CHAPTER-3
RESEARCH METHODOLOGY

3.1 OVERVIEW

Research methodology is a way to systematically solve the research problem. It may be understood as a science of studying how research is done scientifically. In it, the researcher tries to explain the various steps that are generally adopted by him in studying his research problem along with the logic behind them. It is necessary for the researcher to know not only the research techniques but also the methodology. Researchers not only need to know how to develop certain indices or tests or how to apply particular research techniques but they also need to know which of these techniques are relevant and which are not, and what would they mean and indicate. Researchers also need to understand the assumptions underlying various techniques and they need to know the criteria by which they can decide that certain techniques and procedures will be applicable to certain problems and others will not. This means that it is necessary for the researcher to design a methodology suitable for his research problem as the same may differ from problem to problem (Kothari, 1985).

Research methodology has many dimensions and research methods do constitute a part of the research methodology. The scope of research methodology is wider than that of research methods. Thus, research methodology includes not only research methods but also the logic behind the chosen research methods and explains why researchers are using a particular method or technique and why researchers are not using others so that research results are capable of being evaluated either by the researchers themselves or by others.

3.2 NEED FOR THE STUDY

There are several reasons to take up this study. Volatility represents risk and is a matter of concern for anyone who is dealing with money or investing in the stock market or any other financial instruments. Hence, the issue of volatility has become
increasingly significant in recent times for the financial practitioners, market participants, retail investors, regulators and researchers.

Volatility is a matter of great concern for market participants for the simple reason that as an investor one would like to know how much volatility or risk, he or she is exposed to, as more volatile a stock is, the more risky it is. Knowing the volatility of a stock provides some idea about what possible range of values it will take on some future date and can make informed decisions on his investments. Nonetheless, it is hard to predict with any certainty the price of a volatile stock. In general, people dislike risk and would like to have less risk or no risk while investing. Secondly, organizations entrusted with the job of regulating the market also need a clear idea regarding the pattern of volatility for framing policies to protect the interests of investors. Hence, an understanding of the market volatility is, thus, important from the regulatory policy perspective as well. Lastly, it is perceived as an indicator of market inefficiency and potential threat to the very integrity of market mechanism.

Stock return volatility hinders economic performance through consumer spending (Garner, 1988). For example, immediately after a persistent drop in stock prices, economic forecasts generally predict sharply weaker economic growth. They believe that the fall in stock prices would reduce consumer spending. The sizeable fall in consumer wealth as a result of fall in stock prices is expected to directly lower consumer spending. In addition, a weakening in consumer confidence could contribute to a further spending reduction.

Stock return volatility may also affect business investment spending (Gertler and Hubbard, 1989). Investors may perceive a rise in stock market volatility as an increase in the risk in equity investments. If so, investors may shift their funds to less risky assets. This reaction would tend to raise the cost of funds to firms issuing stock. Moreover, small firms and new firms might gravitate towards the purchase of stock in larger well-known firms.

Further, extreme stock return volatility could disrupt the smooth functioning of the financial system and lead to structural or regulatory changes. Systems that work well with normal return volatility may be unable to cope with the extreme price changes. Changes in the market rules of regulations may be necessary to increase the resiliency of the market in the face of greater volatility.
Although the issue of stock market volatility is of vital importance, yet limited very rare efforts have been made in India to examine it empirically and adequately. A plethora of research has been done on this subject in developed markets such as USA, Australia, UK and other European markets, and some emerging markets of Southeast Asia and Latin America. However, only one or two studies have been done on the magnitude and pattern of volatility in the Indian stock market. The present study is an attempt to fill this gap and aims to empirically investigate stock market volatility in India.

Hence, the empirical findings of the present study can be considered useful would be useful for financial practitioners, market participants, regulators and researchers, especially investors as these provide evidence of stock market volatility and its pattern prevailing in India. Investors aim at making more profitable and less risky investments. Therefore, they need to have a deep knowledge about stock market volatility before making their investment decisions.

However, from an investor’s perspective, it would be immensely useful if the future stock return volatility could be predicted from the past data. Such forecasting capability is useful for the pricing of sophisticated financial instruments such as futures and options. An attempt is made to model the stock market volatility process from the past stock return data. It is generally said stock market volatility is predictable. This observation has important implications for asset pricing and portfolio management. Investors seeking to avoid risk, for example, may choose to adjust their portfolios by reducing their commitments to assets whose volatilities are predicted to increase or by using more sophisticated dynamic diversification approaches to hedge predicted volatility increases. In a market in which such strategies operate, equilibrium asset prices should respond to forecasts of volatility, as well as to the risk aversion of investors. This is particularly true of the markets for derivative assets such as options and swaps, where the volatility of the underlying asset has a profound effect on the value of the derivative.

3.3 OBJECTIVES OF THE STUDY

Throughout unprecedented share market boom, there was enough indication that the market is behaving irrationally, it is extremely volatile thus its future course of movement is simply unpredictable. Stock market volatility has been the subject matter of many studies conducted abroad. However, no significant efforts have been
made in India to examine the issue of stock market volatility. No systematic study is available about the magnitude, components and pattern of volatility in the Indian stock market. Thus, the present study attempts to fill this void. It attempts to measure stock return volatility in quantitative terms. Some of the important issues related to volatility such as factors causing for volatility, relationship between volatility and stock market have also been examined. An effort has been made to model the stock return volatility. The study, thus, aims to evaluate current level of volatility in historical perspective and examines the effect of some factors on volatility.

