} = \mu + \phi_1 y_{t-1}^{'} + \ldots + \phi_p y_{t-p}^{'} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \quad (2)$$

In this section, in order to analyze, we apply Box-Jenkin methodology for KSE-100 time series data and forecast the values for next 90 days.

We are plotting the data

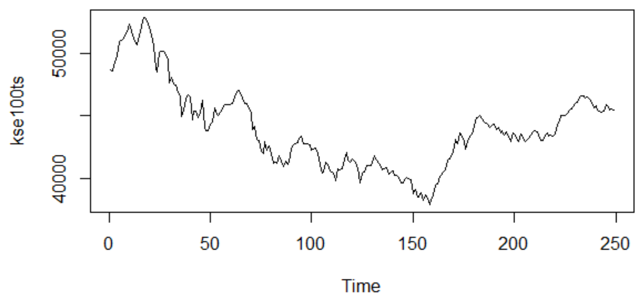

Figure 2. Time Series Plot of KSE-100 (1st May, 2017- 30th April, 2018)

This graph is showing us the index points with respect to time.

We are fitting our model by estimating the parameters by the method of MLE. We are choosing the best model that has smallest AIC and BIC.

Table 2. Summary of ARIMA(0,1,1) model of KSE-100

ARIMA(p,d,q)	AR	MA	AIC	BIC
0,1,1	0	0.0892	3805.12	3815.67

Its mathematical equation becomes

$$y_t' = \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (3)$$

$$y_t' = \varepsilon_t + 0.0892\varepsilon_{t-1} \quad (4)$$

Now, we are checking the normality of our model. For this purpose, we are plotting its histogram.

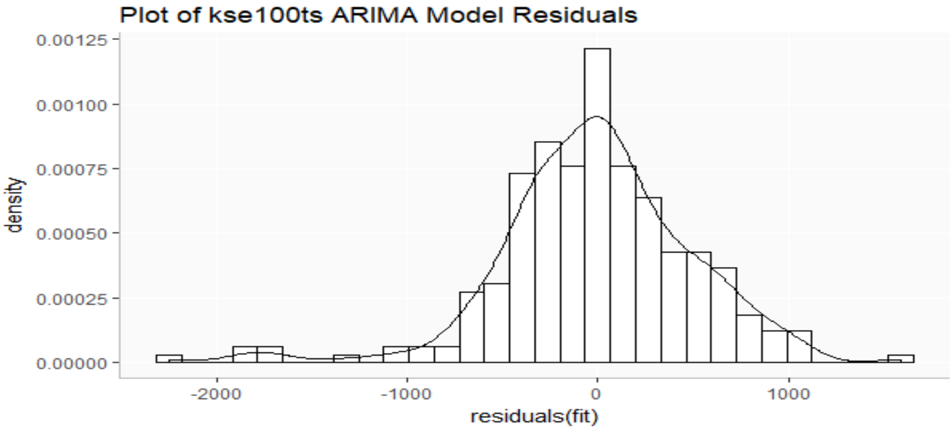

Figure 3. Histogram of residuals of ARIMA (0,1,1) model of KSE-100

The histogram shows that our fit model is normal. A slight inclination in histogram can be ignored because of the behavior of the stock market in one year as it is not certain that it will always perform well.

We are also checking its normality with the help of Normal QQ plot.

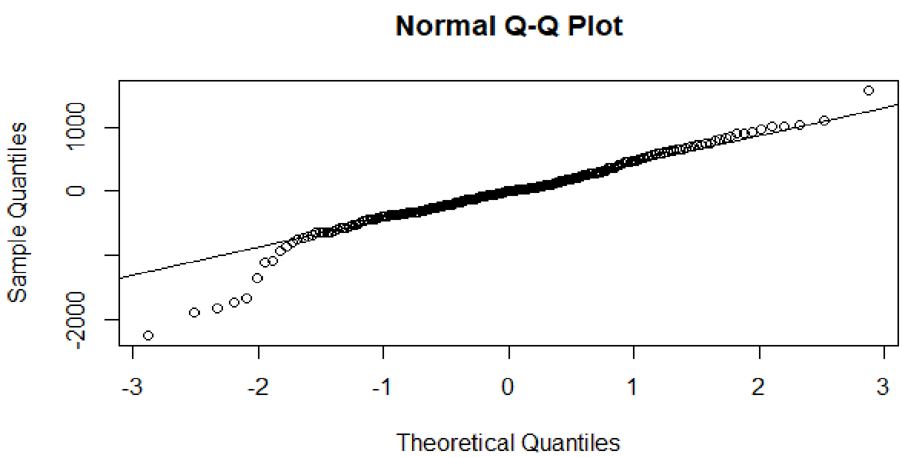

Figure 4. Normal QQ Plot of residuals of ARIMA(0,1,1) model of KSE-100

This plot shows us that our model is normal.

