Bull. Grad. School Educ. Hiroshima Univ., Part Ⅱ, No. 56, 2007, 31－37

Modelling and Forecasting Monthly Highest SENSEX Values

Sanpei Kageyama, Subhabaha Pal[1], Satyabrata Pal[2]

(Received September 28, 2007)

Abstract. First compiled in 1986, SENSEX (Sensitive Index) is a basket of 30 constituent stocks representing a sample of large, liquid and representative companies. The base year of SENSEX is 1978–79 and the base value is 100. The index is widely reported in both domestic and international markets through print as well as electronic media everyday. SENSEX had touched the 10000 landmark in February 2006 and has come within the first three indices in the world in terms of rapid up-down movement. So far as the authors are aware, appropriate powerful models to generate accurate forecasts of SENSEX values are not available in the literature. In this paper the SENSEX monthly highest time-series data from the period, January 1999 to December 2006, have been used to model the monthly highest values of the SENSEX and to generate future forecasts for the subsequent twelve months till December 2007. The special features of this model ([ARIMA$(1,1,6) \times (1,0,3)_{12}$] selected from among fifty competitive models) are that the forecasts are very close to the future values realized later in January 2007. It had predicted that the SENSEX would touch 14000 landmark in the month of January 2007 which, indeed, was the scene. So it can be expected that the remaining forecasts on monthly highest values will also be identical with those SENSEX values representing the real situations destined to prevail in future. The most salient feature is that the R^2 value is found to be 0.983 which indicates that the prediction power of the model is very high, though only 96 monthly SENSEX values have been used for prediction and 3 monthly SENSEX values have been reserved for determination of the accuracy of the forecast model.

Keywords: SENSEX, ARIMA modelling, seasonal model.

1. Introduction

Bombay Stock Exchange Limited is the oldest stock exchange in Asia with a rich heritage. Popularly known as "BSE", it was established as "The Native Share & Stock Brokers Association" in 1875. It is the first stock exchange in India. The Exchange's pivotal and pre-eminent role in the development of the Indian capital market is widely recognized and its index, SENSEX, is tracked worldwide. SENSEX is calculated using the "Free-float Market Capitalization" methodology (BSE Website–SENSEX). With the advent of globalization, capital markets have started playing a prominent role in the economy of all countries. The market capitalization of each company in a free-float index is reduced to the extent of its readily available shares in the market. Appropriate powerful models to generate accurate forecasts of SENSEX values are not available in the literature. In February 2006 the SENSEX touched the 10000 landmark for the first time in the twenty-six year history of SENSEX, which fetched mixed reaction from different sectors of Indian economy and the government.

[1]Post Graduate Student, University of Calcutta, W.B., Pin-700019, India
[2]Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, W.B., Pin-741252, India

Free-float methodology refers to an index construction methodology that takes into consideration only the free-float market capitalization of a company for the purpose of index calculation and assigning weight to stocks in index. Free-float market capitalization is defined as that proportion of total shares issued by the company that are readily available for trading in the market. It generally excludes promoters' holding, government holding, strategic holding and other locked-in shares that will not come to the market for trading in the normal course. In other words, the market capitalization of each company in a free-float index is reduced to the extent of its readily available shares in the market.

The auto-regressive integrated moving average (ARIMA) approach was first popularized by Box et al. (1994), and is often referred to as Box-Jenkins models. In this paper, a seasonal ARIMA modelling technique is applied on the long-term monthly highest SENSEX data for the last six years. The data corresponding to the last six years have been employed because during the last six years, India has been experiencing great changes in economy due to the advent of globalization. Before 1999, the behaviour of SENSEX was more or less stable and with little variation over time during the previous 19 years and as such the future values could be predicted with little effort using simple mathematical models. But after 1999, the daily SENSEX values started showing significant up-down movement in very short intervals of time bringing in immense uncertainty about the situation of the capital market in these later years. Here an attempt has been made to estimate the sudden up-down behaviour of SENSEX with considerable success with the help of auto-regressive (AR) models. An ARIMA model predicts a value in a response time series as a linear combination of its own past values, past errors (also called shocks or innovations), and current and past values of other time series. This procedure (with seasonal adjustment) is applied on the monthly highest SENSEX data and it is found that the generated model yields a fairly good fit (R^2 value, 0.983, defined later) and also the future monthly highest SENSEX forecasts, which are close to the actually obtained monthly highest SENSEX values in later months. The most notable feature in the fitted model is that the predicted value of the monthly highest SENSEX also touches 14000 in the month of January 2007 which is the actual situation so far seen. Another notable feature in the model is that the model predicts the SENSEX to touch 15000 landmark in the month of April 2007.

