A Robust Analysis and Forecasting Framework for the Indian Mid Cap Sector Using Times Series Decomposition Approach

Jaydip Sen 1

1Praxis Business School

November 1, 2023

Abstract

Prediction of stock prices using econometrics and machine learning based approaches poses significant challenges to the research community since the movement of stock prices are essentially random in its nature. However, significant development and rapid evolution of sophisticated and complex algorithms which are capable of analyzing large volume of time series data, coupled with availability of high-performance hardware and parallel computing architecture over the last decade, has made it possible to efficiently process and effectively analyze voluminous stock market time series data in an almost real-time environment. In this paper, we propose a decomposition-based approach for time series analysis of the Indian mid cap sector and also present a highly robust and accurate prediction framework consisting of six forecasting methods for predicting the future values of the time series. Extensive results are presented on the performance of each forecasting method and the reasons why a particular method has performed better than the others have been critically analyzed.
A Robust Analysis and Forecasting Framework for the Indian Mid Cap Sector Using Times Series Decomposition Approach

Jaydip Sen
Praxis Business School, Bakrahat Road, Off Diamond Harbor Road, Kolkata – 700104
West Bengal, INDIA
e-mail: jaydip.sen@acm.org

ABSTRACT

Prediction of stock prices using econometrics and machine learning based approaches poses significant challenges to the research community since the movement of stock prices are essentially random in their nature. However, significant development and rapid evolution of sophisticated and complex algorithms which are capable of analyzing large volume of time series data, coupled with availability of high-performance hardware and parallel computing architecture over the last decade, has made it possible to efficiently process and effectively analyze voluminous stock market time series data in an almost real-time environment. In this paper, we propose a decomposition-based approach for time series analysis of the Indian mid cap sector and also present a highly robust and accurate prediction framework consisting of six forecasting methods for predicting the future values of the time series. Extensive results are presented on the performance of each forecasting method and the reasons why a particular method has performed better than the others have been critically analyzed.

Key words: Time Series Decomposition, Trend, Seasonal, Random, Holt Winters Forecasting model, Auto Regression (AR), Moving Average (MA), Auto Regressive Integrated Moving Average (ARIMA), Partial Auto Correlation Function (PACF), Auto Correlation Function (ACF).

JEL Classification: G 11, G 14, G 17, C 63

1. Introduction

Developing an accurate and efficient forecasting framework for robust prediction of stock prices has been one of the most exciting yet difficult challenges faced by the researchers working in the field of machine learning and analytics. Numerous technical, fundamental and statistical indicators have been proposed by researchers for accurately predicting the prices of stocks. Sen & Datta Chaudhuri (2016a; 2016b; 2016c; 2016d) proposed a novel approach based on time series decomposition and analysis for efficient portfolio diversification and prediction of stock prices. The authors have hypothesized all sectors of an economy do not exhibit similar pattern of variations in their stock prices. In fact it is more usual to find that different sectors exhibiting different patterns in their trend, different characteristics in their seasonality behavior, and varying degree of randomness in their time series values. While on one side the efficient market hypothesis has argued for the randomness aspect of stock price movements, on the other side, there are
propositions to counter the hypothesis by delving into various fundamental characteristics of different sectors and different stocks in those sectors. We argue that in addition to the differences in the fundamental characteristics among stocks belonging of different companies, performances of different stocks are also very much dependent and coupled to the sectors to which the stocks belong. Since behavior of each sector of economy is influenced by its unique set of factors, the pattern of price movement of stocks belonging to different sectors are also determined and influenced by these factors.

In this work, our goal is to study the behavioral pattern exhibited by the time series of the mid cap sector of India, so that the salient properties of that sector can be better understood. By its definition, a mid cap company has a market capitalization between Indian Rupees (INR) 50 billion to INR 200 billion. For the purpose of our study, the monthly average index values of the mid cap sector are used for the period January 2010 - December 2016 as per the Bombay Stock Exchange (BSE). The monthly time series data is decomposed into its three components using functions defined in the R programming language. Based on the results of decomposition, we demonstrate how several interesting characteristics of the time series can be extracted to gain useful insights into its behavioral pattern. We particularly illustrate how a deeper analysis of the trend, seasonal and random components provides us with useful information about the growth pattern, seasonal properties and randomness exhibited by the time series index values. For predicting future behavioral pattern, we also propose an extensive framework for time series forecasting consisting of six methods of prediction of time series index values. The six forecasting methods are critically analyzed in terms of their accuracy in forecasting and why some methods have performed better than the others are also explained.

The organization of the paper is as follows. In Section 2, we present a detailed description of the methodology that we have followed in this work. We discuss in detail the method of decomposition of the mid cap sector time series into its various components. Section 3 presents an extensive results of decomposition of the time series values into its trend, seasonal and random components. The decomposition results are analyzed in depth in order to understand several important characteristics and behavior revealed by the time series. In Section 4, we propose a set of six forecasting methods for predicting the future values of the time series. Section 5 provides extensive results on the performance of the six forecasting methods on the mid cap sector time series data. Each of the proposed algorithms are evaluated on the basis of six metrics of performance in their forecasting accuracy and a comparative analysis of the methods is presented. These six metrics are: maximum error, minimum error, mean error, standard deviation of error, the root mean square error (RMSE), and the ratio of the RMSE value and the mean index value. Section 6 presents a brief literature survey on some of the existing work on time series analysis and forecasting. In Section 7, we conclude the paper and highlight some future scope of work.

2. Methodology

This section presents a brief discussion on the methodology that has been followed in this work. The rich features of the programming language R have been exploited in all three activities of analytics: data management, data analysis and presentation of analysis results. Ihaka & Gentleman (1996) provided a detailed description of various features of the R programming language and its power and capabilities in data management and data analysis work. R is an open source language with a very large and rich collection of libraries with numerous useful in-built functions that makes it one of the most powerful tools for complex analytics projects. In this work, the monthly average index values have been used from the Bombay Stock Exchange (BSE) of India for the mid cap sector for the period January 2010 - December
2016. The average monthly index values of the mid cap sector time series for the 7 year period are then stored in a plain text (.txt) file. Hence, the plain text file is now populated with 84 records, with each record referring to the average index value for one among the 84 months in the 7 year period under our investigation. The text file is then converted into an R data object by reading it into the object using the \texttt{scan()} function. The resultant R data object is then transformed into an R time series object by invoking the \texttt{ts()} function in R with its three parameters as: (i) the R data object, (ii) a frequency value of 12 and (iii) the starting month of the time series, i.e., January 2010. The frequency value is fixed at 12 so that the monthly seasonality characteristics of the time series can be analyzed. Finally, using the \texttt{decompose()} function in R that is defined in the library TTR, the time series object is decomposed into its three components: trend, seasonal and random. The aggregate time series of the mid cap sector and its three components are plotted using the \texttt{plot()} function in R. Based on the numeric values of the components and the graphs, various important characteristics of the time series are analyzed.

We make a detailed analysis of the decomposition results of the times series of the mid cap sector so that various important properties revealed by the time series can be understood properly. Then, we propose six robust forecasting methods that can be applied on the time series so that its future values and behavior can be efficiently and accurately forecasted. In order to verify the accuracy in forecasting for each of the six methods, the forecast models are built using the mid cap time series data for the period January 2010 - December 2015. Once the models are built, they are used for forecasting the time series values for each month of the year 2016. Since the actual values of the time series for all months of 2016 are already available with us, the error in forecasting is easily computed by computing the percentage by which the predicted values deviate from their corresponding actual time series values for each forecasting methods proposed. We also carry out a detailed comparative analysis of the six forecasting methods using various useful metrics proposed by us. Based on the values of the metrics for each forecasting method, we critically analyze the reason why a particular method performs better than the other methods for the mid time series.

Sen & Datta Chaudhuri (2016a; 2016b) illustrated how time series decomposition approach can be effectively used in analyzing and forecasting time series of the Indian auto sector. In another work, Sen & Datta Chaudhuri (2016c) investigated the behavior of two different sectors of the Indian economy – the small cap sector and the consumer durable sector. The study revealed the fact that while the small cap sector time series is dominated by a strong random component, the consumer durable sector times series exhibits the presence of a significant seasonal component. Following another approach of time series analysis, Sen & Datta Chaudhuri (2016d) analyzed the behavior of the Indian information technology (IT) sector time series and the Indian capital goods sector time series. In yet another work, using the time series decomposition-based approach, Sen & Datta Chaudhuri (2016e) illustrated how time series analysis enables us to check the consistency between the fund style and actual fund composition of a mutual fund. Sen & Datta Chaudhuri (2017a; 2017b) also analyzed the characteristics of the Indian healthcare and FMCG sector time series and proposed robust and efficient forecasting techniques to accurately predict the future values of the time series index of the two sectors. Sen (2017a; 2017b) carried out studies on the realty sector and the metal sector of the Indian economy and identified some interesting characteristics of these two sectors.

