CHAPTER-2
RESEARCH METHODOLOGY

2.0 INTRODUCTION

Stock Market volatility is unavoidable. It is the nature of the stock markets to fluctuate and turn red and green within short span of time. Volatility is an essential part of the stock market because it checks the nerve of the market. As a coin has two sides, the same way market has two aspects the positive and the negative. Any information in the market will result into changes in prices of any stock which is the cause of fluctuations in the market and hence volatility. At present, stock market is much volatile because of the impact of weak rupee against dollar. Volatility increases during period of recession and it’s essential to know the causes and extent of volatility so that it can be controlled to some extent and future activity can be better. Indian capital market has grown exponentially in the last few decades. Further research provides evidence of volatility caused by a host of factors, including information contained in news, the financial performance of organizations, and even investor behavior. There are evidences which prove that the flow of information found in macroeconomic news and any other public information having a direct impact on stock return volatility (Ross, 1989; Andersen and Bollerslev, 1998; Andersen et al., 2006, among others).

The growth has been in every area from significant transformation brought through electronic form of trading. Capital rose through primary market, market indices and market capitalization etc. that has not changed is high level of volatility. Day-to-day price swings are often large. So, it is felt that there is a need to analyze the volatility and return in the Indian capital market from the perspective of understanding market behavior. Under this situation, present study was conceptualized to re-examine the time variation in volatility.

2.1 NEED OF THE PRESENT STUDY

It can be seen that volatility has its long term impact in the market so an investor is required to take all possible measures to design his portfolio. Stock returns bear a good relationship with volatility as with increase in financial volatility stock prices fluctuates. According to one study, an average investor gets very less returns as compared to the average market returns. So, he is required to understand the fluctuations in the bourses to earn the maximum profits out of his investment.
Information plays a vital role and negative information can have more impact as compared to the positive news of Indian bourses, which happened in 2004 when stock market crashed due to unexpected defeat of NDA or in 2008 in anticipation of recession in USA. Weaken Indian policies and regulations are another reason to be considered with. Government has not taken any major steps to strengthen the economy to protect it from the macroeconomic factors. Therefore, it can be said that persistence level of volatility seems to be in Indian stock market in future as well.

2.2 OBJECTIVES OF THE STUDY
The main objective of the study is to assess whether there is any return and volatility relationship exists in Indian stock market. The purpose is to examine the shifts in stock price volatility and nature of the events that cause the shifts in volatility. To be more precise, following objectives are conceptualized in the study:

1. To see the stock market volatility patterns in Indian stock market and behavior of volatility after the introduction of derivatives.
2. To study the stock price movements to show that any trend or movements in the market are interdependent and to understand the weak form efficiency of the Indian stock market.
3. To identify the day-of-the-week effect and month-of-the-year effect in the Indian stock market.
4. To investigate and compare the stock returns and volatility behavior of the Indian stock market as compared to International stock markets.

2.3 HYPOTHESIS OF THE STUDY
Hypothesis of the study are as follows:

1. There is no predictability in volatility in Indian stock market and volatility is not reduced after the introduction of derivatives.
2. That stock price movements and share price changes on the Indian stock market are not independent which means Indian stock market is not weak form efficient.
3. There is no difference in the returns across the days of the week and month of the year.
4. There is no correlation of returns between Indian stock market and International stock markets.
2.4 RESEARCH DESIGN

2.4.1 Sample Selection

Indian Stock Market is one of the most dynamic and efficient markets in Asia. The Bombay stock exchange and National Stock Exchange are two major stock exchanges in India as most of the share transactions are done by the investors in these two exchanges, Kaur (2004). These exchanges are well equipped with Electronic Trading Platforms and handle large volume of transactions on a daily basis. As on 30th September 2010, there were eighteen indices in NSE and twenty four indices in BSE. But for the purpose of this study, only two indices from each i.e. SENSEX and BSE100 indices in Bombay Stock Exchange and NIFTY and CNX500 indices in National Stock Exchange were considered as sample for this study. These four indices are important in the Indian Stock Market. The NIFTY is well diversified, with 50 stocks accounting for 22 Sectors of the Economy. It is used for Benchmarking Fund Portfolios, Index Based Derivatives and Index Funds. Further, Nifty Stocks represent about 56% of the Free Float Market Capitalization as on September 30th, 2010. The CNX500 is India’s first broad based benchmark. It represents about 90% of the Free Float Market Capitalization and about 87% of the total turnover on the NSE as on Sept 30, 2010. The CNX500 Companies are disaggregated into 72 industry indices viz. S&P CNX Industry Indices. SENSEX is the value-weighted index of the companies listed on the stock exchange.

Bombay Stock Exchange (BSE) in 1986 came out with a stock index that subsequently became the Barometer of the Indian Stock Market. On August 9, 1999, the Bombay Stock Exchange constructed a new index, namely, BSE-500, consisting of 500 Scrips in its basket. BSE-500 Index represents nearly 93% of the total market capitalization on Bombay Stock Exchange Limited. The daily closing prices of the four indices is taken for the period of the study. Brazil, Russia, India and China countries collectively would play an increasingly important role in the global economy. Goldman Sachs (2003) predicted that over the next 50 years, the Brazil, Russia, India and China economies could become a major force in the world economy, and that by 2050 the only developed economies among the six-largest global economies would be the US and Japan in US dollar terms.

