Volume 3, No. 8, August 2012

*Journal of Global Research in Computer Science*

ISSN-2229-371X

**RESEARCH PAPER**

**Available Online at www.jgrcs.info**

# PREDICTIVE ANALYSIS OF SEQUENTIAL PATTERN IN TIME SERIES TEMPORAL DATASETS FOR INDIAN STOCK MARKET

Ms. Shilpa Dang[*1], Dr. Naveeta Mehta[2]

*[1]MMICT&BM (MCA), M.M.University, Ambala, Haryana, India
Dangshilpa2@gmail.com
[2] Computer Science, K.U.K/ Govt. college Karnal, Ambala, Haryana, India
Navita80@gmail.com

***Abstract:*** Data Mining is the process of extracting interesting information or pattern from large information of stock market .Therefore, the objective of the research work is to find and characterize interesting sequential pattern in temporal data sets. To accomplish the above said objective time series procedures can be used to analyze data collected over time, commonly called a time series. These procedures include simple forecasting and smoothing methods, correlation analysis methods, and ARIMA modeling. Although correlation analysis may be performed separately from ARIMA modeling, the author presents the correlation methods as part of ARIMA modeling.

## INTRODUCTION

Temporal data stored in a temporal database is different from the data stored in non-temporal database in that a time period attached to the data expresses when it was valid or stored in the database. Conventional databases consider the data stored in it to be valid at time instant now, they do not keep track of past or future database states. By attaching a time period to the data, it becomes possible to store different database states. There are mainly two different notions of time which are relevant for temporal databases. One is called the valid time, the other one is the transaction time. Valid time denotes the time period during which a fact is true with respect to the real world. Transaction time is the time period during which a fact is stored in the database.

A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. For example, measuring the value of retail sales each month of the year would comprise a time series. This is because sales revenue is well defined, and consistently measured at equally spaced intervals. Data collected irregularly or only once are not time series.

An observed time series can be decomposed into four components as explained below:

a. Trend: The long-term tendency of a series to rise or fall (upward trend or downward trend).
b. Seasonality: The periodic fluctuation in the time series within a certain time period. These fluctuations form a pattern that tends to repeat from one seasonal period to another.
c. Cycles: long departures from the trend due to factors others than seasonality. Cycles generally occur over a large time interval, and the lengths of time between successive peaks or troughs of a cycle are not necessarily the same.
d. Irregular movement: the movement left after accounting for trend, seasonal and cyclical movements; random noise or error in a time series

## SEQUENTIAL PATTERN IN TIME SERIES TEMPORAL DATASETS

To carry on this research work trend analysis procedure has been opted. This procedure is used to ascertain the sequential pattern among the temporal datasets. It is the child procedure of time series analysis. In it statistical data is identified and characterized; to retrieve interesting sequential pattern in temporal data sets. For this purpose a secondary database for the period Jan. 2011 to Dec. 2011 is analyzed; which belongs to Bombay Stock Exchange, India. The dataset for Bombay Stock Exchange, India has shown five attributes as date, open, high, low and close with 252 tuples. Where date states the temporal dataset, open shows the open price of the shares in stock market; high the highest bid values of that share and low is the lowest bid value in respect to the concerned shares. Close the close price of that share for that day. The procedure of Trend analysis has been applied in MINITAB 15.0; which is well known data mining tool used for the purpose of forecasting and prediction. After applying this procedure of trend analysis the following results are retrieved. These are interpreted as follow:

## ANALYSIS AND INTERPRETATION OF SEQUENTIAL PATTERN

Figure 1 shows the sequential pattern in the form of trend line for the shares goes on highest price on the day. In it the black line shows the actual trend of the pattern for the share price during the period and the green line is the forecast/ prediction for the next 60 days. The red line shows the goodness of fit line. The trend line is obtained by using linear equation which states as Yt= 2013.4 – 0.111893*t. Where Yt - represents the value of the variable that you

measure, at time t, t - represents the time units (t**2 means t ), coefficients - are constants used to express the variable under consideration as a linear, quadratic, exponential or S-curve function of time. The Figure shows the date at x-axis and High (bid value at y-axis). Along with these values the figure also states the accuracy values. For all three measures, smaller values generally indicate a better fitting model.