The objectives of the study are:

1. To study the trends of volatility in Indian stock market;
2. To examine the factors which influence volatility of stock prices;
3. To study relationship between the volatility and stock price behaviour;
4. To analyze the positive and negative consequence of volatility;
5. To arrive at stock market volatility model.

3.4 SCOPE OF THE STUDY

The scope of the study is restricted to the extent that the BSE (Sensex) and NSE (Nifty) data considered for the present research work covers a period of sixteen financial years only, i.e., from 1996-2011. The daily closing values of each exchange form the data of this study. The crucial factors affecting the stock prices considered for the study are FIIs, Interest Rate, Inflation, Union Budget, and Company Earning. The financial position of a company has been evaluated on the basis of its earning per share, dividend per share, market price, price-earning ratio and book value of share. The study takes into consideration the data on these factors concerning the banking and automobile sectors only.

3.5 SAMPLE SIZE

The study is based on the stock prices of two important stock exchanges in India, i.e., BSE and NSE. For this purpose, one market index is selected from each of these two stock exchanges such as BSE Sensex and S&P CNX Nifty respectively. NSE and BSE are the two major exchanges in terms of volumes traded, delivered and market capitalization. Of these two exchanges, NSE accounts for two-thirds of the turnover and the rest is claimed by BSE. The Wholesale Price Index (WPI) is used as proxy for inflation because it is the most widely used price index in India, and is an indicator of movement in wholesale prices of 435 commodities in all trade and
transactions. WPI is the only price index in India, which is available on a weekly basis with the shortest time lag of two weeks. It is due to these attributes that the index is widely used in business and industry circles, and by the government. It is generally taken as an indicator of the rate of inflation in the economy. Most of the previous empirical studies investigating the macroeconomic factors and stock relationship have used WPI as a measure of inflation. The base year of the WPI series is 1993-94. Bank Rate is used as proxy for interest rate. As many as five companies have been finally selected from the banking and automobile sectors. The companies are HDFC Bank Limited, State Bank of India, Tata Motors Limited, Hero Motor Corporation Limited, Mahindra and Mahindra Limited. While selecting the sample of the companies from industries following criteria was adopted:

- The necessary financial data required for calculating the measures of dependent and independent variable pertaining to all the years, i.e., 1996-20011 is available.
- The listed shares on BSE Sensex and S & P CNX Nifty are considered.

The data relating to the companies was taken from PROWESS database of the Centre for Monitoring of Indian Economy (CMIE).

3.6 SAMPLING TECHNIQUE

Stratified Sampling

Under this technique, the universe or entire population is divided into number of groups or strata, and sample is taken from each group. NSE and BSE have been chosen as sample from the Indian Stock Market for the study because these are the two major stock exchanges in terms of volumes traded, delivered and market capitalization. Of these two exchanges, NSE accounts for two-thirds of the turnover and the remaining one-third is claimed by BSE. Further, one market index is selected from each of these two exchanges such as BSE Sensex and S&P CNX Nifty respectively because these indices give a broad outline of the market movement and represent the market.

3.7 DATA COLLECTION

The closing figures of the two indices, the Sensex and Nifty have been collected from the official websites of BSE and NSE. FII transactions, i.e., purchases, sales and net investment data have been taken from SEBI Handbook. Wholesale Price Index data has been taken from the website of Central Statistics Organisation. Bank rates have been taken from Handbook of Statistics of the Reserve Bank of India.
(RBI). The Budget dates and the names of their respective presenters has been gathered from Finance Ministry website. The company financial data have been taken from PROWESS, the online database maintained by Centre for Monitoring of Indian Economy (CMIE). Business Line, Economic Times, Capital Market, and Money Control are the other sources used for collecting the required information.

3.8 STATISTICAL TOOLS

3.8.1 Measures of Central Tendency

Measures of central tendency provide information about a representative value of the data set. Arithmetic mean (simply called the mean), median and mode are the most common measures of central tendency.

3.8.1.1 Mean or average is the sum of the values of a variable divided by the number of observations.

If \( n \) numbers are given, each number denoted by \( x_i \), where \( i = 1, \ldots, n \), the arithmetic mean is the sum of the \( x_i \)'s divided by \( n \), or

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} (x_1 + \cdots + x_n)
\]

3.8.1.2 Median refers to the middle value in a distribution. In the case of median, one-half of the items in the distribution have a value the size of the median value or smaller; and one-half have a value the size of the median value or larger. The median is just the \( 50^{th} \) percentile value below which 50 per cent of the values in the sample fall. It splits the observation into two halves. Hence, when \( N \) (number of observation) is odd, the median is an actual value, with remainder of the series in two equal parts on either side of it. If \( N \) is even, the median is a derived figure, i.e., half the sum of the middle values.

Median = Size of \( \left( \frac{N + 1}{2} \right) \)th item

3.8.1.3 Mode is that value in a series of observations which occurs with the greatest frequency. For determining mode count the number of times the various values repeat themselves and the value occurring maximum number of times is the modal value or mode. The more often the modal value appears relatively, the more valuable the measure is an average to represent data.

3.8.2 Standard Deviation
Standard deviation is the most widely used measure of dispersion of a series and is commonly denoted by the symbol \( \sigma \) (pronounced as sigma). It shows how much variation or "dispersion" there is from the average (mean, or expected value). It measures the absolute dispersion, the greater the standard deviation, the greater will be the magnitude of the deviations of the values from their mean. A small standard deviation means a high degree of uniformity of the observation as well as homogeneity of a series; a large standard deviation means just the opposite. Standard deviation is defined as the square root of the average of squares of deviations, when such deviations for the values of individual items in a series are obtained from the arithmetic average. It is worked out as under:

\[
s_\text{SN} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}
\]

In this formula, \( S_\text{N} \) is standard deviation, \( \bar{x} \) is the value of the mean, \( N \) is the sample size, and \( x_i \) represents each data value from \( i=1 \) to \( i=N \).