In this work, we illustrate how time series decomposition-based approach can be utilized in an effective and accurate analysis of the behavior of the mid cap sector time series of the Indian economy. Based on the time series values for the mid cap sector during the period January 2010 - December 2016, we carry
out a decomposition exercise on the time series and make a detailed analysis of the decomposition results so that the salient properties of the sector can be understood. In addition, we propose six different approaches to forecasting that can be applied on the mid cap time series for predicting its future values. Based on several metrics, we carry out a detailed comparative analysis of the forecasting methods, and examine which forecasting method performs most efficiently for the mid cap time series. We also analyze the reason why some methods perform very well and produce small error in forecasting, while the others produce higher margin of error.

3. Time Series Decomposition Results

We present the decomposition results for the time series of the mid cap sector index values based on the records of the BSE for the period January 2010 - December 2016. First, we create a plain text (.txt) file that contains the monthly average index values of the mid cap sector for the period January 2010 - December 2016. This file includes 84 records each record representing the average index value for a month for the 84 months in the 7 years under our study. The `scan( )` function in R language is used to read the text file and store it in an R data object. The resultant R data object is converted into a time series object using the R time series function - `ts( )`. The value of the `frequency` parameter for the `ts( )` function is taken as 12 so that the decomposition of the time series is done on a monthly basis. Once the time series data object is created, the graphic function `plot( )` in R is used to draw the plot of the mid cap sector time series during the period January 2010 - December 2016. Figure 1 depicts the graph of the mid cap sector time series.

![Figure 1: Time series of the mid cap sector index in India (Period: Jan 2010 – Dec 2016)](image)

In order to gain deeper insights into the characteristics of the time series, the time series object is decomposed into its three components: trend, seasonal and random. For decomposing the time series object, the `decompose( )` function is invoked that is defined in the TTR library in R programming environment. The `decompose( )` function is invoked with the mid cap time series object as the actual parameter of the function. Hence, the three components of the time series are obtained. Figure 2 depicts
the plots of mid cap sector time series and its three components. It may be noted that Figure 2 consists of four boxes arranged in the form of a stack. The boxes display the plot of the overall time series, the trend, the seasonal and the random component respectively arranged from top to bottom in that order.

From Figure 1, it may be seen that the time series of the mid cap sector experienced a rise during the period January 2010 - October 2010. During the period November 2010 - December 2011, the time series decreased from a value of 7764 to 5135. It started rising again in January 2012 from a value of 5872 and attained a value of 7113 in December 2012 before it started falling again. The time series exhibited a fall during the period January 2013 till September 2013, decreasing from a value of 6971 to a value of 5606. The period October 2013 to December 2015, the time series experienced a consistent rise from a value of 6117 to a value of 11143. The months of January and February 2016 witnessed a fall in the time series value. However, the time series consistently increased in its value during the period March 2016 till December 2016 from an initial value of 10619 attaining a value of 12031. Figure 2 shows the decomposition results of the mid cap time series. The three components of the time series are depicted separately so that their relative behavior can be effectively observed.

![Figure 2: Decomposition results of the mid cap sector index time series into its three components (Period: Jan 2010 – Dec 2016)](image)

Table 1 lists the numerical values of the time series index values and its three components. It may be observed that the values for the trend and the random components are unavailable for the period January 2010 – June 2010 and also for the period July 2016 – December 2016. The reason for this data unavailability stems from the fact that the computation of trend requires long term data. Coughlan (2015) shows that the computations in the `decompose()` function in R is based on a 12-month moving average method for determining the trend values in a time series. Thus, for computation of trend values for the period January 2010 – June 2010, the `decompose()` function requires time series information for the period July 2009 – December 2009. As the data for the period July 2009 – December 2009 are not available in the dataset that we have used, accordingly it was not possible to compute the trend values for the period January 2010 – June 2010. In the same line of logic, the trend values for the period July 2016 - December 2016, could not be determined due to non-availability of the time series information for the period January 2017 – June 2017 in our dataset. It is observed from Table 1 that the seasonal value for a
given month remains unchanged throughout the entire time horizon of analysis. For example, corresponding to the month of January, the seasonal component remains constant at a value of 22 for every year during the period January 2010 - December 2016. It is also worth observing that because of the unavailability of the trend values for the period January 2010 – June 2010 and July 2016 – December 2016, the values of the random components are also missing. The unavailability of the random component values is due to the following reason. The aggregate time series values are given by the sum of the corresponding trend, seasonal and random component values, and because of the fact that the seasonal value for a given month remains the same throughout the horizon of analysis, if the trend values are missing for a certain period, the random components values also become indeterminable for the same period.

Table 1: Time series index values of the Indian mid cap sector and its components
(Period: Jan 2010 – Dec 2016)

| Year | Month   | Aggregate Index | Trend  | Seasonal | Random |
|------|---------|-----------------|--------|----------|--------|
| 2010 | January | 6510            | -102   | -381     | -171   |
|      | February| 6398            | -318   | -171     | -102   |
|      | March   | 6806            | -102   | -102     | -381   |
|      | April   | 7185            | -96    | -96      | -171   |
|      | May     | 6835            | 39     | -96      | -171   |
|      | June    | 7149            | 142    | 39       | -96    |
|      | July    | 7408            | 100    | 142      | 39     |
|      | August  | 7597            | -138   | 100      | 142    |
|      | September| 8084           | 72     | -138     | 142    |
|      | October | 8303            | 239    | 72       | 142    |
|      | November| 7764            | 239    | 239      | 142    |
|      | December| 7803            | 239    | 239      | 142    |
| 2011 | January | 6868            | 7307   | -102     | -337   |
|      | February| 6373            | 7231   | -381     | -477   |
|      | March   | 6873            | 7095   | -171     | -50    |
|      | April   | 7094            | 6930   | -96      | 260    |
|      | May     | 6910            | 6757   | 39       | 114    |
|      | June    | 6854            | 6557   | 142      | 155    |
|      | July    | 6915            | 6405   | 100      | 411    |
|      | August  | 6274            | 6364   | -138     | 48     |
|      | September| 6130           | 6342   | 72       | -284   |
|      | October | 6298            | 6288   | 239      | -229   |
|      | November| 5628            | 6214   | 147      | -733   |
|      | December| 5135            | 6143   | 149      | -1157  |
| 2012 | January | 5872            | 6076   | -102     | -102   |
|      | February| 6387            | 6027   | -381     | 741    |
|      | March   | 6346            | 6036   | -171     | 482    |
|      | April   | 6316            | 6067   | -96      | 345    |
|      | May     | 5908            | 6131   | 39       | -262   |
|      | June    | 6154            | 6267   | 142      | -255   |
|      | July    | 6012            | 6395   | 100      | -482   |
|      | August  | 6005            | 6437   | -138     | -294   |
|      | September| 6607           | 6425   | 72       | 111    |
|      | October | 6566            | 6418   | 239      | -91    |
|      | November| 6902            | 6439   | 147      | 316    |
|      | December| 7113            | 6451   | 149      | 512    |
|        | January | February | March | April | May | June | July | August | September | October | November | December |
|--------|---------|----------|-------|-------|-----|------|------|--------|-----------|---------|----------|----------|
| 2013   | 6971    | 6424     | -102  | 649   |     |      |      |        |           |         |          |          |
|        | 6303    | 6375     | -381  | 309   |     |      |      |        |           |         |          |          |
|        | 6142    | 6304     | -171  | 10    |     |      |      |        |           |         |          |          |
|        | 6344    | 6243     | -96   | 197   |     |      |      |        |           |         |          |          |
|        | 6389    | 6200     | 39    | 151   |     |      |      |        |           |         |          |          |
|        | 5964    | 6159     | 142   | -336  |     |      |      |        |           |         |          |          |
|        | 5543    | 6114     | 100   | -671  |     |      |      |        |           |         |          |          |
|        | 5300    | 6095     | -138  | -657  |     |      |      |        |           |         |          |          |
|        | 5606    | 6142     | 72    | -608  |     |      |      |        |           |         |          |          |
|        | 6107    | 6222     | 239   | -354  |     |      |      |        |           |         |          |          |
|        | 6326    | 6350     | 147   | -171  |     |      |      |        |           |         |          |          |
|        | 6706    | 6578     | 149   | -22   |     |      |      |        |           |         |          |          |
|        |         |          |       |       |     |      |      |        |           |         |          |          |
| 2014   | 6308    | 6873     | -102  | -463  |     |      |      |        |           |         |          |          |
|        | 6500    | 7191     | -381  | -309  |     |      |      |        |           |         |          |          |
|        | 7083    | 7521     | -171  | -267  |     |      |      |        |           |         |          |          |
|        | 7323    | 7840     | -96   | -421  |     |      |      |        |           |         |          |          |
|        | 8467    | 8160     | 39    | 269   |     |      |      |        |           |         |          |          |
|        | 9379    | 8477     | 142   | 760   |     |      |      |        |           |         |          |          |
|        | 9188    | 8814     | 100   | 274   |     |      |      |        |           |         |          |          |
|        | 9299    | 9178     | -138  | 258   |     |      |      |        |           |         |          |          |
|        | 9530    | 9504     | 72    | -46   |     |      |      |        |           |         |          |          |
|        | 9835    | 9779     | 239   | -184  |     |      |      |        |           |         |          |          |
|        | 10271   | 10002    | 147   | 121   |     |      |      |        |           |         |          |          |
|        | 10373   | 10150    | 149   | 73    |     |      |      |        |           |         |          |          |
| 2015   | 10739   | 10291    | -102  | 549   |     |      |      |        |           |         |          |          |
|        | 10811   | 10438    | -381  | 755   |     |      |      |        |           |         |          |          |
|        | 10592   | 10550    | -171  | 213   |     |      |      |        |           |         |          |          |
|        | 10416   | 10651    | -96   | -139  |     |      |      |        |           |         |          |          |
|        | 10716   | 10728    | 39    | -51   |     |      |      |        |           |         |          |          |
|        | 10680   | 10790    | 142   | -252  |     |      |      |        |           |         |          |          |
|        | 11273   | 10809    | 100   | 364   |     |      |      |        |           |         |          |          |
|        | 10734   | 10744    | -138  | 128   |     |      |      |        |           |         |          |          |
|        | 10799   | 10694    | 72    | 34    |     |      |      |        |           |         |          |          |
|        | 10975   | 10721    | 239   | 14    |     |      |      |        |           |         |          |          |
|        | 10990   | 10774    | 147   | 68    |     |      |      |        |           |         |          |          |
|        | 11143   | 10844    | 149   | 149   |     |      |      |        |           |         |          |          |
| 2016   | 10417   | 10945    | -102  | -427  |     |      |      |        |           |         |          |          |
|        | 9575    | 11107    | -381  | -1150 |     |      |      |        |           |         |          |          |
|        | 10619   | 11309    | -171  | -519  |     |      |      |        |           |         |          |          |
|        | 11043   | 11512    | -96   | -373  |     |      |      |        |           |         |          |          |
|        | 11366   | 11679    | 39    | -351  |     |      |      |        |           |         |          |          |
|        | 11717   | 11778    | 142   | -203  |     |      |      |        |           |         |          |          |
|        | 12661   | 100      |       |       |     |      |      |        |           |         |          |          |
|        | 13217   | -138     |       |       |     |      |      |        |           |         |          |          |
|        | 13167   | 72       |       |       |     |      |      |        |           |         |          |          |
|        | 13473   | 239      |       |       |     |      |      |        |           |         |          |          |
|        | 12499   | 147      |       |       |     |      |      |        |           |         |          |          |
|        | 12031   | 149      |       |       |     |      |      |        |           |         |          |          |