Remark: SENSEX is a financial index and its value changes with financial market condition with sharp rise and fall within a very short span of time. Within a short span of a month, SENSEX figures take different values at different instants of time and here only the highest value reached by SENSEX figures over a period of one month is considered for modelling and forecasting.

2. Data set and objectives

We use the SENSEX monthly highest time-series data from the period, January 1999 to December 2006 (collected from BSE Website). The objectives of the paper are to model the monthly highest values of SENSEX, and to generate future forecasts for the next twelve months (future SENSEX values, excepting the first three, yet to come in reality), and to judge the competence of the model to be used for forecasting purpose (also to compare its forecasting power corresponding to the months, January – March 2007, the values of which are available but not used in building up of the model).

3. Method

The order of an ARIMA model is usually denoted by the notation ARIMA(p, d, q), where p is the order of the AR part, d is the order of the differencing, q is the order of the moving-average (MA) process. If no differencing is done, i.e., $d = 0$, the model is usually referred to an ARMA(p, q) model. Mathematically, the pure ARIMA model is written as

$$w_t = \mu + \frac{\theta(B)}{\phi(B)}\alpha_t,$$

where t indexes time, w_t is the response series or a difference of the response series y_t, μ is the mean term, B is the backshift operator, that is, $Bx_t = x_{t-1}$, $\phi(B)$ is the AR operator, represented as a polynomial in the backshift operator, $\phi(B) = 1-\phi_1 B-\cdots-\phi_p B^p$, $\theta(B)$ is the MA operator, represented as a polynomial in the backshift operator, $\theta(B) = 1-\theta_1 B-\cdots-\theta_q B^q$, α_t is the independent disturbance, also called the random error .

For simple (non-seasonal) differencing $w_t = (1-B)^d y_t$ and for seasonal differencing $w_t = (1-B)^d(1-B^s)^D y_t$, where d is the degree of non-seasonal differencing, D is the degree of seasonal differencing, and s is the length of the seasonal cycle. For example, the mathematical form of the ARIMA(1,1,1) model is given by

$$(1-B)y_t = \mu + \frac{1-\theta_1(B)}{1-\phi_1(B)}\alpha_t.$$

Alternatively, the ARIMA model can also be written as $\phi(B)(w_t-\mu) = \theta(B)\alpha_t$ or $\phi(B)w_t = const + \theta(B)\alpha_t$, where $const = \phi(B)\mu = \mu - \phi_1\mu - \phi_2\mu - \cdots - \phi_p\mu$, since here $B^i\mu = \mu$ for $1 \le i \le p$.

Thus, when an AR operator and a mean term are both included in the model, the constant term for the model can be represented as $\phi(B)\mu$. This model expresses the response series as a combination of past values of random shocks and past values of other input series. The response series and input time series are also called the *dependent series* or *output series* and *independent series* or a *predicto series*, respectively. Box et al. (1994) and Pankratz (1983) discuss the ARIMA procedure.