When we divide the standard deviation by the arithmetic average of the series, the resulting quantity is known as coefficient of standard deviation which happens to be a relative measure and is often used for comparing the variability in two or more series. When this coefficient of standard deviation is multiplied by 100, the resulting figure is known as coefficient of variation.

### 3.8.3 Skewness

A distribution is said to be skewed if the observations above and below the mean are not symmetrically distributed. A zero value of skewness implies a symmetric distribution. The distribution is positively skewed when the mean is greater than the median and negatively skewed when the mean is less than the median.

\[
SKp = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}
\]

\( SKp \) = Karl Pearson's coefficient of skewness

### 3.8.4 Kurtosis

Kurtosis is a measure of how peaked or flat a distribution is. A distribution could be mesokurtic, leptokurtic or platykurtic. The absolute value of kurtosis for a mesokurtic or normal distribution is 3; kurtosis for other distributions is always measured relative to this value. Platykurtic distribution has a negative kurtosis,
implying a flatter distribution than the normal distribution while leptokurtic
distribution has positive kurtosis, implying a more peaked distribution than normal
distribution.

3.8.5 Multiple Regressions

Regression analysis is used to assess the relationship between one dependent
variable (DV) and several independent variables (IVs). However, for the purpose of
this study, multiple regression has been used taking only three independent variables
and one dependent variable. In this situation, the results are interpreted as shown
below:

Multiple regression equation assumes the form

\[ Y = a + b_1X_1 + b_2X_2 + b_3X_3 \]

Where, \( X_1, X_2 \) and \( X_3 \) are three independent variables and \( Y \) being the dependent
variable, and \( a, b_1, b_2 \) and \( b_3 \) are the constants.

Regression Coefficient

Regression coefficient is a measure of how strongly each IV (independent
value also known as predictor variable) predicts the DV (dependent value). There are
two types of regression coefficients – unstandardized coefficients and standardized
coefficients, also known as beta values. The unstandardized coefficients can be used
in equation as coefficients of different IVs along with the constant term to predict
value of DV. The standardized coefficient (beta) is, however, measured in standard
deviations. A beta value of 2 associated with a particular IV indicates that a change of
1 standard deviation in that particular IV will result in change of 2 standard deviations
in the DV.

R-value

\( R \) represents the correlation between the observed values and the predicted
values (based on the regression equation obtained) of the DV. \( R^2 \) is square of R and
gives the proportion of variance in the dependent variable accounted for by the set of
IVs chosen for the model. \( R^2 \) is used to find out how well the IVs are able to predict
the DV. However, the \( R^2 \) value tends to be a bit inflated when the number of IVs is
more or when the number of cases is large. The adjusted \( R^2 \) takes into account these
things and gives more accurate information about the fitness of the model. For
example, an adjusted \( R^2 \) value of 0.70 would mean that IVs in the model can predict
70\% of the variance in the DV.
3.8.6 Paired t-test

Paired t-test is a method for comparing two related samples, involving small values of n that does not require variances of the two populations to be equal, but the assumption that the two populations are normal must continue to apply. For a paired t-test, it is necessary that the observations in the two samples be collected in the form of what is called matched pairs, i.e., each observation in the one sample must be paired with an observation in the other sample in such a manner that these observations are somehow matched or related, in an attempt to eliminate extraneous factors which are not of interest in test. Such a test is generally considered appropriate in a before and after treatment study. To apply this test, first work out the difference score for each matched pair, and then find out the average of such differences, $\bar{D}$, along with the sample variance of the difference score. If the values from the two matched samples are denoted as $X_i$ and $Y_i$ and the differences by $D_i (D_i = X_i - Y_i)$, then the mean of differences i.e.,

$$\bar{D} = \frac{\sum D_i}{n}$$

and the variance of the differences or

$$(\sigma_{diff})^2 = \frac{\sum D_i^2 - (\bar{D})^2 \cdot n}{n - 1}$$

Assuming the said differences to be normally distributed and independent, we can apply the paired t-test for judging the significance of mean of differences and work out the test statistic $t$ as under:

$$t = \frac{\bar{D} - 0}{\sigma_{diff} / \sqrt{n}} \text{ with (n-1) degrees of freedom}$$

where $\bar{D} = \text{Mean of differences}$

$\sigma_{diff} = \text{Standard deviation of differences}$

$n = \text{number of matched pairs}$

This calculated value of $t$ is compared with its table value at a given level of significance as usual for testing purposes.

3.8.7 ANOVA

ANOVA or analysis of variance is used to compare the means of more than two populations. ANOVA analysis uses the F-statistics, which tests if the means of
the groups, formed by one independent variable or a combination of independent variables are significantly different. It is based on the comparison of two estimates of variances: one representing the variance within groups, often referred to as error variance; and the other representing the variance due to differences in group means. If the two variances do not differ significantly, one can believe that all group means come from the same sampling distribution of means and there is no reason to claim that the group means differ. If, however, the group means differ more than can be accounted for due to random error, there is reason to believe that they were drawn from different sampling distributions of means. The F-statistics calculates the ratio between the variance due to difference between groups and the error variance.

$$F = \frac{\text{Variance due to difference between groups}}{\text{Error variance}}$$

The larger the F-ratio, the greater is the difference between groups as compared to within group differences. An F-ratio equal to or less than one indicates that there is no significant difference between groups and the null hypothesis is correct. If the null hypothesis (that the group means do not differ significantly) is correct, then it can be concluded that the independent variables do not have an effect on the dependent variable. However, if F-test proves that the null hypothesis to be wrong, multiple comparison tests are used to further explore the specific relationships among different groups.