From Table 1, some important observations are made. First, we can see that seasonal component has the maximum value of 239 in the month of October, while the lowest value of seasonality -381 is observed in the month of February. The seasonal component is found to have positive values during the months of May, June, July, September, October, November and December. The months of January, February,
March, April, and August are found to experience negative values for their seasonality components. In order to analyze the impact of seasonality on the mid cap time series, we computed some statistics of the seasonal component values. We computed the percentage contribution of the seasonal component on the aggregate time series values and found the following. The maximum, the minimum, the mean of the absolute values, and the standard deviation of the percentage of the seasonal components with respect to the aggregate time series values were found to be 3.91, -6.04, 1.94 and 2.35 respectively. The maximum percentage of seasonal component was found in the month of November 2013, while the minimum was observed in the month of February 2013. While the low value of the mean percentage indicated that the time series was not seasonal, the high value of standard deviation in comparison to the mean value implied that the seasonal percentages exhibited high level of dispersion among themselves.

Second, the trend component of the time series exhibited a sluggish nature with a slow rate of fall during the period July 2010 till February 2012. However, the trend started increasing at a very slow rate during the period March 2012 till August 2012. Since September 2012 the trend started falling slowly again and that small rate of decrease continued till August 2013. The trend consistently increased, however, during September 2013 till June 2016. The maximum, the minimum, the mean, and the standard deviation of the percentage of the trend component with respect to the aggregate time series were found to be 119.63, 88.52, 100.69 and 6.25 respectively indicating that the trend was the single most dominant component in the time series. The maximum percentage of trend component was found in the month of December 2011, while the minimum was found in the month of October 2010.

Third, the maximum and the minimum values of the random component of the time series were found to be 760 and -1157 respectively. These values are quite modest in comparison to the aggregate time series values. In order to understand the contribution of the random component on the overall time series, we computed the maximum, the minimum, the mean of the absolute values and the standard deviation of the percentage of random component values with respect to the aggregate time series values. These values were found to be 11.60, 22.53, 4.59 and 6.04 respectively. It indicated that while the mean random component is not dominant in the time series, occasionally the series exhibited high randomness. The high value of standard deviation with respect to the mean value of the percentage of random component values implied that the random component exhibited large deviations across their mean value. The random component contributed its maximum percentage to the aggregate time series in the month of February 2012, while the lowest percentage was found in the month of December 2011.

The overall conclusion is that the mid cap time series is primarily dominated by its trend component, while seasonal and random components are having not significant contributions to the aggregate time series. However, the series exhibited high randomness occasionally, and the seasonal and the random components showed significant variations across their mean values.

4. Proposed Forecasting Methods

This section presents a collection of forecasting approaches that can be effectively applied on the mid cap sector time series data. Six different methods are proposed for predicting the future time series values of the mid cap sector and their performances are analyzed. For comparing the efficiency and accuracy of the proposed forecasting methods, six different metrics are identified. Based on the values of these metrics, we identify method leads to lowest value of the forecasting error. The forecasting methods are also critically analyzed and reasons are identified for why one particular method performs better than the others on the given dataset of mid cap sector time series for the period January 2010 – December 2016. In
the following, we first briefly describe the six forecasting methods, and then in Section 5 we present the detailed results of their performance in forecasting on the mid cap sector dataset.

**Method I**: In this method, we use the mid cap sector time series data for the period January 2010 till December 2015 for the purpose of forecasting the monthly index values for each month of the year 2016. The `HoltWinters()` function defined in the in R library `forecast` has been used for the purpose of forecasting. In order to account for both the trend and the seasonal component in the forecasted value, the `HoltWinters` function is used with a changing trend and an additive seasonal component as two parameters in the function call. The forecast horizon in the `HoltWinters` model is chosen to be 12 so that the forecast for all months of the year 2016 can be made by using this method at the end of the year 2015. Forecast error for each month of 2016 and an overall RMSE value is also computed.

**Method II**: Unlike Method I that uses forecast horizon of 12 months, this approach forecasts the mid cap sector monthly time series values for the year 2016 using the `HoltWinters()` function with a forecast horizon of 1 month. For example, in order to forecast the time series value for the month of March 2016, the time series data for the period January 2010 - February 2016 are used to develop the predictive model. Similar to Method I, the `HoltWinters` function is used with a changing trend and an additive seasonal component parameters. Since the forecast horizon is short, it is expected that the model will produce higher accuracy in forecasting most often in comparison to Method I that uses a much longer forecast horizon of 12 months. The forecast error corresponding to each month of 2016 and an overall RMSE value for the model is computed.

**Method III**: In this forecasting approach, using the time series data for the mid cap sector index for the period January 2010 - December 2015, we first compute the values of its trend and seasonal components. The computation produces the values of the trend component of the time series for the period July 2010 - June 2015. We now use the computed trend values for predicting the trend values for the period July 2015 - June 2016 using the `HoltWinters()` function in R with a changing trend component level but no seasonal component. In other words, in order to forecast the trend values, in the `HoltWinters()` function in R, we set the parameter `beta` = TRUE and the parameter `gamma` = FALSE. The forecasted trend values are now added to the seasonal component values (computed based on the time series data for the period January 2010 - December 2015) of the corresponding months so as to arrive at the forecasted aggregate of the trend and seasonal components. Next, the time series of the entire period, i.e., from January 2010 till December 2016 is considered. We decompose it into its three components, and compute the aggregate of the trend and the seasonal component values for the period July 2015 - June 2016. The forecasting accuracy of this method is computed by finding the percentage of deviation of the aggregate of the actual trend and the actual seasonal component values with respect to the corresponding aggregate values of the forecasted trend and past seasonal components for each month for the period July 2015 - June 2016. The RMSE and the percentage of the RMSE with respect to the mean aggregate of the trend and the seasonal component is also computed for this method as explained in Section 5.