Seasonal models: ARIMA models for time series with regular seasonal fluctuations often use differencing operators and AR and MA parameters at lags that are multiples of the length of the seasonal cycle. When all the terms in an ARIMA model factor refer to lags that are a multiple of a constant s, the constant is factored out and suffixed to the ARIMA(p, d, q) notation. Thus, the general notation for the order of a seasonal ARIMA model with both seasonal and non-seasonal factors is ARIMA$(p, d, q) \times (P, D, Q)_s$. The term (p, d, q) gives the order of the non-seasonal part of the ARIMA model and the term $(P, D, Q)_s$ gives the order of the seasonal part. The value of s is the number of observations in a seasonal cycle: 12 for monthly series, 4 for quaterly series, 7 for daily series with day-of-week effects, and so on. (See also Section 4.)

For example, the notation ARIMA$(0,1,2)\times(0,1,1)_{12}$ describes a seasonal ARIMA model for monthly data with the following mathematical form:

$$(1-B)(1-B^{12})y_t = \mu + (1-\theta_{1,1}B - \theta_{1,2}B^2)(1-\theta_{2,1}B^{12})\alpha_t,$$

or

$$(1-B)(y_t - y_{t-12}) = \mu + (1-\theta_{1,1}B - \theta_{1,2}B^2)(\alpha_t - \theta_{2,1}\alpha_{t-12}),$$

where $\theta_{1,1}$ is a coefficient in ARIMA model (non-seasonal MA 1 with respect to B), $\theta_{1,2}$ is a coefficient in ARIMA model (non-seasonal MA 1 with respect to B^2) and $\theta_{2,1}$ is a coefficient in ARIMA model (seasonal MA 1 with respect to B^{12}). Note that for a seasonal MA with 12 as the subscript $B^{12}y_t = y_{t-12}$.

If y_t represent the set of observations, then the total sum of squares for the series, SST, corrected for the mean is given by

$$SST = \sum_{t=1}^{n}(y_t - \bar{y})^2.$$

The sum of the squared prediction errors, SSE, is the one-step predicted value and it is given by

$$SSE = \sum_{t=1}^{n}(y_t - \widehat{y_t})^2,$$

where $\widehat{y_t}$ is the predicted value at the t-th time point. The R^2 statistic is given by $R^2 = 1 - SSE/SST$. The mean absolute prediction error is given by

$$\frac{1}{n}\sum_{t=1}^{n}|y_t - \widehat{y_t}|.$$

The mean absolute percent prediction error (MAPE) is shown as

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \widehat{y_t}}{y_t}\right|,$$

where the summation ignores observations when $y_t = 0$.

4. Results

A good number of AR and ARIMA models have been exploited and the best one with the required features has been presented. The best seasonal ARIMA model [ARIMA$(1,1,6) \times (1,0,3)_{12}$] is fitted on the SENSEX data and the findings are listed in Table 1 (the seasonality component $(1,0,3)_{12}$ has been introduced in the model to increase the precision of fitting, i.e., see the value of $R^2 = 0.983149$ later).

Table 1

Parameters of fitted ARIMA	Estimate
MA, Lag 1	0.7812
MA, Lag 2	0.1063
MA, Lag 3	0.1150
MA, Lag 4	0.0130
MA, Lag 5	−0.0011
MA, Lag 6	−0.2460
Seasonal MA, Lag 12	−0.5820
Seasonal MA, Lag 24	−0.3984
Seasonal MA, Lag 36	0.1114
AR, Lag 1	0.9532
Seasonal AR, Lag 12	−0.6567

Table 2 gives the predicted and forecast values along with upper and lower limits. Predicted values are given from January 2004 and future monthly highest SENSEX forecasts are given for the period, January 2007 to December 2007. However, the calculation is done based on the monthly highest SENSEX values from January 1999 to December 2006.