### 3.8.8 Unit Root Test

A unit root test is a statistical test for the proposition that is an autoregressive statistical model of a time series, the autoregressive parameter is one. It is a test for detecting the presence of stationarity in the series. A non-stationary variable has a definite positive or negative trend over time, and is said to have a unit root. A variable whose time series is represented by a first-order, autoregressive scheme, AR (1) for short, i.e.

$$Y_t = \alpha Y_{t-1} + \varepsilon_t \quad \text{Where, } \varepsilon_t \text{ is a standard normal variable.}$$

If $\alpha$ is less than one, then the time path is stationary, i.e., although it will fluctuate the value of $Y_t$ will tend to keep coming back to its mean value. Graphically, the time path of $Y_t$ will have no upward or downward trend, and the fluctuations around the constant mean value will be contained within constant bounds.
However, if \( \alpha \) is equal to one, the time path of \( Y_t \) is non-stationary, and it will have an upward trend. As \( \alpha \) is known as the root of the process, non-stationarity implies having a unit root.

So, if a variable is non-stationary, how can it be made stationary? The answer is differencing. To remove the trend in a series takes the first difference, i.e.

\[
\Delta Y_t = Y_t - Y_{t-1}
\]

The first difference will usually be stationary, i.e., have no unit root. If so, the variable is said to be integrated of order 1, written I(1), meaning that the variable has to be differenced once to make it stationary. Sometimes a variable may have to be differenced twice to make it stationary, i.e., \( Y_t \) and \( \Delta Y_t \) are non-stationary, but \( \Delta^2 Y_t \) is stationary. In this case, \( Y_t \) is said to be integrated of order 2, written I(2).

**Testing for Stationarity**

There are two approaches for testing a variable to see if it is stationary, an informal and a formal method. The informal approach is to look at a time plot of the variable and see if there is any obvious trend in the data. The second, preferable method is to perform a formal test of stationarity known as the Augmented Dickey-Fuller (ADF) test.

Augmented Dickey-Fuller (ADF) test is most frequently used test of unit root. It is based on simple logic. A non-stationary process has infinite memory as it does not show decay in a shock that takes place in the process. Every random shock carries away the process from its earlier level not to return back again unless another random shock pushes it towards its previous level. Therefore, it behaves like AR (1) process with \( \rho = 1 \). Dickey-Fuller test is designed to examine if \( \rho = 1 \).

Let;

\[
y_t = \rho y_{t-1} + \varepsilon_t
\]

\[
\Rightarrow y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \varepsilon_t
\]

\[
\Rightarrow \Delta y_t = (1 - \rho) y_{t-1} + \varepsilon_t
\]

\[
\Rightarrow \Delta y_t = \delta y_{t-1} + \varepsilon_t
\]

In ADF test, the hypothesis if \( \delta = 0 \) has been tested. The test procedure is similar to usual t-test but standard critical values of the t-test are not valid in this case.
To get white noise the lagged terms of $\Delta y_t$ are also included in the regression. The results may be sensitive to number of lag terms included.

Similarly, deterministic terms such as intercepts and trends may also be included in the model. The complete model will look like this:

$$\Delta y_t = \alpha + \pi + \delta \Delta y_{t-1} + \sum_{i=1}^{m} \beta_i \Delta y_{t-i} + \varepsilon_t$$

Depending on which terms we include in model specification, the following hypotheses are tested:

(i) Unit root vs stationary process with zero mean.
(ii) Unit root with drift vs stationary process with constant mean.
(iii) Unit root with drift and trend vs. stationary process with trend.

3.8.9 Granger Causality

A statistical approach proposed by Granger (1969) to infer cause and effect relationship between two or more time series is known as Granger causality. Granger causality is based on the simple logic that effect cannot precede cause.

It is important to note that the statement $\bar{x} \rightarrow \bar{y}$ Granger causes $y \rightarrow \bar{y}$ does not imply that $y$ is the effect or the result of $x$. Granger causality measures precedence and information content but does not by itself indicate causality in the more common use of the term.

Procedure of Traditional Granger Non-Causality Test

In its original form it is based on following bivariate regression model (there are some other procedures used for causality testing such as Sims Causality test, Hasiao Causality Test etc.)

$$y_t = \alpha_0 + \sum_{i=1}^{l} \alpha_i y_{t-i} + \sum_{j=1}^{l} \beta_j x_{t-j} + \varepsilon_t$$

$$x_t = \omega + \sum_{i=1}^{l} \gamma_i x_{t-i} + \sum_{j=1}^{l} \theta_j y_{t-j} + \varepsilon_i$$

If all the coefficients of $x$ in first regression equation of $y$, i.e. $\beta_j$ for $i = 1 \ldots l$ are significant that the null hypothesis that $x$ does not cause $y$. However, the significance of the coefficient cannot be evaluated based on usual t-statistic. For this purpose the following procedure of testing the models is used.
(i) Estimate the model without including lagged values of variable $x$. Suppose the $R^2$ from this estimate is $R^2_1$.

(ii) Now estimate the model including lagged values of variable $x$. Suppose, the $R^2$ from this estimate is $R^2_2$.