**Method IV**: The method follows the same approach as Method III except for the way in which the trend values are forecasted. In contrast to Method III that uses `HoltWinters()` function with a changing trend component and a zero seasonal component, Method IV builds a linear regression model for forecasting the trend component values for the period July 2015 - June 2016. The `lm()` function in R is used for building the bivariate linear model using the "trend component" as the response variable and "time" as the predictor variable. The regression model is built using the trend values for the period July 2010 - June 2015 as the training data set. As in Method III, the aggregate of the predicted trend values and the past
seasonal values are compared with the aggregate of the actual trend values and the actual seasonal values for the period July 2015 - June 106. The error in forecasting and an overall RMSE value is also computed.

Method V: This method is based on an Auto Regressive Integrated Moving Average (ARIMA) approach to forecasting. The ARIMA forecasting model is built using the mid cap sector time series data for the period January - December 2015. The \texttt{auto.arima()} function defined in the forecast package in R is invoked with the above time series data as the parameter of the function for finding the values of the three parameters of the Auto Regressive Moving Average (ARMA) model, i.e. the auto regression parameter \((p)\), the difference parameter \((d)\), and the moving average parameter \((q)\). Next, the ARIMA model of forecasting is built by invoking the \texttt{arima()} function in R with its two parameters as: (i) the mid cap time series R object (based on data for the period January 2010 - December 2015), (ii) the order of the ARMA i.e., the values of the three parameters \((p, d, q)\). With the resultant ARIMA model, we execute the function \texttt{forecast.Arima()} in the forecast library with the parameters as: (i) ARIMA model object, and (ii) forecast horizon = 12 months. Since the method is based on a forecast horizon of 12 months, it forecasts the time series values for all the twelve months of the year 2016, the forecasting being made at the end of the year 2015. The error in forecasting for each month and the overall RMSE value are also computed.

Method VI: Similar to Method V, this forecasting method is also based on an ARIMA model. However, in contrast to Method V that is based on a forecast horizon of 12 months, Method VI uses a shorter forecast horizon of 1 month. The ARIMA forecasting model is built using time series data for the period January 2010 till the month previous to the month for which forecasting is being made. For example, in order to predict the time series index for the month of May 2016, the time series data from January 2010 till April 2016 is used as a parameter to the \texttt{auto.arima()} function. Since the training data set for building the ARIMA model constantly changes in each iteration of the model building, the ARIMA parameters (i.e., \(p, d,\) and \(q\)) are evaluated accordingly before every round of forecasting. In other words, for each month of the year 2016, before he forecast for the next month is being made, the parameters of the ARIMA model are computed from the scratch.

5. Forecasting Results

In this section, we present detailed results on the forecasting accuracies of the six methods that we discussed in Section 4.

Method I: Table 2 lists the forecasted values of the time series for each month of the year 2016 and the corresponding error in forecasting. Figure 3 shows the plot of the actual values of the time series index and their corresponding predicted values for each month of the year 2016.

Observations on Method I: It is clearly evident from Table 2 that the forecasted values are very close in magnitude to the actual values of the mid cap sector time series for large majority of the months in the year 2016. The error percentage value exceeded the threshold value of 10 only for three months - July, August and September. The month of August yielded the highest value of the error percentage of 15.78, while the month of December produced the smallest error in forecasting, the magnitude of the error being 0.81 percent. The reason for such high accuracy in results for this method can be attributed to the slow rate of growth of the mid cap sector during major parts of the year 2016. Since there was no abrupt rise or fall in the time series for most of the months of 2016, the HoltWinters method of forecasting using a long prediction horizon of 12 months that effectively forecasted a smoothened average value turned out to be very accurate. The RMSE value for the forecasted results is found out to be 979, which is only 8.29
This small percentage clearly indicates that Method I is quite effective in forecasting the time series.

Table 2: Computation Results using Method I of forecasting

| Month | Actual Index | Forecasted Index | % Error | RMSE |
|-------|--------------|------------------|---------|------|
| Jan   | 10417        | 10711            | 2.82    |      |
| Feb   | 9575         | 10253            | 7.08    |      |
| Mar   | 10619        | 10329            | 2.73    |      |
| Apr   | 11043        | 10676            | 3.32    |      |
| May   | 11366        | 11104            | 2.31    |      |
| Jun   | 11717        | 11263            | 3.87    |      |
| Jul   | 12661        | 11260            | 11.07   |      |
| Aug   | 13217        | 11131            | 15.78   |      |
| Sep   | 13167        | 11658            | 11.46   |      |
| Oct   | 13473        | 12146            | 9.85    |      |
| Nov   | 12499        | 12175            | 2.59    |      |
| Dec   | 12031        | 12129            | 0.81    |      |

Figure 3. Actual and predicted values of the mid cap sector index using Method I of forecasting (Period: Jan 2016 – Dec 2016)

**Method II:** In Table 3, we have listed the results of forecasting for the mid cap sector time series for the year 2016 using Method II. Figure 4 shows the plot of the actual time series values and their corresponding forecasted values.

**Observations on Method II:** The observations in Table 3 and the plot in Figure 4 clearly elicit the fact that the forecasted values very closely match with the actual values of the time series for all the months of the year 2016. The forecast for the month of May 2016 yielded the smallest error, the value being 0.13 percent. The highest value of error was observed in the forecast for the month of March 2016 - the...
magnitude of the error being 8.36 percent. Computations have produced the RMSE value for this method as 584 and the mean value of the mid cap time series in 2016 as 11815. Thus, the RMSE value is only 4.94 percent of the mean value of the actual index. This evidently validates the fact that the HoltWinters additive model with a small forecast horizon of 1 month very effectively models and accurately forecast the mid cap sector time series index values.

Table 3: Computation Results using Method II of forecasting

| Month | Actual Index (A) | Forecasted Index (B) | % Error \(\frac{(C-B)}{B} \times 100\) | RMSE |
|-------|------------------|----------------------|----------------------------------------|------|
| Jan   | 10417            | 10711                | 2.82                                   |      |
| Feb   | 9575             | 9994                 | 4.38                                   |      |
| Mar   | 10619            | 9731                 | 8.36                                   |      |
| Apr   | 11043            | 10969                | 0.67                                   |      |
| May   | 11366            | 11381                | 0.13                                   | 584  |
| Jun   | 11717            | 11535                | 1.55                                   |      |
| Jul   | 12661            | 11667                | 7.85                                   |      |
| Aug   | 13217            | 12550                | 5.05                                   |      |
| Sep   | 13167            | 13833                | 5.06                                   |      |
| Oct   | 13473            | 13738                | 1.97                                   |      |
| Nov   | 12499            | 13390                | 7.13                                   |      |
| Dec   | 12031            | 12538                | 4.21                                   |      |

Figure 4. Actual and predicted values of the mid cap sector index using Method II of forecasting (Period: Jan 2016 – Dec 2016)

Method III: Table 4 presents the results of forecasting produced by this method. The actual index values and their corresponding predicted values for all months of the year 2016 using this method are presented in Figure 5. The actual monthly values of the trend and the seasonal component for the period from July 2015 - June 2016 (computed based on the time series data for the period January 2010 - December 2016) and their aggregate are noted in Columns B, C and D respectively in Table 4. The monthly forecasted trend values (computed using HoltWinters method with a trend component and no seasonal component and having a forecast horizon of 12 months), the past seasonal component values (based on time series
data for the period from January 2010 - December 2015), and their aggregate values are listed in columns $E$, $F$ and $G$ respectively in Table 4. The error in forecast for each month and an overall RMSE value for this method are also computed and listed Table 4.