Table 2

Date	Actual	Predicted	Upper Limit	Lower Limit	Error
Jan 2004	6194.11	5965.39	6693.50	5237.28	228.72
Feb 2004	6035.80	6378.36	7101.67	5655.06	−342.56
Mar 2004	5935.19	6138.05	6860.27	5415.83	−202.86
Apr 2004	5925.58	6079.90	6801.92	5357.88	−154.32
May 2004	5757.30	6057.13	6779.15	5335.11	−299.83
Jun 2004	4963.75	5969.14	6691.15	5247.14	−1005.39
Jul 2004	5170.32	5191.70	5913.68	4469.71	−21.38
Aug 2004	5252.78	5321.85	6043.35	4600.35	−69.07
Sep 2004	5616.87	5486.92	6208.41	4765.43	129.95
Oct 2004	5776.85	5847.12	6568.59	5125.65	−70.27
Nov 2004	6234.29	5871.82	6593.29	5150.36	362.47
Dec 2004	6602.69	6184.17	6905.62	5462.72	418.52
Jan 2005	6679.20	6649.62	7371.00	5928.24	29.58
Feb 2005	6713.86	6689.85	7408.79	5970.92	24.01
Mar 2005	6915.09	6525.15	7243.88	5806.42	389.95
Apr 2005	6606.41	6817.40	7536.12	6098.69	−210.99
May 2005	6715.11	6269.04	6987.75	5550.32	446.07
Jun 2005	7193.85	7107.09	7825.80	6388.38	86.76
Jul 2005	7635.42	7320.30	8038.99	6601.60	315.13
Aug 2005	7859.53	7883.58	8602.04	7165.12	−24.05
Sep 2005	8650.17	8097.68	8816.03	7379.33	552.49
Oct 2005	8799.96	8921.83	9640.08	8203.58	−121.87
Nov 2005	8994.94	8938.05	9656.18	8219.92	56.89
Dec 2005	9397.93	9377.89	10095.91	8659.87	20.04
Jan 2006	9919.89	9703.85	10421.79	8985.91	216.04
Feb 2006	10370.24	9956.22	10674.15	9238.28	414.02
Mar 2006	11307.04	10712.50	11430.39	9994.61	594.54
Apr 2006	12042.56	11673.94	12391.82	10956.07	368.62
May 2006	12612.38	12493.84	13211.71	11775.97	118.54
Jun 2006	10609.25	12272.46	12990.33	11554.59	−1663.21
Jul 2006	10930.09	10495.49	11213.36	9777.62	434.60
Aug 2006	11723.92	11091.44	11809.31	10373.58	632.48
Sep 2006	12454.42	12360.37	13078.13	11642.61	94.05
Oct 2006	13024.26	13014.42	13732.07	12296.76	9.84
Nov 2006	13773.59	13749.87	14467.44	13032.3	23.72
Dec 2006	13972.03	13740.32	14457.85	13022.78	231.71
Jan 2007	14282.72	14059.32	14774.80	13343.83	223.40
Feb 2007		14457.60	15559.93	13355.28	
Mar 2007		14858.46	16268.85	13448.08	
Apr 2007		15007.98	16647.94	13368.02	
May 2007		15357.42	17176.24	13538.61	
Jun 2007		16329.45	18292.81	14366.10	

Table 2 (continued)

Date	Actual	Predicted	Upper Limit	Lower Limit	Error
Jul 2007		16784.18	18931.82	14636.54	
Aug 2007		17089.99	19457.04	14722.93	
Sep 2007		17551.09	20167.33	14934.85	
Oct 2007		17801.84	20691.86	14911.82	
Nov 2007		17921.92	21105.79	14738.05	
Dec 2007		18242.29	21736.28	14748.30	

