Prediction of Stock Market Indices – Using SAS

B.Siddhartha Reddy
Research Facilitation Center
IBS Bangalore
Bangalore, Karnataka (state), INDIA
E-mail:Siddhartha.burri@gmail.com

Abstract —**The SAS© System has a powerful suite of tools for analyzing and forecasting data taken over a selected time period. The paper concentrates more on Stock Market (NSE-Nifty, India) & its prediction, by and large a risky venture. Knowledgeable investors base their predictions either on the basis of Fundamental Analysis, or Technical Analysis, or both. But most of the investors rely on the tips given by the experts for Stock Market Predictions. However there are many such models available such as Interrupted Time Series, Auto Regression (AR), Exponential Smoothening, Moving Average (MA), and Distributed Lags Analysis. The procedures FORECAST, ARIMA process will be illustrated.**

Keywords - forecasting, prediction, stock market prediction, Time series analysis.

I. INTRODUCTION

Stock market is the market for securities where organized issuance and trading of Stocks takes place either through exchanges or over-the-counter in electronic or physical form. It plays an important role in channelizing capital from the investors to the business houses, which consequently leads to the availability of funds for business expansion.

Determination of stock price considers that market is supreme and it discounts everything (economical, political and all related factors). It presumes that all the investors behave rationally and the value of the asset is estimated based on future expectations. Hence, with every new information, the future expectation of the market is liable to change and consequently the stock prices**.** As the new information is erratic in nature so it influences the price in a random way.

Stock Market Investments are subjected to market risks and the returns are variable in nature because it is an uncertain existence. Hence Stock Market Prediction is not also a guaranteed one. It is based on certain analyses namely as shown below provided there are still other analyses available such as traditional time series analysis and machine learning methods.

A. Why Analyze?

Stock Market Investments are subjected to market risks and the returns are variable in nature because it is an uncertain existence. Hence Stock Market Prediction is not a guaranteed one. It is based on certain analyses linked to the Macro & Micro economic factors or variables such as demand & supply. There are certain other analyses available such as fundamental analysis, traditional time series analysis and machine learning methods.

The analyzing and prediction of the indices is in one line-To reap returns while investing on the index derivatives, *Index derivatives* provide investors the exposure to price movements of entire indices through a single futures or options contract.

Using index options, a very interesting kind of "portfolio insurance" can be obtained, whereby an investor gets paid only if the market index drops. If one does not want to bear index fluctuations in the coming weeks, then the index futures or index options can be used to reduce (or even eliminate) the consequent index exposure. This is far more convenient than distress selling of the underlying equity in the portfolio. Thus, such prediction of indices would help the investor to play safe in the dynamic volatile market arena.

B. Analyzing What?

Analyzing the indices as in NIFTY MIDCAP 50, one can determine various trends the market has seen on day-to-day basis-the process which includes determining the trends underneath the data series (at a stretch of 13 years from 1997-2010) brought down from various secondary sources. Moreover, these trends can be used to identify the patterns in the series and thus laying the foundation for forecasting adopting a couple of iterative processes which will be explained in the following sections.

Details of the data: Observations-3282, Type-Daily Tools used: Eviews and SAS.

II. DATA FEATURES AND EXAMPLE

The first step in time series analysis, or any good statistical analysis for that matter, is to plot the data. However, unless the user knows what to look for in the plots this exercise is futile. Features of interest usually include:

A. Trend

It is an overall long term upward or downward movement in the data. Do we handle this by differencing the data or fitting a low order polynomial in time?

B. Seasonality

978-1-4244-6928-4/10/$26.00 ©2010 IEEE

A component of the series repeats periodically, for example, retail sales have a tendency to be high around November and December and lower near the first of the year. Do we look at seasonal differences or put in seasonal "indicator variables"?

C. Variability

Is the variation additive or multiplicative (in which case we might want to analyze the logarithms of the data).

D. Special events

Was there a strike sometime during the period? Does the time period include the dotcom bubble and the recession in 2008? Is there an effect of the same or some other impingements?