We are again confirming the normality by applying Ljung-Box Test on the residuals of our model. For this purpose, we are building our hypothesis:

H_0: The data is independently distributed

H_1: The data is not independently distributed

Its test statistic is

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k} \quad (5)$$

We are calculating its p-value.

Table 3. Summary of Ljung-Box Test on residuals of ARIMA (0,1,1) model of KSE-100

X-squared	Df	p-value
4.1432	6	0.6573

As its p-value > 0.05, therefore the residuals are independent, and the model is normal.

Now we are forecasting our future values of KSE-100 for next 90 days and plotting it afterwards.

Table 4. Forecasted values of KSE-100 by ARIMA (0,1,1) model

Point	Forecast	Lo 99.5	Hi 99.5
250	45471.33	44025.02	46917.63
251	45458.36	43319.79	47596.93
252	45445.40	42789.25	48101.55
253	45432.43	42344.26	48520.60
254	45419.47	41952.70	48886.24
255	45406.50	41598.60	49214.41
256	45393.54	41272.64	49514.44
257	45380.57	40968.83	49792.32
258	45367.61	40683.04	50052.18
259	45354.65	40412.29	50297.00
260	45341.68	40154.34	50529.02
261	45328.72	39907.45	50749.99
262	45315.75	39670.24	50961.27
263	45302.79	39441.60	51163.98
264	45289.82	39220.62	51359.03
265	45276.86	39006.54	51547.18
266	45263.90	38798.72	51729.07
267	45250.93	38596.59	51905.27
268	45237.97	38399.70	52076.23
269	45225.00	38207.63	52242.38
270	45212.04	38020.01	52404.06
271	45199.07	37836.54	52561.61
272	45186.11	37656.93	52715.29
273	45173.14	37480.92	52865.36
274	45160.18	37308.31	53012.05
275	45147.22	37138.87	53155.56
276	45134.25	36972.43	53296.07
277	45121.29	36808.83	53433.74
278	45108.32	36647.91	53568.74
279	45095.36	36489.53	53701.19
280	45082.39	36333.57	53831.22
281	45069.43	36179.91	53958.95
282	45056.47	36028.44	54084.49
283	45043.50	35879.06	54207.94
284	45030.54	35731.69	54329.39
285	45017.57	35586.23	54448.92
286	45004.61	35442.60	54566.62
287	44991.64	35300.74	54682.55
288	44978.68	35160.57	54796.79
289	44965.71	35022.02	54909.41
290	44952.75	34885.05	55020.45
291	44939.79	34749.58	55129.99
292	44926.82	34615.57	55238.08
293	44913.86	34482.96	55344.76
294	44900.89	34351.71	55450.08
295	44887.93	34221.77	55554.09
296	44874.96	34093.10	55656.83
297	44862.00	33965.65	55758.35
298	44849.04	33839.40	55858.67
299	44836.07	33714.30	55957.84
300	44823.11	33590.33	56055.89
301	44810.14	33467.43	56152.85
302	44797.18	33345.60	56248.76
303	44784.21	33224.78	56343.64
304	44771.25	33104.97	56437.53
305	44758.28	32986.13	56530.44
306	44745.32	32868.22	56622.42
307	44732.36	32751.24	56713.47
308	44719.39	32635.16	56803.63
309	44706.43	32519.94	56892.91
310	44693.46	32405.58	56981.35
311	44680.50	32292.05	57068.95
312	44667.53	32179.32	57155.75
313	44654.57	32067.39	57241.75
314	44641.61	31956.23	57326.98
315	44628.64	31845.82	57411.46
316	44615.68	31736.15	57495.20
317	44602.71	31627.20	57578.22
318	44589.75	31518.96	57660.54
319	44576.78	31411.41	57742.16
320	44563.82	31304.53	57823.11
321	44550.86	31198.31	57903.40
322	44537.89	31092.74	57983.04
323	44524.93	30987.80	58062.05
324	44511.96	30883.48	58140.44
325	44499.00	30779.77	58218.22
326	44486.03	30676.66	58295.41
327	44473.07	30574.13	58372.01
328	44460.10	30472.17	58448.03
329	44447.14	30370.78	58523.50
330	44434.18	30269.94	58598.41
331	44421.21	30169.64	58672.78
332	44408.25	30069.88	58746.62
333	44395.28	29970.63	58819.93
334	44382.32	29871.90	58892.74
335	44369.35	29773.67	58965.03
336	44356.39	29675.94	59036.84
337	44343.43	29578.70	59108.15
338	44330.46	29481.93	59178.99
339	44317.50	29385.63	59249.36

The Table 3 along with the Figure 5 indicates the predicted values. It indicates that in the next 90 days the KSE-100 Index is going to be bearish with constant declination.