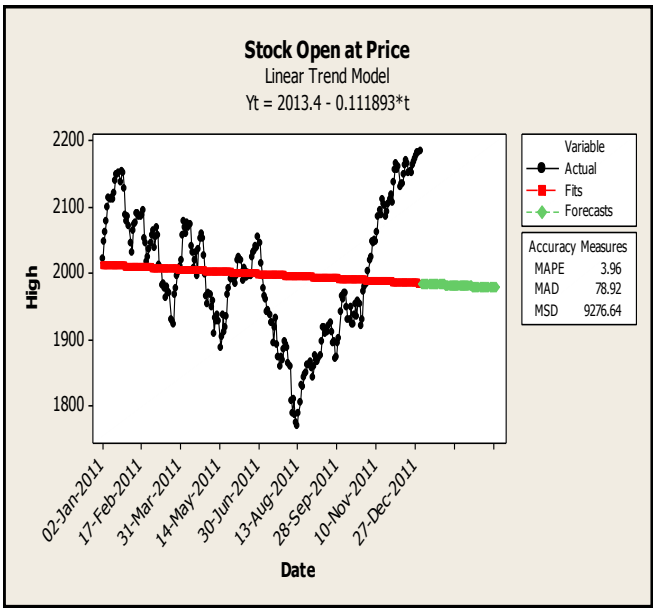


Figure 1: Trend Line/ Forecasting for Stock price reached to High Bid Value

Figure 2 shows the Normal Probability Plot when share bid value is high. For the stock data, the residuals do not appear to follow a straight line. Some evidence of non normality exists at the tails, although it is not extreme. Therefore a normality test to determine whether the residuals are normal may be conducted.

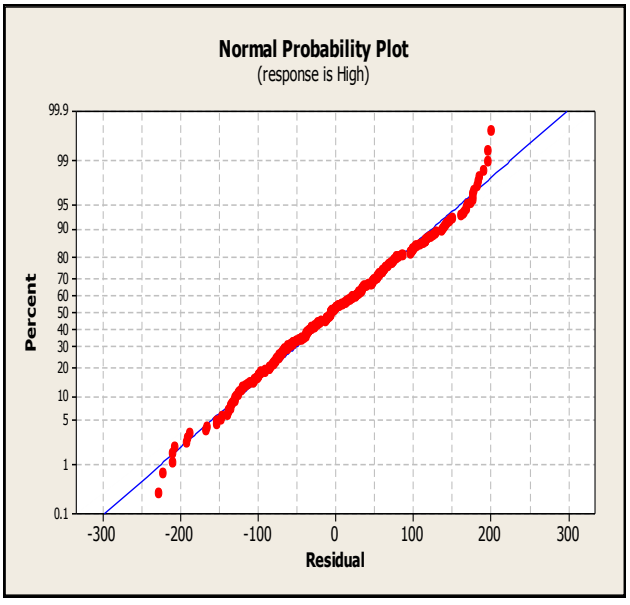


Figure 2: Normal Probability Plot when share bid value is high

Figure 3 examines Residual versus Fit Plot when share bid value is high. Based on this plot, the residuals appear to be randomly scattered about zero. No evidence of non constant variance, missing terms, or outliers exists.

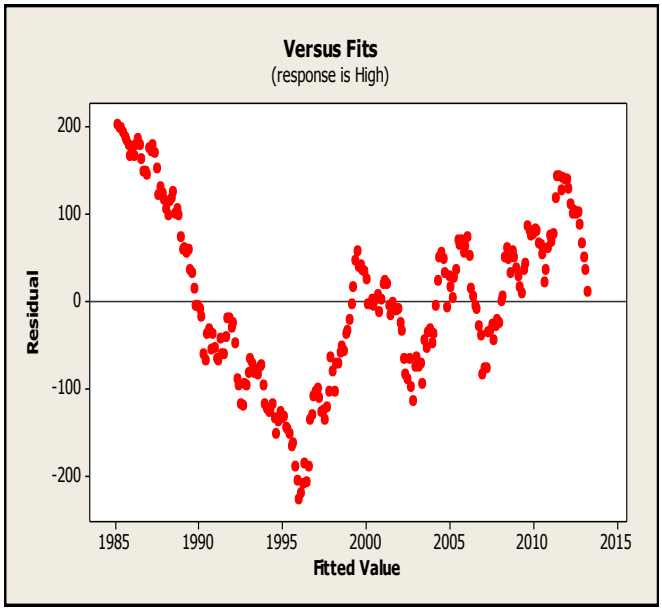


Figure 3: Residual versus Fit Plot when share bid value is high

Figure 4 provides the histogram for the equipment data, no evidence of skewness or outliers exists as per the diagram. Pattern of this figure was low to high from earlier date to mid of year whereas high to low in the last half year of the year.