The features of the data determine which procedures are appropriate and identify those that are not appropriate. There are statistical checks available in the system for deciding, for example, the best approach for modeling the trend.

As an example, Fig. 1 shows the change in the Index values of NIFTY MIDCAP 50 in INDIA. While not exactly steady, we can see an overall increase/decrease in market with time. If we are thinking in terms of polynomials, it seems that at least a cubic polynomial might be needed to approximate the curvature in the historic data plot. The flexibility of high degree polynomials allows them to fit historic data well, but becomes a liability in forecasting in that such polynomials can, by their inherent flexibility, produce unreasonable increases or decreases as they are projected out into the future.

Having checked with the seasonality and stationary test of the data, the logarithmic value of the data CLOSE (closing point of NIFTY) is taken and rechecked for the stationary test. If the data is a stationary, then the data is ready to model into a time series. The stationarity tests are done by using the Time Series Analysis in SAS, preferably The Fullers Test.

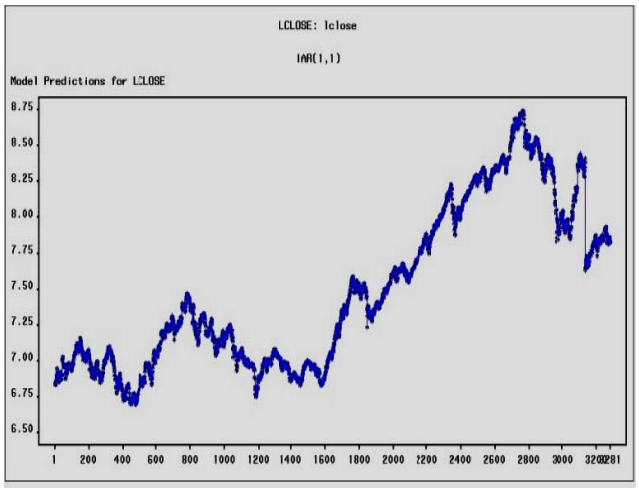

Figure 1. Log Transformed Flows

(Data Series: Obs - 3282 Type - Daily)

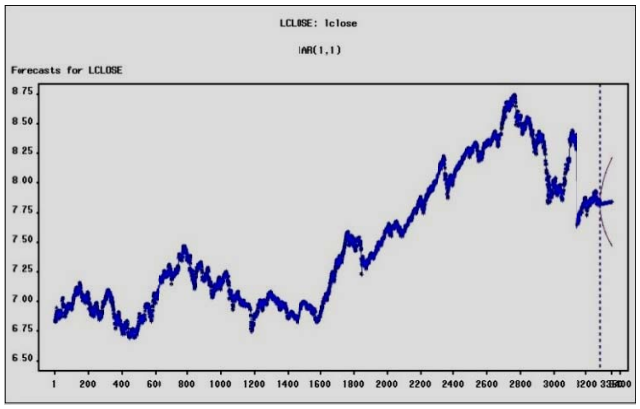

Figure 2. Log Transformed Forecast

The accuracy of the above forecast shown in fig. 2 can be viewed in the fig 3 which shows the difference of the actual values and the predicted values of the market by considering the entire data of 13years at once and later on the same difference in case of the splitted graph.

Fig 3 shows the statistics of differences in forecasting of the data: the graph (trend) with split and without split.

Table 1
Differencing of Forecasted Data with Split & Without Split

WITHOUT SPLIT			
7.8317	2519.208	7.8267	2506.644
7.8336	2536.651	7.8318	2519.46
	17.44265		12.81654
WITH SPLIT			
7.8317	2519.208	7.8267	2506.644
7.8324	2520.972	7.829	2512.416
	1.764063		5.771916

Thus, from Table 1 it is evident that the data series has to be split as per the trends and seasonality available.

Having inspected the trends available in the above data series, it is found that there are various trends in the market index (closing) and had resulted in the 10 splits.

The entire project has been verified while considering each and every split that shows some or the other trend and/or the impact of global impacts such as the Dotcom bubble and followed by the Recession in the year of 2008.