Table 4: Computation Results using Method III of forecasting

| Month | Actual Trend | Actual Seasonal | Actual (Trend + Seasonal) | Forecasted Trend | Past Seasonal | Forecasted (Trend + Seasonal) | % Error | RMSE |
|-------|--------------|----------------|---------------------------|------------------|---------------|-------------------------------|---------|------|
| A     | 10809        | 100            | 10909                     | 10852            | -11           | 10841                         | 0.62    | 230  |
| Jul   | 10744        | -138           | 10606                     | 10914            | -201          | 10713                         | 1.01    |      |
| Aug   | 10794        | 72             | 10766                     | 10976            | 27            | 11003                         | 2.20    |      |
| Sep   | 10721        | 239            | 10960                     | 11038            | 199           | 11237                         | 2.53    |      |
| Oct   | 10774        | 147            | 10921                     | 11100            | 96            | 11196                         | 2.52    |      |
| Nov   | 10844        | 149            | 10993                     | 11162            | 82            | 11244                         | 2.28    |      |
| Dec   | 10945        | -102           | 10843                     | 11224            | -54           | 11170                         | 3.02    |      |
| Jan   | 11107        | -381           | 10726                     | 11286            | -189          | 11097                         | 3.46    |      |
| Feb   | 11107        | -381           | 10726                     | 11286            | -189          | 11097                         | 3.46    |      |
| Mar   | 11309        | -171           | 11138                     | 11348            | -106          | 11242                         | 0.93    |      |
| Apr   | 11512        | -96            | 11416                     | 11410            | -59           | 11351                         | 0.57    |      |
| May   | 11679        | 39             | 11718                     | 11472            | 71            | 11543                         | 1.49    |      |
| Jun   | 11778        | 142            | 11920                     | 11534            | 145           | 11679                         | 2.02    |      |

Observations on Method III: The margin of error for all months during the period July 2015 - June 2016 as produced by Method III of forecasting have been almost negligible. This can be easily verified from Table 4 and Figure 5. The forecast for the month of April 2016 yielded the lowest value of error with a magnitude of 0.57 percent, while the month of February 2016 produced the highest value of error percentage of 3.46. The computation of the RMSE yields a value of 230, while the mean value of the sum of the actual trend and the actual seasonal components during the period July 2015 - June 2016 is 11076. It implies that the RMSE is only 2.08 percent of the mean value of the sum of the actual trend and the
actual seasonal component. This makes it amply clear that Method III is very effective in forecasting the mid cap time series values. Let us understand why Method III is so accurate. It is clear from Table 4 that the actual trend values exhibited very sluggish growth over the period July 2015 - June 2016. The actual seasonal values being smaller in magnitude in comparison to their trend counterparts, the sum of the actual trend and the actual seasonal values exhibited the same pattern as that of the trend. In fact, the sum of the actual trend and the actual seasonal values increased slowly from a value of 10909 in July 2015 to 11920 in June 2016. Since the trend values are predicted using HoltWinters method with a long forecast horizon of 12 months, the forecasted values are smoothed out over the period. This resulted in a forecasted series of trend that exhibited a very slow rate of increase - the value of forecasted trend series increased from 10852 in July 2015 to 11534 in June 2016. Since the past seasonal values computed on the mid cap sector time series for the period January 2010 - December 2015 were much smaller in magnitude in comparison to the corresponding trend values, the sum of the forecasted trend values and the respective past seasonal values also increased very sluggishly exhibiting the same pattern as the forecasted trend time series values. It is the slow rate of increase for both the actual series and the forecasted series that made the patterns of the two series identical. This, in turn, has resulted into a highly of accuracy in forecasting for Method III.

Table 5: Computation Results using Method IV of forecasting

| Month | Actual Trend | Actual Seasonal | Actual (Trend + Seasonal) | Forecasted Trend | Past Seasonal | Forecasted (Trend + Seasonal) | % Error | RMSE |
|-------|--------------|----------------|--------------------------|------------------|--------------|-------------------------------|---------|------|
|       | A            | B              | C                        | D                | E            | F                             |         |      |
| Jun   | 11778        | 142            | 11920                    | 9633             | 145          | 9778                          | 17.97   | 1468 |
| May   | 11679        | 39             | 11718                    | 9578             | 71           | 9649                          | 17.14   | 1469 |
| Apr   | 11512        | -56            | 11416                    | 9523             | -59          | 9464                          | 17.10   | 1469 |
| Mar   | 11309        | -171           | 11138                    | 9468             | -106         | 9362                          | 15.95   | 1473 |
| Feb   | 11107        | -381           | 10726                    | 9413             | -189         | 9224                          | 14.00   | 1473 |
| Jan   | 10945        | -102           | 10843                    | 9358             | 54           | 9412                          | 13.20   | 1473 |
| Dec   | 10844        | 149            | 10993                    | 9303             | 82           | 9385                          | 14.63   | 1473 |
| Nov   | 10774        | 147            | 10921                    | 9248             | 96           | 9344                          | 14.44   | 1473 |
| Oct   | 10721        | 239            | 10960                    | 9193             | 199          | 9392                          | 14.31   | 1473 |
| Sep   | 10694        | 72             | 10766                    | 9138             | 27           | 9165                          | 14.87   | 1473 |
| Aug   | 10744        | -138           | 10606                    | 9083             | -201         | 8882                          | 16.25   | 1473 |
| Jul   | 10809        | 100            | 10909                    | 9028             | 11           | 9039                          | 17.14   | 1473 |

Method IV: We present the results of forecasting for Method IV in Table 5. The actual index values and their corresponding predicted values for all months during the period July 2015 - June 2016 are depicted in Figure 6. The values of the actual trend component and the actual seasonal component for the period July 2015 - June 2016 are computed using the time series data for the period January 2010 - December 2016. Columns B, C and D in Table 4 list the actual trend, the actual seasonal and their sum respectively for the period July 2015 - June 2016. In Method III, the trend values were forecasted using the HoltWinters( ) function. However, in Method IV, we forecast the trend values using a linear regression approach with the mid cap index values as the response variable and the month as the predictor variable. The predicted values of the trend component, the past seasonal component values (based on time series data for the period January 2010 - December 2015) and their corresponding sum are listed in columns E, F and G respectively in Table 5. The percentage of error in forecasting for each month during the period July 2015 till June 2016, and an overall RMSE value are also listed.
Observation on Method IV: Results presented in Table 5 and Figure 6 make it evident that unlike Method III, Method IV yields high error values. The month of January 2016 produced the lowest error per cent value of 13.20, while the highest error per cent of 17.97 was observed in the forecasted value corresponding to the month of June 2016. It is also observed in Figure 5 that the error margin increased with time. Computations have yielded the RMSE value for this method as 1749, while the mean value of the actual sum of trend and seasonal components during the period July 2015 - June 2016 is found to be 11076. Hence, the RMSE is found to be 15.79 percent of the mean value of the index. This indicates that
the method of forecasting is effective although its error margin is higher than those produced by some of the methods that we have discussed before. It is not difficult to understand why this method has produced higher error margins. The reason is as follows. We have used a linear regression model to forecast the trend values for the period July 2015 - June 2016. Based on the historical trend values of the time series for the period July 2010 - June 2015, the linear regression model produced the values of the regression parameters, and used those values to determine the trend values of the time series for the period July 2015 - June 2106. Since, the trend values for the period July 2010 - June 2015 exhibited a very small positive slope, the value of the regression parameter for the time variable is of relatively smaller magnitude in comparison to the value of the intercept. The regression model yielded an intercept value of 5673 and the value of the coefficient for time as 55. The smaller value of the regression coefficient for the time variable ensured that the rate of increase of the forecasted trend values was very slow which is evident from the fact that the forecasted trend values started with an initial value of 9028 for the month of July 2015, increased only to a value of 9633 in the month of June 2016. Since the magnitudes of the seasonal component computed on the basis of the time series for the period January 2010 - December 2015 (referred to as "past seasonal values" in Table 5) were small compared to the trend values, the sum of the forecasted trend and the past seasonal values followed the same pattern as the forecasted trend values. This is validated from the fact that the sum of the forecasted trend and the past seasonal values increased very slowly from a value of 9039 in July 2015 to 9778 in June 2016. However, since the actual trend values of the mid cap time series during the target period of July 2015 - June 2016 were higher in magnitude due to a rise in the mid cap time series during the year 2016, the actual trend values were consistently higher than the forecasted trend values. In fact, starting with a value of 10809 in July 2015, the actual trend attained a value of 11778 in June 2016. This difference in the values of the actual trend and forecasted trend adversely affected the accuracy of Method III.