(iii) F-ratio for improvement in the model is worked out as follows:

$$F = \frac{(R^2_2 - R^2_1)/k^*}{(1-R^2_2)/(n-k)}$$

Where $k^*$ are the number of lag orders $l$ of variable $x$, $k$ is the total number of the parameters estimated and $n$ is the number of observations. The null hypothesis of non-causality is rejected if F-statistic is greater than its critical value at $k^*$ and $(n-k)$ degree of freedom. Econometric software packages such as Eviews routinely test Granger causality.

Similarly from the second equation above, we can test the null hypothesis that $\dot{O}y$ does not cause $x \dot{O}$. If only one of the two variables causes the second variable but the second variable does not cause the first variable it is called one-way causality. If both the variables cause each other it is called the feedback causality.

3.8.10 ARCH and GARCH

Engle (1982) suggests that the conditional variance $h_t$ can be modelled as a function of the lagged $\hat{\sigma}_t^2$, i.e. the predictable volatility is dependent on past news. The most detailed model he develops is the $q$th order ARCH model, the ARCH($q$):

$$h_t = \chi + \bar{\Upsilon}_1 \hat{\sigma}_{t-1}^2 + \bar{\Upsilon}_2 \hat{\sigma}_{t-2}^2 + \ldots + \bar{\Upsilon}_q \hat{\sigma}_{t-q}^2$$

where,

$\chi, \bar{\Upsilon}_1, \bar{\Upsilon}_q = \text{parameters to be estimated}$

$h_t =$ conditional variance at period $t$

$q =$ number of lags included in the model

$\hat{\sigma}_t =$ innovation in return at time $t$

where, $\chi > 0$, $\bar{\Upsilon}_l, \bar{\Upsilon}_q > 0$. In an ARCH($q$) model, an old news which arrived at the market more than $q$ periods ago has no effect at all on current volatility. Alternatively, if a major market movement occurred yesterday, the day before or up to $q$ days ago, the effect will be to increase today's conditional variance.

Bollerslev (1986) generalized the ARCH ($q$) model to the GARCH ($p,q$) model, such that:
\[ h_t = \omega + \tilde{\epsilon}_t + \tilde{\epsilon}_t^2 + \tilde{\epsilon}_t \tilde{\epsilon}_{t-1}^2 + \ldots + \tilde{\epsilon}_t \tilde{\epsilon}_{t-q}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \ldots + \beta_p h_{t-p} \]

\( \omega, \tilde{\epsilon}_t, \tilde{\epsilon}_{t-1}, \ldots, \tilde{\epsilon}_{t-q}, \beta_1, \beta_2, \ldots, \beta_p \) are parameters to be estimated.

- \( h_t \) = conditional variance at period \( t \)
- \( q \) = number of return innovation lags included in the model
- \( p \) = number of past volatility lags included in the model
- \( \tilde{\epsilon}_t \) = innovation in return at time \( t \)

where \( \omega > 0, \alpha_1 \geq 0, \ldots, \alpha_q \geq 0, \beta_1 \geq 0, \ldots, \beta_p \geq 0 \).

The GARCH \((p, q)\) process defined above is stationary when \((\alpha_1 + \alpha_2 + \ldots + \alpha_q) + (\beta_1 + \beta_2 + \ldots + \beta_p) < 1\). The simplest but often very useful GARCH process is the GARCH (1,1) process which is also called the generic or 'vanilla' GARCH model given by:

\[ h_t = \omega + \tilde{\epsilon}_t + \tilde{\epsilon}_t^2 + \beta_1 h_{t-1} \]

where \( \omega > 0, \tilde{\epsilon}_0 = 0, \beta_1 = 0 \). The stationary condition for GARCH (1, 1) is \( \tilde{\epsilon}_t + \beta_1 < 1 \) (Karmakar, 2003).

In the GARCH (1,1) model, the effect of a return shock on current volatility declines geometrically over time. As referred earlier, the GARCH (1,1) model is found to be an excellent model for a wide range of financial data. The sizes of the parameters \( \tilde{\epsilon}_t \) and \( \beta_1 \) determine the short-run dynamics of the resulting volatility time series. Large GARCH lag coefficients \( \beta_1 \) indicate that shocks to conditional variance take a long time to die out, so volatility is \( \omega \) persistent. Large GARCH error coefficient \( \tilde{\epsilon}_t \) means that volatility reacts quite intensely to market movements and so if \( \tilde{\epsilon}_t \) is relatively high and \( \beta_1 \) is relatively low, then volatilities tend to be more \( \omega \) spikey. In financial markets, it is common to estimate lag (or \( \omega \) persistence) coefficients based on daily observation in excess of 0.8 and error (or \( \omega \) action) coefficients not more than 0.2. If \( \tilde{\epsilon}_t + \beta_1 \) is close to unity, then a \( \omega \) shock at time \( t \) will persist for many future periods. A high value of \( \tilde{\epsilon}_t + \beta_1 \), therefore, implies a \( \omega \) long memory. For \( \tilde{\epsilon}_t + \beta_1 = 1 \), any shock will lead to a permanent change in all future values of \( h_t \), hence, shock to the conditional variance is \( \omega \) persistent. For \( \tilde{\epsilon}_t + \beta_1 = 1 \) integrated GARCH process (i.e., IGARCH) is used. For IGARCH, the conditional variance is non-stationary and the unconditional variance is unbounded.