The above impingements referring the dotcom bubble and recession are clearly observed globally and their impacts on the Indian Market are analyzed and assessed.

It can be inferred that though Indian markets have shown a slight drift in its market during the above impacts, the overall effect of the recession on India is far negligible compared to the other developed countries (in reference to

the BPO sector: India at its best) and we can say India along with other Pan-Asia countries have stood back from such wallops.

III. RELATED WORK

The Time Series Prediction analyzes historic data and attempts to approximate future values of a time series as a linear combination of these historic data. In econometrics there are two basic types of time series forecasting: univariate (simple regression) and multivariate (multivariate regression). These types of regression models are the most common tools used in econometrics to predict time series. The way they are applied in practice is that firstly a set of factors that influence (or more specific is assumed that influence) the series under prediction is identified. These factors are the explanatory variables x_i of the prediction model.

A. Methodology:

The methodology includes

- Collecting the Secondary data from the market sources such as National Stock Exchange (Nifty).
- Checking the seasonality in the data series (if any) and minimizing the same for an accurate prediction.
- Claiming the data series to be precise – i.e. checking whether the data series obtained is stationary and raising the steps to make it a stationary series. Having checked with the stationarity, the data would be crisp and apparently ready to fetch a good forecast.
- Subjecting the resultant series into various models such as AR, MA etc.
- Forecasting the index and plotting the change in behavior of the market indices which is a real time analysis. This will help analysts/investors to make good and accurate investment decisions.

B. Limitations:

- Access to the market's real time data would be time variant and a slight change in the market may change the trend.
- Time series forecasting heavily depends upon the availability of the data. Long time series give the flexibility of verifying alternative models and choose the best.
- Exact modeling needs lots of experience and trend identification will be heavily depend upon this experience.
- Though the study gives a trend analysis it might not work in a uniform way for all the investors, therefore the behavior of the investor also plays a vital role in the modeling.
- The resultant analysis may work only for a specific period and may always subject to change depend on other externalities which will impact the economy of the country.
- Limited time may or may not fetch good results in implementation of the analysis since share market cannot be cracked in the real time.

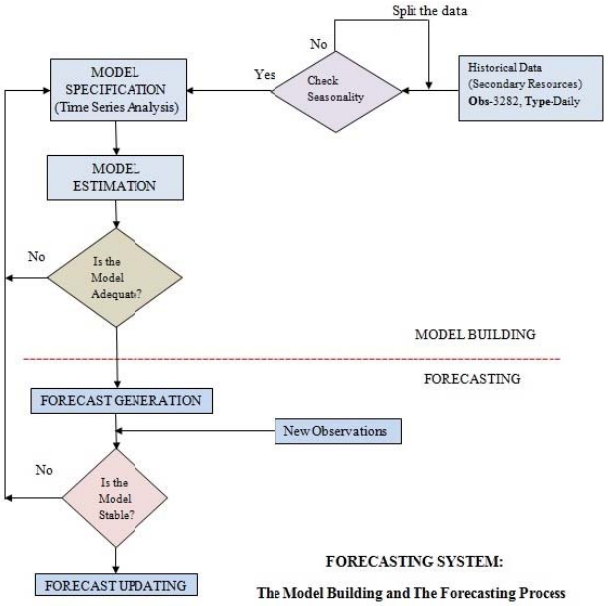

Figure 3. Forecasting System
The Model Building and The Forecasting Process

IV. PERFORMANCE ANALYSIS

The Performance analysis is been considered for the 10th split of the above data series since the entire paper concentrates on the Real-Time analysis of the NIFTY market, while comparing the predicted values with that of the Real-Time Index values and the same difference of Prediction is shown in the below Excel table.

Fig. 4 shows the graph (LCLOSE) for a time period of July 09 - Feb 10 and is not stationary. It can be determined by just viewing its trend and moreover the data is not moving in and out of the MEAN at equal intervals. However, the series clearly shows the trend with a non-stationary series. Such a trend does require a differencing while fitting into the model.