![Figure 8. Plot of the partial auto correlation function (PACF) of the mid cap sector time series with max lag of 2 years (Period: Jan 2016 – Dec 2016)](image-url)
Method V: This method of forecasting utilizes an ARIMA technique with a forecast horizon of 12 months. We apply the `auto.arima()` function defined in the library `forecast` in R, on the mid cap sector time series index values for the period January 2010 till December 2015. The application of `auto.arima()` function on the mid cap time series data yields the parameter values for the time series as: $p = 0$, $d = 1$, and $q = 1$. The values of $p$, $d$, and $q$ are also cross-verified by plotting the *partial auto correlation function* (PACF), the *first-order difference*, and the *auto correlation function* (ACF) for the time series respectively. Figure 7 depicts the first order difference of the mid cap sector time series. It is evident that the first-order difference of the mid cap sector time series is a stationary one, as its mean and variance are approximately time-invariant. Thus the value of $d = 1$ is cross-verified to be correct. Next, the *partial auto correlation function* (PACF) and the *auto correlation function* (ACF) are plotted to cross-check the values of the parameters $p$ and $q$ respectively. The PACF of the mid cap sector time series if depicted in Figure 8. It is observed that except for lag = 0, the partial correlation values at all lags are statistically insignificant. Hence the value of $p = 0$ is also cross-verified. We observe in Figure 9 that the minimum integral value of lag beyond which all autocorrelation values are statistically insignificant is 1. Therefore, the value of the parameter $q = 1$ is also cross-verified. Hence, we confirm that the mid cap sector time series for the period January 2010 - December 2015 can be modeled as an ARMA $(0, 1, 1)$ model. We build the ARIMA model using the `arima()` function in R with two parameters as: (i) the mid cap sector time series object and (ii) the order $(0, 1, 1)$ of ARMA. Finally, for the purpose of forecasting the time series index values for all the twelve month of the year 2016, the function `forecast.Arima()` is used with two parameters as: (i) the ARIMA model and (ii) the time horizon of forecast = 12 months. Table 6 presents the results of forecasting, while Figure 10 depicts the actual index values and their corresponding predicted values for the mid cap sector corresponding to each month of the year 2016.
Table 6: Computation Results using Method V of forecasting

| Month | Actual Index | Forecasted Index | % Error | RMSE |
|-------|--------------|------------------|---------|------|
|       | (A)          | (B)              | (C)     |      |
| Jan   | 10417        | 11184            | 7.36    |      |
| Feb   | 9575         | 11184            | 16.80   |      |
| Mar   | 10619        | 11184            | 5.32    |      |
| Apr   | 11043        | 11184            | 1.28    |      |
| May   | 11366        | 11184            | 1.60    |      |
| Jun   | 11717        | 11184            | 4.55    |      |
| Jul   | 12661        | 11184            | 11.67   |      |
| Aug   | 13217        | 11184            | 15.38   |      |
| Sep   | 13167        | 11184            | 15.06   |      |
| Oct   | 13473        | 11184            | 16.99   |      |
| Nov   | 12499        | 11184            | 10.52   |      |
| Dec   | 12031        | 11184            | 7.04    |      |

**Figure 10.** Actual and predicted values of the mid cap sector index using Method V of forecasting (Period: Jan 2016 – Dec 2016)

**Observations on Method V:** It is evident from Table 6 that the error percentage values are moderate but higher than those produced in some of the methods discussed earlier. In fact, the error percentage values are found to be high for the period July 2016 to November 2016. The forecast for month of April 2016 yielded the lowest error percentage value of 1.28. The highest value of the error percentage of 15.38 was found corresponding to the month of August 2016. We computed the RMSE value for this method and found it to be 1347. The RMSE value was 11.40 per cent of the mean value of the mid cap sector index during the period January 2016 - December 2016, since the mean value of the mid cap sector index was 11815. When we consider the fact that this method uses a long forecast horizon of 12 months, the error percentage values are moderate. However, as mentioned earlier, the error margins are found to be higher during the period July 2016 to November 2016. The reason for such high error is that the mid cap time series experienced a sudden jump in its value from 11717 in June 2016 to 12661 in July 2016. For Method V that uses a long forecast horizon of 12 months, it was impossible to predict this behavior in
December 2015 when the forecasted values were actually computed. The increasing trend of the time series continued till October 2016 and this caused the error margin to increase till October 2016. However, the error started falling in November 2016 as the time series experienced a fall and that behavior continued in the month of December 2016 as well. Figure 10 presents the actual index values and their corresponding predicted values using Method V. It may be noted that the forecasted values remained constant throughout the entire period as the underlying ARIMA model is (0,1,1).

Table 7: Computation Results using Method VI

| Month | Actual Index | Forecasted Index | % Error | RMSE |
|-------|--------------|------------------|---------|------|
| Jan   | 10417        | 11184            | 7.36    |      |
| Feb   | 9575         | 10404            | 8.66    |      |
| Mar   | 10619        | 9388             | 11.59   |      |
| Apr   | 11043        | 10619            | 3.84    | 685  |
| May   | 11366        | 10913            | 3.99    |      |
| Jun   | 11717        | 11441            | 2.36    |      |
| Jul   | 12661        | 11693            | 7.65    |      |
| Aug   | 13217        | 12907            | 2.35    |      |
| Sep   | 13167        | 13306            | 1.06    |      |
| Oct   | 13473        | 13124            | 2.59    |      |
| Nov   | 12499        | 13578            | 8.63    |      |
| Dec   | 12031        | 12222            | 1.59    |      |

Figure 11. Actual and predicted values of the mid cap sector index using Method VI of forecasting (Period: Jan 2016 – Dec 2016)

Method VI: In this approach, we build an ARIMA model using a forecast horizon of one month. The approach that we follow for building the ARIMA model, however, is almost identical to that used in Method V. The difference in Method V and Method VI lies in different values of forecast horizon used in
these methods. While Method V uses a forecast horizon of 12 months, we use a forecast horizon of 1 month in Method VI. Since, in Method VI, we make forecast only one month in advance, the training data set used for building the ARIMA model consistently increases by one record in each iteration, and hence, we re-evaluate the parameters of the ARIMA model after every iteration of the ARIMA forecasting. Stating in other words, in each month of 2015, before we make the forecast for the next month, we compute the values of the three parameters of the ARIMA model. The computation of the values of ARIMA parameters $p$, $d$, and $q$ revealed that for the months of January, March, June and August - December 2016, the ARIMA model was $(0, 1, 1)$. While the months of February, May and July yielded an ARIMA model of $(1, 1, 1)$, the month of April exhibited an $(0, 1, 0)$ ARIMA model. Table 7 presents the forecasting results for Method VI. Figure 11 depicts the actual index values of the mid cap sector and their corresponding predicted values for this method of forecasting.

**Observations on Method VI:** From Table 7, it is evident that the percentage of error values for all months of the year 2016 are quite moderate. The month of September 2016 produced the lowest value of the error percentage which was found to be 1.06, while the month of March 2016 yielded the highest value of error - 11.59 per cent. The RMSE value for Method VI of forecasting, after computation, was found to be 685. The mean value of the index of the mid cap sector for the period January 2016 to December 2016 was 11815. Hence, it can be easily verified that the RMSE value is 5.80 per cent of the mean value of the actual index of the mid cap sector. This indicates that Method VI has been highly accurate in forecasting the mid cap sector index values. The high level of accuracy of Method VI may be attributed to its short forecast horizon of one month. The short forecast horizon is able to catch the changing pattern of the time series very effectively. This has resulted into small error in forecasting. It is clearly evident from Figure 11 that the forecasted time series values very closely followed the pattern of the time series of the actual index values of the mid cap sector.

**Summary of Forecasting Results**

In Table 8, we summarize the performance of the six forecasting methods that we have used. For the purpose of comparing these six different approaches, we have used six metrics: (i) minimum (Min) error in per cent, (ii) maximum (Max) error in per cent, (iii) mean error in per cent, (iv) standard deviation (SD) of error in per cent, (v) root mean square error (RMSE), and (vi) the ratio of RMSE to the mean of the actual index values in per cent. It may be noted that for Method I, II, IV and V the mean of the index values are identical, which is the mean of the actual index values of the mid cap sector for the twelve months in the year 2016. This mean value is computed and found to be equal to 11815. For Method III and IV, however, the mean of the index values are the mean of the sum of the actual trend and seasonal values during the period July 2015 till June 2016. The mean value of the sum of the trend and seasonal values is found to be equal to 11076. Since the RMSE values for Methods I, II, IV and V and those for Method III and IV are computed based on different raw data set, instead of the raw RMSE value, the percentage of RMSE to the mean value of the index serves as a better metric for comparing different methods of forecasting. Accordingly, we have ranked the forecasting methods based on this metrics only.