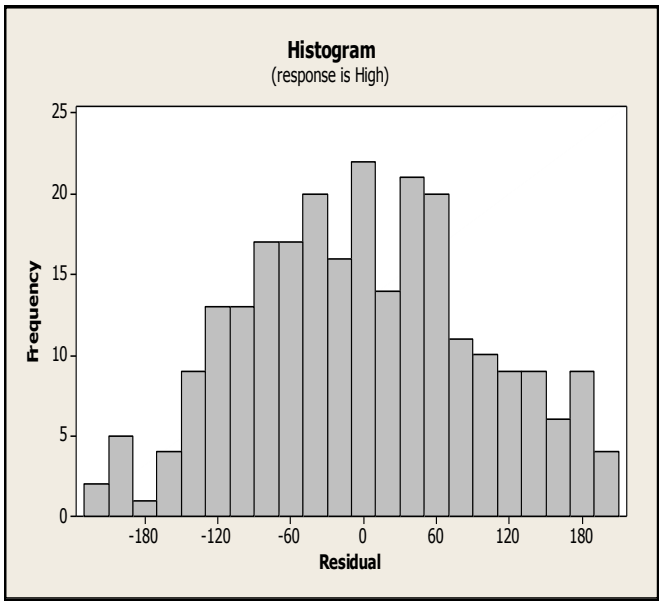


Figure 4: Histogram of Residual when Share bid in High.

Figure 5 examines the residual versus order. Here, the residuals appear to be randomly scattered about zero. No evidence exists that the error terms are correlated with one another.

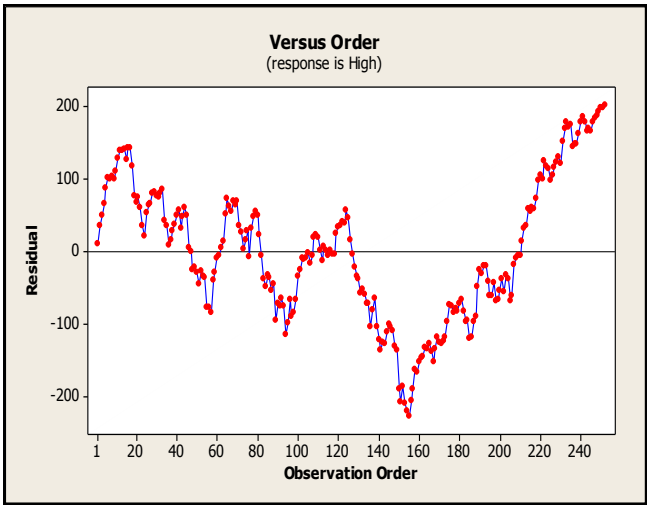


Figure 5: Residual versus order when stock price is high

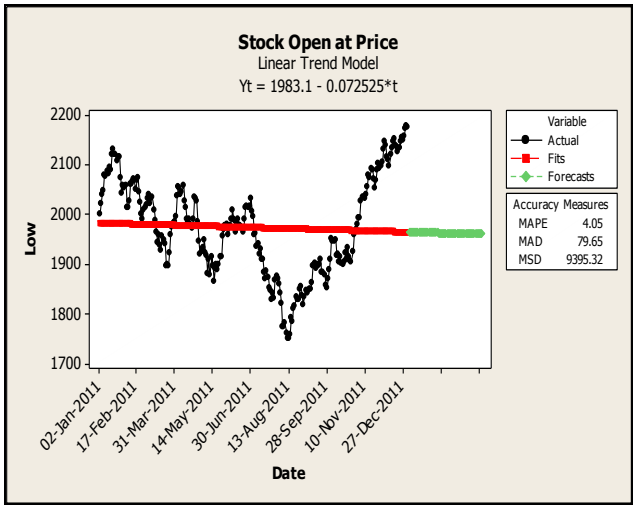


Figure 6: Trend Line/ Forecasting for Stock price reached to Low Bid Value

Figure 6 repeats the complete procedure of Trend sequential chart for the Sock value price goes on low value throughout the year. Along with actual trend line in black color, fits curve in red color and forecast value for next 60 days in green color represents the accuracy measures which are smaller and accurate.

Keeping in view the above chart/ graph as shown in Figure 1 to Figure 6, the predictive value has been retrieved as shown in the following table1. This table show close price of share for the next year i.e. Jan 2012 to Dec. 2012. For a review/ speciman the predicted price pattern has been shown only for the month of January 2012 and Dec. 2012.