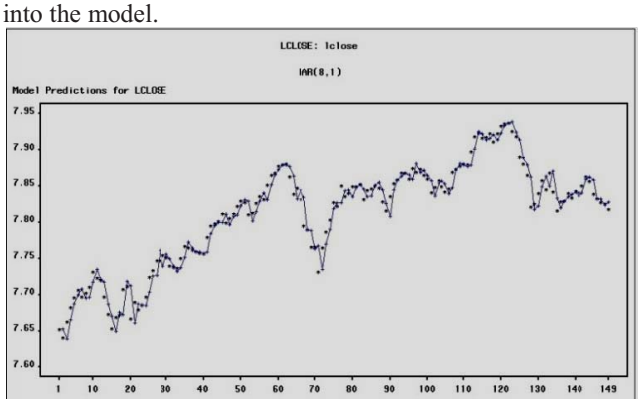

Figure 4 Log Transformed Close Values (LCLOSE)

The stationary series (ref fig. 4) is best fitted into the AR Model (Auto Regressive Model) and is fitted for eight lags. Fig 4 shows AR (8, 1) since the parameter estimates show that:

The probability is 0.0221 which says that the series (LCLOSE) is more significant at lag 8 and the coefficient of the standard error is almost zero (0.0839) as obtained during the evaluation of the above model leading to the best fit than any other model.

Table 2 shows the statistics of fit of LCLOSE graph with R-Square as 0.961 which says that the series is been fitted into the model with the 96.1% of confidence with the adjusted R-Square of 0.958 (almost same) and with the mean absolute error being almost zero thus leading to the best fit.

Table 2
Statistics of Fit

Statistics of Fit
LCLOSE: lclose
IAR(8,1)

Statistic of Fit	Value
Mean Square Error	0.0001930
Root Mean Square Error	0.01389
Mean Absolute Percent Error	0.13784
Mean Absolute Error	0.01075
R-Square	0.961
Adjusted R-Square	0.958

The reason for this best fit of almost 96.1% is due to the upward trend with the elevation of the data series which could make the prediction more accurate as it is a stationary.

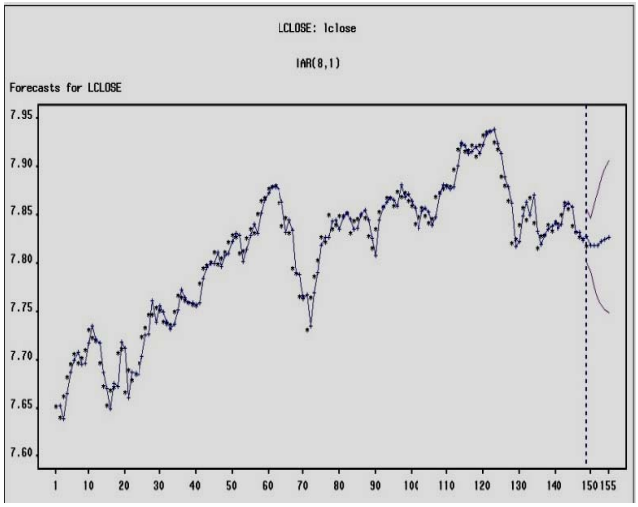

Figure 5 96.1% Confidence Interval Forecast (LCLOSE)

Fig. 5 is obtained on this day of 24th Mar 2010 (the updated data) and the values of the above forecasted graph are fed into the excel sheet to calculate the anti-log of the same in order to come up with the predicted values for the NIFTY closing.

Table 3
Predictions and Data Overlaid

Forecast Data Set
LCLOSE: lclose
IAR(8,1)

TIME	ACTUAL	PREDICT	U95	L95	ERROR	PERROR
145	7.8389	7.8586	7.8866	7.8305	-0.0197	-1.3755
146	7.8317	7.8320	7.8600	7.8039	-0.000234	-0.0164
147	7.8267	7.8325	7.8606	7.8045	-0.005838	-0.4080
148	7.8246	7.8241	7.8522	7.7961	0.000458	0.0320
149	7.8173	7.8285	7.8565	7.8005	-0.0112	-0.7821
150	.	7.8190	7.8470	7.7909	.	.
151	.	7.8182	7.8610	7.7753	.	.
152	.	7.8182	7.8735	7.7630	.	.
153	.	7.8223	7.8877	7.7569	.	.
154	.	7.8247	7.8977	7.7516	.	.
155	.	7.8274	7.9067	7.7482	.	.