Table 8 presents the comparative analysis of the six forecasting methods based on the six metrics that we have already mentioned. It may be seen that Method III that computes the aggregate of the forecasted trend values using \textit{HoltWinters( )} function of horizon 12 months and the past seasonal values to predict
the aggregate of the future trends values and the new seasonal values, has yielded the lowest forecasting error. This method has produced the lowest percentage value of the ratio of RMSE to the mean index value. In fact, Method III has produced lowest values for all other metrics as well except for the metric min error rate. Hence, Method III clearly turns out to be the best method among all the six approaches. On the contrary, Method IV that forecasts the aggregate of the trend and the seasonal component values using a linear regression on the trend values and then summing up the old seasonal values with the forecasted trend values to predict the aggregate of the future trend and new seasonal values has performed worst. Method IV has yielded the highest values of errors corresponding to four metrics: ratio of RMSE to the mean value, min errors, max error and mean error. In terms of their forecasting accuracy, Method II, Method VI and Method I have secured the second, the third and the fourth rank respectively yielding RMSE to mean index values of 4.94, 5.80 and 8.29 percent respectively. While Method V performed worse than all other methods except Method IV, its performance also has been quite satisfactory with RMSE to mean index value of 11.40 per cent.

Table 8: Comparison of the performance of the forecasting methods

| Methods  | Min Error | Max Error | Mean Error | SD of Errors | RMSE  | RMSE / Mean Index Value Percentage |
|----------|-----------|-----------|------------|--------------|-------|-----------------------------------|
| Method I | 0.81      | 15.78     | 6.14       | 4.78         | 979   | 8.29                              |
| Method II| 0.13      | 8.36      | 4.10       | 2.76         | 584   | 4.94                              |
| Method III| 0.57     | 3.46      | 1.89       | 0.95         | 230   | 2.08                              |
| Method IV| 13.2      | 17.97     | 15.63      | 1.59         | 1749  | 15.79                             |
| Method V | 1.28      | 16.99     | 9.46       | 5.75         | 1347  | 11.40                             |
| Method VI| 1.06      | 11.59     | 5.14       | 3.46         | 685   | 5.80                              |

6. Related Work

In the literature, researchers have proposed several approaches and techniques for forecasting the daily prices of stocks and index values of different sectors of economy. Neural network-based approaches are most popular among these propositions. Mostafa (2010) presented a method of forecasting the movement of stock prices in Kuwait that utilized the concepts of artificial neural networks (ANNs). Kimoto et al. (1990) utilized neural networks that used historical accounting data and various macroeconomic factors to predict movement patterns in several stock returns. Leigh et al. (2005) illustrated various approaches to prediction of stock prices and stock market index movements in the New York Stock Exchange (NYSE) during the period 1981 – 1999. The authors used linear regression and simple neural network models for in developing their proposed approaches. Hammad et al. (2009) demonstrated how the output of an ANN model can be forced to converge after the execution of a finite number of iterations and then produce highly accurate predicted values of stock prices. Dutta et al. (2006) designed predictive models using ANNs for the purpose of forecasting of the closing index values the BSE during the period January 2002 till December 2003. Tsai & Wang (2009) carried out a study that investigated how and why the forecasting accuracy produced by a BN-based approach usually is higher than that yielded by a traditional regression and neural network-based method. Trseng et al. (2012) illustrated how techniques like
traditional time series decomposition, *HoltWinters* models, Box-Jenkins method and artificial neural networks can be applied in predicting the prices of a randomly selected set of 50 stocks during the period September 1998 till December 2010. The authors found that while the forecasting errors were small for Box-Jenkins method, *HoltWinters* model and normalized neural network model, higher values of error were observed in time series decomposition method and non-normalized neural network model. Moshiri & Cameron (2010) proposed a *back propagation network* (BPN) with econometric models for predicting the inflation level of economy using the following techniques: (i) Box-Jenkins Autoregressive Integrated Moving Average (BVARMA) model, (ii) Vector Autoregressive (VAR) model and (ii) Bayesian Vector Autoregressive (BVAR) model. Thenmozhi (2001) demonstrated how the chaos theory can be applied for identifying the pattern of changes of stock prices in Bombay Stock Exchange (BSE) during the period August 1980 till September 1997. The author discovered that the daily and weekly returns of the BSE index exhibited nonlinear trends, while the BSE index showed a movement pattern that was weakly chaotic. Hutchinson et al. (1994) proposed a novel approach based on the *principle of learning networks* for accurately predicting the price of a derivative.

Predictive models built on ANN architecture and algorithms are found most effective in discovering the non-linear nature of the stock price movements and hence these methods are found to be most accurate in forecasting stock prices. Shen et al. (2007); Jaruszewicz & Mandziuk (2004); Ning et al. (2009); Pan et al. (2005); Hamid & Iqbal (2004); Chen et al. (2005); Chen et al. (2003); Haniaa et al. (2007); de Faria et al. (2009); Zhu et al. (2008)) illustrated the efficacy and effectiveness of ANN-based models in their forecasting ability of stock price movements. Many applications of hybrid systems in stock market time series data analysis have also been proposed in the literature. Wu et al. (2008), Wang & Nie (2008), Perez-Rodriguez et al. (2005), Leung et al. (2000) and Kim (2004) proposed applications of hybrid systems in stock price prediction.

In contrast to the work mentioned above, our approach in this paper is based on structural decomposition of time series of the mid cap sector index in India during the period January 2010 till December 2016. Based on the decomposition of the time series, we identified several important features and characteristics of the mid cap sector India. In particular, we investigated the nature of the trend, seasonality pattern and degree of randomness of the time series. After analyzing the nature of the mid cap time series, we presented six forecasting techniques for predicting the index values of the sector for every months of the year 2016. We computed the accuracies of each of the forecasting techniques, and critically analyzed under what situations a particular technique performs better than the other techniques.

### 7. Conclusion

In this paper, we have presented a time series decomposition-based approach for analyzing the salient characteristics of the time series of the mid cap sector of the Indian economy during the period January 2010 till December 2016. The rich features of the R programming language in data management, analytics and visualization have been used to decompose the time series index values of the Indian mid cap sector into its three components - trend, seasonal, and random. Several interesting characteristics of the time series were revealed by the decomposition results. Using the results of decomposition of the mid cap sector time series, we computed the amount of seasonality and randomness that were found in the time series. In particular, we could identify the months of in which the time series exhibited its strongest seasonality and the months during which the seasonal component was the weakest. The month of September was found to exhibit the highest level of seasonality for the mid cap time series, while the
month of February experienced the lowest effect of seasonality during the period of our study. The time series was found to be consisted of a very moderate random component with the mean value of the percentage of the random component to the aggregate value of the time series being as low as 2.4. The trend was the most dominant component in the time series. The value of the trend component increased at a rather slow rate over the seven year period under our study, i.e., during January 2010 to December 2016. After analyzing the characteristics of the mid cap sector time series based on the time series decomposition results, we present six methods for forecasting the future index values of the time series. The six forecasting methods were based on different algorithms and used different lengths of the forecasting horizons. Among the six methods, Method III was found to have produced the highest level of accuracy in the forecasting. This method used the sum of the forecasted trend values using \textit{HoltWinters( )} function of horizon 12 months and the past seasonal values to predict the sum of the future trends values and the new seasonal values. However, Method IV yielded the maximum error and the lowest accuracy in forecasting. Method IV attempted to predict the trend values using a linear regression model and based on the predicted trend values and the past seasonal values it forecasted the sum of the actual trend values and the current seasonal values. This method produced the highest value of the ratio of the RMSE to the mean index value, and hence performed worst among all the six forecasting methods. The forecasting methods that used \textit{HoltWinters} algorithm with forecast horizons of 12 months and 1 month and the method that used ARIMA technique with a forecast horizon of 1 month have also performed quite satisfactorily with none of these methods producing the RMSE to mean index ratio more than 4.7. However, the method that used ARIMA technique with a forecast horizon of 12 months yielded a relatively higher value of 6.83 percent for the RMSE to mean index value.

While the results in this work have yielded valuable insights into the characteristics of the mid cap index time series in India, these results also serve as guidelines for choosing an appropriate forecasting framework for predicting the future index values of the time series. Moreover, the observations in the time series decomposition can be very useful in forming an optimized portfolio of stocks, if similar such exercise are performed on different sectors of the economy. With the knowledge of the time of highest and lowest seasonality, the degree of randomness, and the behavior of the trends of different sectors, it is possible to identify suitable time for buy/sell decisions of the stocks belonging to different sectors. and With a robust decision making in buy/sell of stocks and their appropriate timing it is possible to design an efficient and optimized portfolio.