Now we are going to discuss the results of forecast and actual values of KSE-100 by fitting the ARIMA (0,1,1) model.

Table 5. Analysis of Forecast and actual values of KSE-100

Date	Actual	Forecast	Date	Actual	Forecast
01/05/2018	45,488.86	45471.33	06/07/2018	40,284.14	44887.93
02/05/2018	45,196.37	45458.36	09/07/2018	39,288.48	44874.96
03/05/2018	44,746.63	45445.40	10/07/2018	39,452.81	44862.00
04/05/2018	44,536.91	45432.43	11/07/2018	39,586.76	44849.04
07/05/2018	44,378.52	45419.47	12/07/2018	39,875.12	44836.07
08/05/2018	44,066.96	45406.50	13/07/2018	40,271.00	44823.11
09/05/2018	43,795.00	45393.54	16/07/2018	39,665.77	44810.14
10/05/2018	43,855.78	45380.57	17/07/2018	39,932.98	44797.18
11/05/2018	43,594.79	45367.61	18/07/2018	40,897.90	44784.21
14/05/2018	42,498.86	45354.65	19/07/2018	41,795.59	44771.25
15/05/2018	42,459.53	45341.68	20/07/2018	41,221.75	44758.28
16/05/2018	42,301.20	45328.72	23/07/2018	40,463.98	44745.32
17/05/2018	41,869.65	45315.75	24/07/2018	41,339.22	44732.36
18/05/2018	41,623.52	45302.79	26/07/2018	42,089.16	44719.39
21/05/2018	41,648.65	45289.82	27/07/2018	42,786.45	44706.43
22/05/2018	42,744.82	45276.86	30/07/2018	43,556.63	44693.46
23/05/2018	42,772.25	45263.90	31/07/2018	42,712.43	44680.50
24/05/2018	42,536.16	45250.93	01/08/2018	42,810.04	44667.53
25/05/2018	42,074.09	45237.97	02/08/2018	42,330.32	44654.57
28/05/2018	42,138.57	45225.00	03/08/2018	42,505.05	44641.61
29/05/2018	42,622.74	45212.04	06/08/2018	42,808.66	44628.64
30/05/2018	42,546.48	45199.07	07/08/2018	42,760.13	44615.68
31/05/2018	42,846.64	45186.11	08/08/2018	42,731.86	44602.71
01/06/2018	42,912.81	45173.14	09/08/2018	42,923.95	44589.75
04/06/2018	43,268.29	45160.18	10/08/2018	42,842.18	44576.78
05/06/2018	43,702.58	45147.22	13/08/2018	42,637.58	44563.82
06/06/2018	44,144.20	45134.25	15/08/2018	42,446.56	44550.86
07/06/2018	43,948.11	45121.29	16/08/2018	41,960.80	44537.89
11/06/2018	43,931.16	45108.32	17/08/2018	42,446.62	44524.93
12/06/2018	43,228.90	45095.36	20/08/2018	42,425.10	44511.96
13/06/2018	43,507.50	45082.39	24/08/2018	42,588.29	44499.00
14/06/2018	43,680.68	45069.43	27/08/2018	42,745.78	44486.03
19/06/2018	43,682.84	45056.47	28/08/2018	42,544.47	44473.07
20/06/2018	43,002.83	45043.50	29/08/2018	42,249.44	44460.10
21/06/2018	42,358.61	45030.54	30/08/2018	41,863.52	44447.14
22/06/2018	41,637.38	45017.57	31/08/2018	41,742.24	44434.18
25/06/2018	40,978.23	45004.61	03/09/2018	41,581.96	44421.21
26/06/2018	41,246.09	44991.64	04/09/2018	41,753.89	44408.25
27/06/2018	41,717.99	44978.68	05/09/2018	41,620.96	44395.28
28/06/2018	41,997.85	44965.71	06/09/2018	41,266.39	44382.32
29/06/2018	41,910.90	44952.75	07/09/2018	40,854.77	44369.35
02/07/2018	41,734.05	44939.79	10/09/2018	40,684.05	44356.39
03/07/2018	41,564.42	44926.82	11/09/2018	40,759.53	44343.43
04/07/2018	40,345.68	44913.86	12/09/2018	40,522.04	44330.46
05/07/2018	40,238.81	44900.89	13/09/2018	41,049.91	44317.50