### 3.8.11 EGARCH

The EGARCH model specifies the conditional variance as follows:

\[ \log \sigma_t^2 = w^* + \beta \log \sigma_{t-1}^2 + a|\varepsilon_{t-1}| + \delta \varepsilon_{t-1} \]
where, $\gamma^* = \gamma - \beta b \mu$. The specification of the volatility in terms of its logarithmic transformation implies that there are not restrictions on the parameters to guarantee the positivity of the variance. Nelson (1991) establishes the conditions for covariance stationarity of the EGARCH model under particular specifications of the error distribution. Furthermore, a sufficient condition for the stationarity of the EGARCH model is $\beta b < 1$ when $\hat{\beta}$ has a distribution. Nelson establishes the conditions for covariance stationarity of the EGARCH model under particular specifications of the error distribution. In particular, assuming that $\beta b < 1$, the EGARCH model is always stationary if $\hat{\beta}$ has a Normal or a Generalized Error Distribution (GED) with parameter $\kappa > 1$. However, when $\hat{\beta}$ has a Student-$t$ or a GED distribution with parameter $\hat{\mu}$, $\gamma$ is stationary if $\hat{\mu} > 1$, which is a rather implausible restriction when dealing with real time series of returns.

3.9 METHODOLOGY

3.9.1 Methodology Concerning First Objective

In this objective, volatilities of market indices, i.e., BSE Sensex and S&P CNX Nifty have been calculated over different observation periods. Yearly volatility of monthly returns, Yearly volatility of daily returns and Monthly volatility of daily returns have been calculated. In Yearly volatility of daily returns, the rates of returns have been computed by taking a logarithmic difference (The logarithmic difference is symmetric between up and down movements and expressed in percentage terms for ease of comparability with the straightforward idea of a percentage change) of prices of two successive periods. Symbolically, it may be stated as follows:

$$r_t = \log_e (p_t/p_{t-1}) = \log_e (p_t) - \log_e (p_{t-1})$$

where $\log_e$ is natural logarithm, $p_t$ and $p_{t-1}$ are the closing prices for the two successive periods. In yearly volatility of monthly returns, the monthly returns have been calculated by natural logarithmic differences between the monthly opening and closing price of BSE Sensex and S&P CNX Nifty. Daily and monthly returns have been annualized by multiplying them with number of trading days and the number of months in a year respectively.

Volatility has been measured as standard deviation of the rates of return. The standard deviation is also based on logarithmic units. The standard deviation of return $r_t$ from a sample of $n$ observations is the square root of the average squared deviation.
of returns from the average return in the sample. Thus, standard deviation $S_N$ ($\bar{\sigma}$) is defined as:

$$S_N = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

While daily and monthly volatility have been annualized by multiplying them with square root of the number of trading days and the number of months in a year respectively.

Then spike volatility has been found and in last in order to understand the relative contribution of \textit{noise} and \textit{news} in causing abnormal price oscillations in the Indian stock market, list of events reported around extreme daily returns has been prepared.

### 3.9.2 Methodology Concerning Second Objective

FIIs, Interest Rate, Inflation, and Union Budget - all these crucial factors which influence the stock prices have been taken in this objective.

#### 3.9.2.1 Foreign Institutional Investors

The prime intention is to test the causality between net FII investment and stock market volatility. Hence four sets of hypotheses have been formulated.

(A) $H_{0i}$: Net FII investment does not Granger cause Volatility in the Stock Market (BSE).

$H_{1i}$: Net FII investment Granger causes Volatility in the Stock Market (BSE).

(B) $H_{0j}$: Stock Market Volatility (BSE) does not Granger cause Net FII investment.

$H_{1j}$: Stock Market Volatility (BSE) Granger causes Net FII investment.

(C) $H_{0k}$: Net FII investment does not Granger cause Volatility in the Stock Market (NSE).

$H_{1k}$: Net FII investment Granger causes Volatility in the Stock Market (NSE).

(D) $H_{0l}$: Stock Market Volatility (NSE) does not Granger cause Net FII investment.

$H_{1l}$: Stock Market Volatility (NSE) Granger causes Net FII investment.

Hence, the causal relationship between the net FII investment and stock market volatility is tested by applying the Granger Causality Test. But before applying the Granger Causality Test, a unit root test is performed using the Augmented Dicker-Fuller (ADF) method to test whether the series, i.e. the net FII investment and stock market volatility are stationary.
3.9.2.2 Inflation

The prime intention is to test the causality between inflation and stock market volatility. Hence four sets of hypotheses have been formulated.

(A) \(H_0\): Inflation does not Granger cause Stock Market Volatility (BSE).
\(H_1\): Inflation Granger causes Stock Market Volatility (BSE).

(B) \(H_0\): Stock Market Volatility(BSE) does not Granger cause Inflation.
\(H_1\): Stock Market Volatility(BSE) Granger causes Inflation.

(C) \(H_0\): Inflation does not Granger cause Stock Market Volatility (NSE).
\(H_1\): Inflation Granger causes Stock Market Volatility (NSE).

(D) \(H_0\): Stock Market Volatility(NSE) does not Granger cause Inflation.
\(H_1\): Stock Market Volatility(NSE) Granger causes Inflation.

Hence, the causal relationship between the Inflation and Stock market volatility is tested by applying the Granger Causality Test. But before applying the Granger Causality Test, a unit root test is performed using the Augmented Dickey-Fuller (ADF) method to test whether the series, i.e. the Inflation and Stock market volatility are stationary.