The predictions of LCLOSE in table 3 are listed for the next 6 days: 7.8190 to 7.8274 are the forecasted values of the NIFTY closing for the dates above 24th of Feb 2010 that is for the real time of the future. These values are checked with the real time closing values of the NIFTY MIDCAP 50 on the following days. Table 4 shows the predicted values and real time values of the same.

Table 4
Predicted Vs Real Time Values

Date	Predicted Lclose	Predicted NIFTY close	NIFTY Real time values
25-Feb-10	7.819	2487.416747	2483.2
26-Feb-10	7.8182	2485.42761	2515.65
1-Mar-10	7.8182	2485.42761	
2-Mar-10	7.8223	2495.638781	
3-Mar-10	7.8247	2501.635508	

From Table 4 it is evident that the predicted value of the NIFTY MIDCAP 50 as 7.8190 and can be arrived at **2487.416 (anti-log of Predicted)** which has just **4.2167** points of difference with the real time i.e. **2483.20** which is a secondary data downloaded from the NSE site. The real time values for the rest of the periods are written manually and the difference is checked accordingly.

FINDINGS AND CONCLUSIONS

- The above forecasts of the time series models clearly show the behavior pattern of the market (NIFTY MIDCAP 50) in the near future. One may or may not find a difference in the values of predicted to that of the real time.
- There is always a possibility of certain standard errors in case of the time series analysis which may be found due to the absence of the best fit models than which are fitted above.
- Such and such above models namely AR do have some limitations which would result in an inaccurate prediction since these predicted values may sometimes have a huge difference with that of the real time values caused due to the external factors.

These external factors can be:

1. Policies implemented by the GOVERNMENT OF INDIA.
2. Variation in the rates such as
   - Interest Rates.
   - Fluctuations of Indian Rupee with the impact of other currencies.
   - Foreign Exchange Rates etc.
3. Unforeseen issues such as :
   - A sudden decline in the foreign markets.
   - A variation in the chain of demand and supply.
   - Sudden attack by the outlaws.

ACKNOWLEDGMENT

I would like to take this opportunity to thank Prof. Prashobhan Palakkeel and Prof. Veena Bhat (faculties, IBS Bangalore) for their assistance in the concepts of Business Intelligence and statistical tools used in this research namely SAS 9.1 & Eviews 6.0; their ceaseless support has resulted in such a vulnerable & successful explore.

This research paper would not been possible without the untiring support provided by my family and friends. The input and guidance provided by other faculties have been invaluable. I would like to extent my gratitude to all these people.

REFERENCES

[1] www.nse-india.com (NSE National Stock Exchange, India) as a secondary data source for Nifty daily.

[2] http://www.statsoft.com/textbook/time-series-analysis/

[3] http:/webspace.qmul.ac.uk/dsgpollock/public_html/courses/tseries/1trends.pdf

[4] www.nse-india.com/content/fo/fo_NIFTYMIDCAP50.htm

[5] Brocklebank, J. C. and D. A. Dickey (2003) SAS for Forecasting Time Series, SAS Institute, Cary, N.C.

[6] http://nseguide.com/press-releases/nse-index-weights-based-on-nse-daily-bhav-copy/

[7] http://www.hkbu.edu.hk/~billhung/econ3600/application/app03/app03.html

[8] http://webspace.qmul.ac.uk/dsgpollock/public_html/courses/tseries/8idntify.pdf

[9] http://www.duke.edu/~rnau/411arim.htm

"Model Selection is seldom Precise in Time Series modeling, because it's An Art more Than a Science"