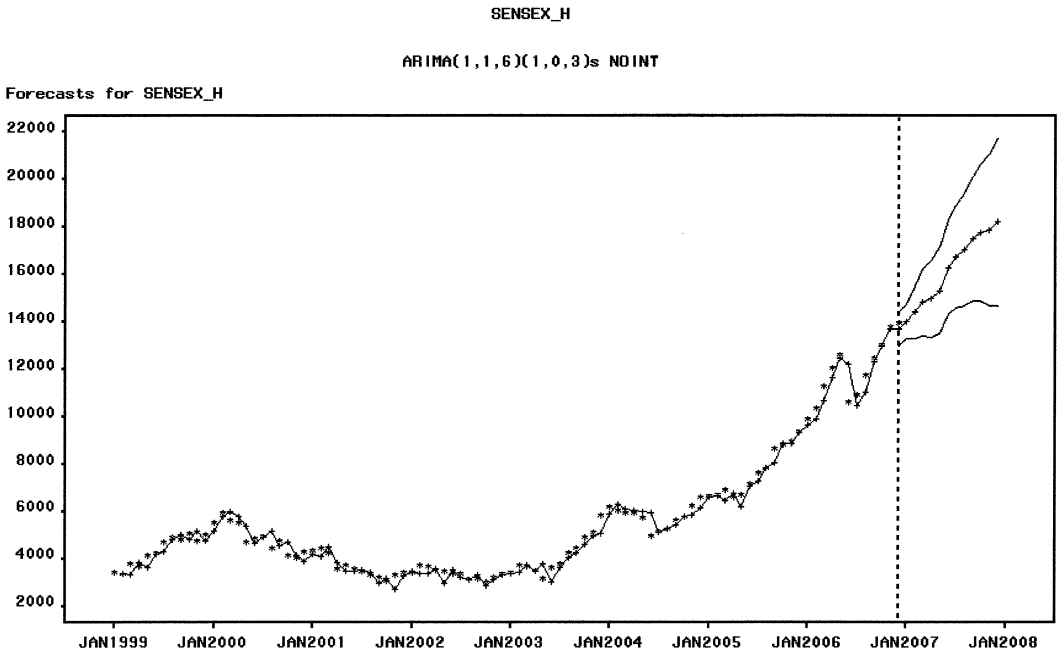
Results of the fit are as follows.

Label	Value
MAPE	5.05681
R^2	0.983149
$(MSE)^{1/2}$	359.6482

Note that here MSE is the square root of Mean Square Error and $MSE = (1/n)SSE$.

5. Some significant observations

The special features of the generated model are as follows. From the forecast values, we find that the monthly highest of January 2007 will reach the 14000 landmark and in reality also it has reached the 14000 landmark in the monthly of January 2007. Among other notable feature, the model predicts the SENSEX to reach the 15000 landmark in the month of April 2007 and 16000 landmark in the month of June 2007.



Figure

ARIMA model fitted to the monthly highest SENSEX values and the graph of forecast values by use of the package SAS. Here the model fitted is $\text{ARIMA}(1,1,6) \times (1,0,3)_{12}$ as can be seen from the parameter chart. The two thin (unspotted) lines correspond to the upper and lower boundaries of the estimate.

6. Conclusion

This paper presents a very useful approach for modelling and forecasting SENSEX data. The forecast values obtained by using the methodology follow closely the actual values (for January 2007). Seasonal ARIMA procedure with necessary transformations adopted in this paper can be well applied to the daily SENSEX data also in order to get useful forecasts on the behaviour of the SENSEX for the coming days. Using this procedure on the hourly SENSEX values in an appropriate manner, the hourly highest values of the SENSEX can also be predicted with precision for the coming hours in any day. Up and down of SENSEX values manifest the condition of the capital market at that moment. If SENSEX values are going up then it shows that the capital market is in good condition, and so then the share-brokers start purchasing shares and if suddenly the SENSEX values start falling then the reverse thing happens, i.e., seeing uncertainty in the capital market the share-brokers start selling shares to take guard from future losses. Hence if they can get any idea about the behaviour of the SENSEX at any time-point it will be of immense use to them and accordingly they can plan their investment. The methodology for forecasting stock indices shown here can also be well used in case of stock exchanges of other countries as well, like Hong Kong, USA, Japan, which shows rapid variation in very short period of time.

References

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Times Series Analysis: Forecasting and Control.* Third Edition. Englewood Cliffs, N. J., Prentice Hall.

Pankratz, Alan (1983). *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases.* John Wiley, New York.