Table 1: Predictive Close Price for the share during Jan. 2012 to Dec. 2012

| Predictive Values | | | Predictive Values | | |
|---|---|---|---|---|---|
| Date | Open | FITSClose (next Year)(Jan-Dec 2012) | Date | Open | FITSClose ( next Year)(Jan-Dec 2012) |
| 02-Jan-2011 | 2011.1 | 1996.7949 | 03-Dec-2011 | 2153.3 | 1977.8622 |
| 05-Jan-2011 | 2020.8 | 1996.7133 | 06-Dec-2011 | 2145.4 | 1977.7806 |
| 06-Jan-2011 | 2044.6 | 1996.6317 | 07-Dec-2011 | 2154.1 | 1977.699 |
| 07-Jan-2011 | 2056.8 | 1996.5501 | 08-Dec-2011 | 2118.1 | 1977.6174 |
| 08-Jan-2011 | 2089.6 | 1996.4685 | 09-Dec-2011 | 2109.6 | 1977.5358 |
| 09-Jan-2011 | 2083.6 | 1996.3869 | 10-Dec-2011 | 2121.2 | 1977.4542 |
| 12-Jan-2011 | 2093.5 | 1996.3053 | 13-Dec-2011 | 2141.2 | 1977.3726 |
| 13-Jan-2011 | 2113.1 | 1996.2237 | 14-Dec-2011 | 2145.1 | 1977.291 |
| 14-Jan-2011 | 2104.3 | 1996.1421 | 15-Dec-2011 | 2159.7 | 1977.2094 |
| 15-Jan-2011 | 2101.9 | 1996.0604 | 16-Dec-2011 | 2160.0 | 1977.1277 |
| 16-Jan-2011 | 2126.1 | 1995.9788 | 17-Dec-2011 | 2142.7 | 1977.0461 |
| 20-Jan-2011 | 2149.0 | 1995.8972 | 20-Dec-2011 | 2142.2 | 1976.9645 |
| 21-Jan-2011 | 2139.3 | 1995.8156 | 21-Dec-2011 | 2134.8 | 1976.8829 |
| 22-Jan-2011 | 2146.3 | 1995.734 | 22-Dec-2011 | 2146.0 | 1976.8013 |
| 23-Jan-2011 | 2124.8 | 1995.6524 | 23-Dec-2011 | 2153.3 | 1976.7197 |
| 26-Jan-2011 | 2120.6 | 1995.5708 | 27-Dec-2011 | 2168.8 | 1976.6381 |
| 27-Jan-2011 | 2148.1 | 1995.4892 | 28-Dec-2011 | 2156.8 | 1976.5565 |
| 28-Jan-2011 | 2125.0 | 1995.4076 | 29-Dec-2011 | 2171.0 | 1976.4749 |
| 29-Jan-2011 | 2085.5 | 1995.326 | 30-Dec-2011 | 2177.5 | 1976.3933 |
| 30-Jan-2011 | 2068.4 | 1995.2444 | 31-Dec-2011 | 2179.0 | 1976.3117 |

## CONCLUSION

In the present research work the dataset for Bombay Stock Exchange, India has been taken care off. This dataset shows five attributes as date, open, high, low and close. We have used date and open attribute has been taken care off for determining the sequential pattern. There were 252 tuples ranging from January 2011 to December 2011. Where date

states the temporal dataset, open shows the open price of the shares in stock market. Various figures have been presented for clear understanding of sequential pattern in time series temporal datasets.

## REFRENCES


[1]. Balaji Padmanabhan and Alexander Tuzhilin, "Pattern Discovery in Temporal Databases: A Temporal Logic Approach", KDD-96 Proceedings, Leonard N. Stern School of Business, New York University, Copyright © 1996, AAAI (www.aaai.org).

[2]. Girish Keshav Palshikar, "Fuzzy Temporal Patterns for Analyzing Stock Market Databases", Tata Research Development and Design Centre, ADVANCES IN DATA MANAGEMENT 2000, Tata McGraw-Hill Publishing Company Ltd. ©CSI 2000.

[3]. Gerasimos Marketos, Konstantinos Pediaditakis, Yannis Theodoridis, and Babis Theodoulidis, "Intelligent Stock Market Assistant using Temporal Data Mining" **,**Database Group, Information Systems Laboratory, Department of Informatics, University of Piraeus, Greece.

[4]. N. Pughazendi and Dr. M. Punithavalli (2011), "Tempopral Databases and Frequent Pattern Mining Techniques", International Journal of P2P Network Trends and Technology- July to Aug Issue 2011.

[5]. Pandya V H (1992), "Securities and Exchange Board of India: Its Role, Powers, Functions and Activities", Chartered Secretary, Vol. 22, No. 9 (Sept), p. 783.

[6]. Srivatsan Laxman and P. S. Sastry, "A survey of temporal data mining", Sadhana, Vol. 31, Part 2, April 2006, pp. 173–198. © Printed in India.

[7]. Salman Azhar ,Greg J. Badros and Arman Glodjo,Ming - Yang Kao and John H .Reif, "Data Compression techniques in Stock Market Prediction,".

[8]. Malhotra, N. K. (2007). Marketing Research: An Applied Orientation (5th Ed.). New Delhi: Prentice Hall. p. 86,284,286,335,339,560,561.


 