3.9.2.3 Union Budget

Event study (An empirical study of prices of an asset just before and after some event, like an announcement, merger, or dividend) has been used to examine the impact of budget on share price. A total of 60 trading days before and after the budget have been considered to study the impact of budget (shown in Appendix 4). This has been done on an assumption, that an impact of budget on share price can be identified on its own for a maximum of 30 days beyond which many other causes may distort the said effect. The study considers only trading days and leaves out any holiday or other days when the market remains closed. These trading days have been segregated into Short-term (3 trading days), Medium-term (15 trading days), and long-term periods (30 trading days) both before and after the budget.

A total of 12 hypotheses tests have been designed to understand the statistical significance of the impact. The test tried to compare the average returns during various time periods with one another and also the budget day impact with average return from previous periods. In the hypotheses tests, \(\mu(RX_1)\), \(\mu(RX_2)\) and \(\mu(RX_3)\) represent the average daily returns during the previous 30, 15 and 3 trading days respectively. \(\mu(RY_1)\), \(\mu(RY_2)\) and \(\mu(RY_3)\) represent the average daily returns during
the next 3, 15 and 30 trading days respectively. R2 represents the budget day (event day-end) return. This is not an average figure as it is a single day's logarithmic return over the previous day's closing figures.

The **null hypothesis** \( H_0 \) has assumed that the budget has no impact on Sensex. All hypotheses have been tested at 5% level of significance at the left tail.

In the first three tests, **alternative hypothesis** \( H_1 \) intends to prove that budget day return i.e., \( (R_2) \) is more than that found during the previous 30, 15 and 3 trading days i.e., \( \mu(RX_1), \mu(RX_2) \) and \( \mu(RX_3) \) for all budgets. In the next set of nine tests, alternative hypothesis \( H_1 \) intends to prove that post-budget average return, i.e., next 3, 15 and 30 trading days [i.e., \( \mu(RY_1), \mu(RY_2) \) and \( \mu(RY_3) \)] are more than pre-budget average return, i.e., previous 30, 15, 3 trading days [i.e., \( \mu(RX_1), \mu(RX_2) \) and \( \mu(RX_3) \)] for all budgets. Paired t-tests have been applied for hypothesis testing.

In next part of the study, the variance of returns (\( \hat{\sigma}^2 \)) have been compared between various time periods in order to find out the extent of volatility (\( \hat{\sigma} \)) in the market around the budget period. F-test has been applied for hypothesis testing. Two sets of hypothesis tests have been conducted in this part of analysis. In the first set, variances of returns during the short-term, medium-term and long-term periods in the post-budget situation have been compared to one another, i.e., variances of returns between \( Y_1 \) & \( Y_2 \), \( Y_2 \) & \( Y_3 \) and \( Y_1 \) & \( Y_3 \) respectively have been examined. These comparisons have been made because variances are expected to rise with the increasing time period and fluctuations in return are also historically found to be continuing in a post-budget situation. The **null hypothesis** in all the three tests assumes no change in variance i.e., the variances are equal. The **alternative hypothesis** \( H_1 \) is that variance during \( Y_2 \) is more than that of during \( Y_1 \), variance during \( Y_3 \) is more than that of during \( Y_2 \) and variance during \( Y_3 \) is more than that of during \( Y_1 \) respectively.

In second set, each of the post-budget short-term, medium-term and long-term periods has been compared to the long-term pre-budget period, i.e., variances of returns between \( X_1 \) & \( Y_1 \), \( X_1 \) & \( Y_2 \) and \( X_1 \) & \( Y_3 \) respectively have been examined. These tests have been framed with a more empirically established fact that the variances in returns after budget for different periods are expected to be greater than the pre-budget long-term variance. The **null hypothesis** in all the three tests assumes
no change in variance, i.e., the variances are equal. The alternative hypothesis $H_1$ wants to prove that variance during $Y_1, Y_2$ and $Y_3$ is more than that of during $X_1$.

### 3.9.2.4 Interest Rate

Event study has been used to examine the impact of announcement of interest rate on share price. A total of 60 trading days before and after the announcement of interest rate have been considered to study the impact of announcement of interest rate (shown in Appendix 5). This has been done on an assumption that an impact of announcement of interest rate on share price can be identified on its own for a maximum of 30 days beyond which many other causes may distort the said effect. The study considers only trading days and leaves out any holiday or other days when the market remains closed. These trading days have been segregated into Short-term (3 trading days), Medium-term (15 trading days), and long-term periods (30 trading days) both before and after the announcement of interest rate.

A total of 12 hypotheses tests have been designed to understand the statistical significance of the impact. The test tried to compare the average returns during various time periods with one another and also the announcement day impact with average return from previous periods. In the hypotheses tests, $\mu(RX_1), \mu(RX_2)$ and $\mu(RX_3)$ represent the average daily returns during the previous 30, 15 and 3 trading days respectively. $\mu(RY_1), \mu(RY_2)$ and $\mu(RY_3)$ represent the average daily returns during the next 3, 15 and 30 trading days respectively. $R_Z$ represents the announcement of interest rate day (event day-end) return. This is not an average figure as it is a single day's logarithmic return over the previous day's closing figures.

The null hypothesis $H_0$ has assumed that the announcement of interest rate has no impact on Sensex. All hypotheses have been tested at 5% level of significance at the left tail.