References

Chen, A.-S., Leung, M. T. & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan stock index. \textit{Operations Research in Emerging Economics}, 30(6), 901–923. DOI: 10.1016/S0305-0548(02)00037-0

Chen, Y., Dong, X. & Zhao, Y. (2005). Stock index modeling using EDA based local linear wavelet neural network. \textit{Proceedings of International Conference on Neural Networks and Brain}, Beijing, China, pp. 1646–1650. DOI: 10.1109/ICNNB.2005.1614946

Coghlan, A. (2015). \textit{A Little Book of R for Time Series}, Release 02. Available at: \texttt{Link} (Accessed on: August 30, 2017)
de Faria, E. L., Albuquerque, M. P., Gonzalez, J. L., Cavalcante, J. T.P., & Albuquerque, M. P. (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. *Expert Systems with Applications*, 36(10), 12506-12509. DOI: 10.1016/j.eswa.2009.04.032

Dutta, G., Jha, P., Laha, A. & Mohan, N. (2006). Artificial neural network models for forecasting stock price index in the Bombay Stock Exchange. *Journal of Emerging Market Finance*, 5(3), 283-295. DOI: 10.1177/097265270600500305

Hamid, S. A., Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 index futures prices. *Journal of Business Research*, 57(10), 1116–1125. DOI: 10.1016/S0148-2963(03)00043-2

Hammad, A. A. A., Ali, S. M. A. & Hall, E. L. (2007). Forecasting the Jordanian stock price using artificial neural network. *Intelligent Engineering Systems through Artificial Neural Networks*, Vol 17, Digital Collection of The American Society of Mechanical Engineers. DOI: 10.1115/1.802655.paper42

Hanias, M., Curtis, P. & Thalassinos, J. (2007). Prediction with neural networks: the Athens stock exchange price indicator. *European Journal of Economics, Finance and Administrative Sciences*, 9, 21–27. Available at: Link (Accessed on: August 30, 2017)

Hutchinson, J. M., Lo, A. W., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, 49(3), 851-889. DOI: 10.3386/w4718

Ihaka, R. & Gentleman, R. (1996). A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299 – 314. DOI: 10.2307/1390807

Jaruszewicz, M. & Mandziuk, J. (2004). One day prediction of NIKKEI index considering information from other stock markets. *Proceedings of the International Conference on Artificial Intelligence and Soft Computing*, 3070, 1130–1135. DOI: 10.1007/978-3-540-24844-6_177

Kim, K.-J. (2004). Artificial neural networks with feature transformation based on domain knowledge for the prediction of stock index futures. *Intelligent Systems in Accounting, Finance & Management*, 12(3), 167–176. DOI: 10.1002/isaf.252

Kimoto, T., Asakawa, K., Yoda, M. & Takeoka, M. (1990). Stock market prediction system with modular neural networks. *Proceedings of the IEEE International Conference on Neural Networks*, San Diego, pp. 1 – 16, CA, USA,. DOI: 10.1109/IJCNN.1990.137535

Leigh, W., Hightower, R. and Modani, N. (2005). Forecasting the New York Stock Exchange composite index with past price and interest rate on condition of volume spike. *Expert Systems with Applications*, 28(1), 1-8. DOI: 10.1016/j.eswa.2004.08.001

Leung, M. T., Daouk, H. & Chen, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting*, 16(2), 173–190. DOI: 10.1016/S0169-2070(99)00048-5

Moshiri, S. & Cameron, N. (2010). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, 19(3), 201-217. DOI: 10.1002/(SICI)1099-131X(200004)19:3<201::AID-FOR753>3.0.CO;2-4

Mostafa, M. (2010). Forecasting stock exchange movements using neural networks: empirical evidence from Kuwait. *Expert Systems with Application*, 37(9), 6302-6309. DOI: 10.1016/j.eswa.2010.02.091

Ning, B., Wu, J., Peng, H. & Zhao, J. (2009). Using chaotic neural network to forecast stock index. *Advances in Neural Networks*, Lecture Notes in Computer Science, Vol 5551, pp. 870–876 Springer-Verlag, Heidelberg, Germany. DOI: 10.1007/978-3-642-01507-6_98
Pan, H., Tilakaratne, C. & Yearwood, J. (2005). Predicting the Australian stock market index using neural networks exploiting dynamical swings and intermarket influences. *Journal of Research and Practice in Information Technology*, 37(1), 43–55. DOI: 10.1007/978-3-540-89378-3_53

Perez-Rodriguez, J. V., Torra, S. & Andrada-Felix, J. (2005). Star and ANN models: forecasting performance on the Spanish IBEX-35 stock index. *Journal of Empirical Finance*, 12(3), 490–509. DOI: 10.1016/j.jempfin.2004.03.001

Sen J. & Datta Chaudhuri, T. (2016a). Decomposition of time series data of stock markets and its implications for prediction – an application for the Indian auto sector. *Proceedings of the 2nd National Conference on Advances in Business Research and Management Practices (ABRMP’16)*, pp. 15 – 28, Kolkata, India, January, 2016. DOI: 10.13140/RG.2.1.3232.0241.

Sen, J. & Datta Chaudhuri, T. (2016b). A framework for predictive analysis of stock market indices – a study of the Indian auto sector. *Calcutta Business School (CBS) Journal of Management Practices*, 2(2), 1-20. DOI: 10.13140/RG.2.1.2178.3448.

Sen, J. & Datta Chaudhuri, T. (2016c). An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice: a comparative study of the Indian consumer durable and small cap sectors. *Journal of Economic Library*, 3(2), 303-326. DOI: 10.1453/jel.v3i2.787.

Sen, J. & Datta Chaudhuri, T. (2016d). An investigation of the structural characteristics of the Indian IT sector and the capital goods sector – an application of the R programming in time series decomposition and forecasting. *Journal of Insurance and Financial Management*, 1(4), 68-132. Available at: Link (Accessed on: August 30, 2017)

Sen, J. & Datta Chaudhuri, T. (2016e). Decomposition of time series data to check consistency between fund style and actual fund composition of mutual funds. *Proceedings of the 4th International Conference on Business Analytics and Intelligence (ICBAI 2016)*, Bangalore, India, December 19-21. DOI: 10.13140/RG.2.2.14152.93443.

Sen, J. & Datta Chaudhuri, T. (2017a). A time series analysis-based forecasting framework for the Indian healthcare sector. *Journal of Insurance and Financial Management*, 3(1), 66 - 94. Available at: Link (Accessed on: August 30, 2017)

Sen, J. & Datta Chaudhuri, T. (2017b). A predictive analysis of the Indian FMCG sector using time series decomposition-based approach. *Journal of Economic Library*, 4(2), 206 - 226. DOI: 10.1453/jel.v4i2.1282.

Sen, J. (2017a). A time series analysis-based forecasting approach for the Indian realty sector. *International Journal of Applied Economic Studies*, 5(4), 8 - 27. Available at: Link (Accessed on: August 30, 2017)

Sen, J. (2017b). A study of the Indian metal sector using time series decomposition-based approach. Book Chapter No 8 in: *Analysis and Forecasting of Financial Time Series Using R: Models and Applications*, Sen, J. & Datta Chaudhuri, T., pp. 223- 255, August 2017, Scholars' Press, Germany. ISBN: 978-3-330-65386-3.

Shen, J., Fan, H. & Chang, S. (2017e). Stock index prediction based on adaptive training and pruning algorithm. *Advances in Neural Networks*, Lecture Notes in Computer Science, Vol 4492, pp. 457–464, Springer-Verlag, Heidelberg, Germany. DOI: 10.1007/978-3-540-72393-6_55

Thenmozhi, M. (2006). Forecasting stock index numbers using neural networks. *Delhi Business Review*, 7(2), 59-69. Available online at: Link (Accessed on: August 30, 2017)

Tsai, C.-F. & Wang, S.-P. (2009). Stock price forecasting by hybrid machine learning techniques. *Proceedings of International Multi Conference of Engineers and Computer Scientists*, pp. 755 – 765, Hong Kong, March 2009. Available at: Link (Accessed on: August 30, 2017)
Tseng, K.-C., Kwon, O., & Tjung, L.C. (2012). Time series and neural network forecast of daily stock prices. *Investment Management and Financial Innovations*, 9(1), 32-54. Available at: Link (Accessed on: August 30, 2017)

Wang, W. & Nie, S. (2008). The performance of several combining forecasts for stock index. *International Seminar on Future Information Technology and Management Engineering*, pp. 450–455, Leicestershire, United Kingdom, November 2008. DOI: 10.1109/FITME.2008.42

Wu, Q., Chen, Y. & Liu, Z. (2008). Ensemble model of intelligent paradigms for stock market forecasting. *Proceedings of the 1st International Workshop on Knowledge Discovery and Data Mining*, pp. 205 – 208, Washington, DC, USA, January 2008. DOI: 10.1109/WKDD.2008.54

Zhu, X., Wang, H., Xu, L. & Li, H. (2008). Predicting stock index increments by neural networks: the role of trading volume under different horizons. *Expert Systems with Applications*, 34(4), 3043–3054. DOI: 10.1016/j.eswa.2007.06.023