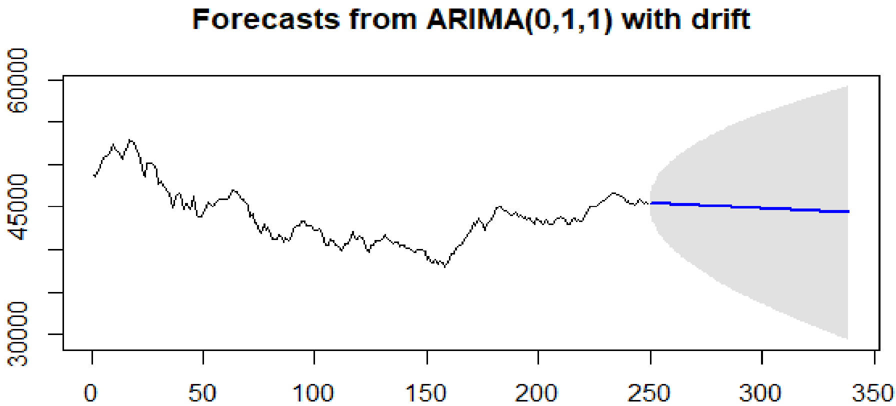

Figure 5. Forecast plot of KSE-100 by ARIMA (0,1,1) model

When examining the forecasts for KSE-100, we notice that the fits had a range of 44317.50-45471.33. However, the actual prices had a range of 39,288.48-45,488.86. There is not much change in the fits because the coefficient of the moving average term is 0.0892 and the model does not contain a constant, so the forecast's price is only 0.0892 of the previous week's price. Modeling the series using previous observation, we see that fits behave more like the actuals, compared to the series that contained all the data points.

5. Conclusions

In this work, we have analyzed the indices of PSX during 1st May 2017 – 30th April 2018. Here, we have index namely KSE-100. By looking at the forecasting results and graphs of KSE-100, we may conclude that in the future, the stock market is going to become bearish with the constant declination in the trend at least for next 90 days. We also compared the forecast values with actual values, and we can observe that the fits behave more like the actuals.

REFERENCES

[1] Ali, A., Kashif, M., and Aslam, M. (2011), "Estimation and Forecast of the Models for Stock Market Performance of the Oil & Gas Companies in Pakistan", Pakistan Journal of Social Sciences (PJSS) Vol. 31, No. 2, pp. 345-363.

[2] Bagnall, A. J. and Janacek, G. J. (2004), "Clustering Time Series from ARMA Models with Clipped Data" in KDD, W. Kim, R. Kohavi, J. Gehrke, and W. DuMouchel, Eds. ACM.

[3] 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.

[4] Box, G., & Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, California, USA.

[5] Chi, S. C., Chen, H. P., and Cheng, C. H. (1999). "A Forecasting Approach for Stock Index Future Using Grey Theory and Neural Networks". IEEE International Joint Conference on Neural Networks, 3850-3855.

[6] Fan, J. and Yao, Q. (2003) "Nonlinear Time Series: Nonparametric and Parametric Methods", Springer-Verlag, New York, NY.

[7] Gujarati, D., & Porter, D. (2004). Basic Econometrics (4th Ed.). McGraw-Hill, New York, USA.

[8] Hipel, K.W., McLeod, A.I. (1994), "Time Series Modelling of Water Resources and Environmental Systems", Amsterdam, Elsevier.

[9] 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.

[10] Pai, P. and Lin, C. (2005), "A hybrid ARIMA and support vector machines model in stock price prediction", Omega vol.33 pp. 497-505.

[11] Raicharoen, T., Lursinsap, C., Sanguanbhoki, P., (2003), "Application of critical support vector machine to time series prediction", Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on Volume 5, pages: V-741-V-744.

[12] Tong, H. (1983), "Threshold Models in Non-Linear Time Series Analysis", Springer-Verlag, New York.

[13] Tseng, K.C., Kwon, O., and Tjung, L.C. (2012), "Time series and neural network forecast of daily stock prices", Investment Management and Financial Innovations, Vol. 9, No 1, pp 32-54.

[14] Wang, J. J, Wang, J. Z., Zhang, Z. G. and Guo, S. P. (2012), "Stock index forecasting based on a hybrid model", Omega vol. 40 pp.758-766.

[15] Zhang, G. P. (2007), "A neural network ensemble method with jittered training data for time series forecasting", Information Sciences 177, pages: 5329–5346.