In the first three tests, alternative hypothesis $H_1$ intends to prove that announcement of interest rate day return i.e. $R_Z$ is more than that found during the previous 30, 15 and 3 trading days [i.e. $\mu(RX_1), \mu(RX_2)$ and $\mu(RX_3)$]. In the next set of nine tests, alternative hypothesis $H_1$ intends to prove that post announcement of interest rate average return i.e. (next 3, 15 and 30 trading days [i.e. $\mu(RY_1), \mu(RY_2)$ and $\mu(RY_3)$] are more than pre announcement of interest rate average return i.e. (previous 30, 15, 3 trading days [i.e., $\mu(RX_1), \mu(RX_2)$ and $\mu(RX_3)$]). Paired t-tests have been applied for hypothesis testing.
In next part of the study, the variance of returns ($\sigma^2$) have been compared between various time periods in order to find out the extent of volatility ($\sigma$) in the market around the announcement of interest rate period. F-test has been applied for hypothesis testing. Two sets of hypothesis tests have been conducted in this part of analysis. In the first set, variances of return during the short-term, medium-term and long-term periods in the post-announcement of interest rate situation have been compared to one another, i.e., variances of returns between $Y_1$ & $Y_2$, $Y_2$ & $Y_3$ and $Y_1$ & $Y_3$ respectively have been examined. These comparisons have been made because variances are expected to rise with the increasing time period and fluctuation in return is also historically found to be continuing in a post-announcement of interest rate situation. The null hypothesis in all the three tests assumes no change in variance, i.e., the variances are equal. The alternative hypothesis $H_1$ is that variance during $Y_2$ is more than that of during $Y_1$, variance during $Y_3$ is more than that of during $Y_2$ and variance during $Y_3$ is more than that of during $Y_1$ respectively.

In second set, each of the post-announcement of interest rate short-term, medium-term and long-term periods has been compared to the long-term pre-announcement of interest rate period, i.e., variances of returns between $X_1$ & $Y_1$, $X_1$ & $Y_2$ and $X_1$ & $Y_3$ respectively have been examined. These tests have been framed with a more empirically established fact that the variances in returns after announcement of interest rate for different periods are expected to be greater than the pre-announcement of interest rate long-term variance. The null hypothesis in all the three tests assumes no change in variance, i.e., the variances are equal. The alternative hypothesis $H_1$, wants to prove that variance during $Y_1$, $Y_2$ and $Y_3$ is more than that of during $X_1$.

### 3.9.2.5 Corporate Fundamental Factors

To determine the factors that are more relevant in determining the stock prices, the use of following step-wise regression has been made:

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + \mu$$

Where, $Y$= Market price, 

$b_0$=Constant term, 

$b_1$ to $b_5$ are the regression coefficients for the respective variables. 

$x_1$= Earnings per share, 

$x_2$= Dividend per share,
\( x_3 = \text{Book value per share}, \)
\( x_4 = \text{Price earning ratio}, \)
\( x_5 = \text{Dividend yield}, \) and
\( \mu = \text{error term}. \)

\( Y \) is dependent variable and \( x_1 \) to \( x_5 \) are independent variables. The independent variables considered for this study have been decided on the basis of existing studies on the subject. Two sets of hypotheses have been formulated.

(A) \( H_{0q}: \) There is no relationship between market price and independent variables (BSE).

\( H_{1q}: \) There exists a relationship between market price and independent variables (BSE).

(B) \( H_{0r}: \) There is no relationship between market price and independent variables (NSE).

\( H_{1r}: \) There exists a relationship between market price and independent variables (NSE).

At the outset, multiple coefficient of determination (R\(^2\)) and adjusted multiple coefficient of determination (Adj. R\(^2\)) were compiled to measure the explanatory power of the multiple regression model used herein. Next, the test of significance of overall multiple regression model was made through F-test. This test has been used to answer the basic question: is there a linear regression relationship between the dependent variable (i.e. stock prices) and any of the independent variables (\( x_i \)) under consideration? To carry out the F-test, analysis of variance (ANOVA) has been performed. With the aim of evaluating the significance of individual regression coefficient (\( \hat{\beta}_i \)), t-test was performed at 5% level of significance.

Durbin-Watson (D.W.) test has been employed to comment on the presence/absence of the problem of auto-correlation in the time series data employed herein.

An important problem that may arise in making inferences about individual regression coefficient is multi-collinearity—the problem of correlations among the independent variables themselves. Due to multi-collinearity the standard errors of the individual slope estimators become unusually high, making the slope coefficients seem statistically not significant (not different from zero). To avoid the problem of
multi-collinearity stepwise-regression is used. The variables causing multi-collinearity were dropped from the model by using backward elimination. The most sophisticated statistical software SPSS 16 has been used to process the data.
3.9.3 Methodology Concerning 3rd Objective

This objective aims to examine a relationship between stock price returns and volatility. The popular member of the GARCH class of models, i.e., EGARCH model has been used to see the asymmetric (bad and good news) impact of stock price volatility on the stock returns. EViews 5 software for model estimation has been used.

3.9.4 Methodology Concerning 5th Objective

In this objective, the aim is to fit an appropriate GARCH model to estimate the conditional market volatility based on S&P CNX Nifty and BSE Sensex. Firstly, the descriptive statistics (Skewness, Kurtosis, Jarque-Bera) for both Sensex and Nifty return have been discussed in order to know whether the daily stock returns are normally distributed or not. Secondly, unit root test is performed using Augmented Dicker-Fuller (ADF) to test whether both the stock return series, i.e., BSE Sensex and S&P CNX Nifty are stationary. Thirdly, the volatility clustering has been investigated by using Box-Jenkins method. Once the volatility clustering is confirmed, then, the most popular member of the ARCH class of models, i.e. GARCH (p,q) model has been used to model volatility of Sensex and Nifty returns. EViews 5 software for model estimation has been used.
REFERENCES