According to the IMF report USA, UK, Australia, Japan, Germany, Hongkong stock market and Singapore are among the 36 countries which are classified as advanced economies. So, Brazil, Russia, India and China countries as emerging
economies and seven developed economies named as USA, UK, Australia, Japan, Germany, Hongkong stock market and Singapore are taken to understand their behavior. The data for India stock market is collected from the official websites of National Stock Exchange and Bombay Stock Exchange i.e. www.nseindia.com and www.bseindia.com. The data of other countries is collected from yahoofinance.com. Daily closing prices are converted into returns. The list of markets is given below:

### Table: 2.1 Stock Markets and their Index

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Country</th>
<th>Stock Market</th>
<th>Indices</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Brazil</td>
<td>Sao Paulo Stock Exchange</td>
<td>BOVESPA</td>
<td>BVSP</td>
</tr>
<tr>
<td>2</td>
<td>Russia</td>
<td>Russia Trading System</td>
<td>RTSI</td>
<td>RTSI</td>
</tr>
<tr>
<td>3</td>
<td>India</td>
<td>Bombay Stock Exchange</td>
<td>SENSEX</td>
<td>SENSEX</td>
</tr>
<tr>
<td>4</td>
<td>China</td>
<td>Shanghai Composite Index</td>
<td>SCI</td>
<td>SCI</td>
</tr>
<tr>
<td>5</td>
<td>Hongkong stock market</td>
<td>Hongkong stock market Stock Index</td>
<td>Hang Seng</td>
<td>HANGSENG</td>
</tr>
<tr>
<td>6</td>
<td>Singapore</td>
<td>Straits Time Index</td>
<td>Straits Time</td>
<td>STI</td>
</tr>
<tr>
<td>7</td>
<td>Japan</td>
<td>Tokyo Stock Exchange</td>
<td>Nikkei 225</td>
<td>NIKKEI</td>
</tr>
<tr>
<td>8</td>
<td>Australia</td>
<td>Australian Stock Exchange</td>
<td>AORD</td>
<td>AORD</td>
</tr>
<tr>
<td>9</td>
<td>Germany</td>
<td>Deutreher Aktien Index</td>
<td>DAX</td>
<td>DAX</td>
</tr>
<tr>
<td>10</td>
<td>USA</td>
<td>Dow Jones</td>
<td>Dow Jones Industrial Average</td>
<td>DJIA</td>
</tr>
<tr>
<td>11</td>
<td>UK</td>
<td>Financial Time Stock Exchange</td>
<td>FTSE</td>
<td>FTSE</td>
</tr>
</tbody>
</table>

### 2.4.2 Sources of Data

The present study mainly depended upon secondary data and used daily index closing values. The required information of every day’s closing values was collected from the websites of respective stock exchanges (www.nseindia.com, www.bseindia.com, www.moneycontrol.com, www.allstocks.com, and www.yahoofinance.com). The software used for analyzing the data is eviews5. The other relevant information for this study was collected from different Websites, Journals, and Books.
2.4.3 Study Period

The study comprised for a period of ten years from January 2003 to December 2012. Engle and Mezrich (1995) suggested that at least eight years of data should be used for proper GARCH estimation. This time period carried lot of significance as major ups and downs happened in the Indian stock market during this period which affected the Indian economy. The Indian stock market fluctuated drastically during this time period. On 17th May, 2004 SENSEX crashed 842 points because of unexpected defeat of the NDA. SENSEX fell 1,111 on May 22, 2006, over 1,700 points on October 17, 2007 and on January 21, 2008 stock market crashed by over 1430 points due to the fear of the United States' economy going into a recession. The SENSEX dropped by 749.05 points on 7 January 2009, when the Satyam fraud came to light. The SENSEX on 15 October 2007 crossed the 19,000 mark for the first time. These years were also important from the point of view of some policy regulations done by the SEBI to control the irregularities of stock market like in 2004, SEBI prohibits issue of PNs to Indians, non-resident Indians or persons of Indian origin and in 2007, SEBI proposed to curb participatory notes. So, these years are very important for the Indian stock market and the present study is done to know the volatility of Indian stock market during this period.

To control the volatility of stock market, derivatives were introduced in the Indian stock market in 2001. It was seen whether the introduction of derivative has impact on the volatility. For this purpose, the pre-derivative and post-derivative time period were taken to see the volatility behavior during both periods. The whole period was from January 1992 to December 2012, pre-derivative period was selected from January 1992 to March 2001 and post-derivative period was selected from April 2001 to December 2012.

2.5 DATA ANALYSIS

The secondary data collected is analyzed with the help of different statistical tools. These statistical tools are as follows:

2.5.1 Statistical tools used to see the trends in stock market returns and volatility patterns in post liberalization period.

To calculate the returns, logarithmic difference of two periods is taken by using the following:

\[ R_t(\ln(P_t) - \ln(P_{t-1})) \]
where $R_t$ is the return in period $t$, $P_t$ and $P_{t-1}$ are the daily closing prices of the index at time $t$ and $t-1$ respectively.

2.5.1.1 Unit Root Test

For testing stationarity, let us consider an AR (1) model:

$$Y_t = p_1 Y_{t-1} + \epsilon_t$$  

The simple AR (1) model represented in above equation is called a random walk model. In this AR (1) model if $|p_1|<1$, then the series is I(0) i.e. stationary in level, but if $p_1 = 1$ then there exist what is called unit root problem. In other words, series is non-stationary. Most economists think that differencing is warranted if estimated $p > 0.9$; some would difference when estimated $p > 0.8$. Besides this, there are some formal ways of testing for stationarity of a series.

2.5.1.1.1 Augmented Dickey Fuller Test

Dickey-Totaler test involve estimating regression equation and carrying out the hypothesis test. The simplest approach to testing for a unit root is with an AR(1) model. AR(1) process:

$$\Delta y_t = c + p y_{t-1} + \epsilon_t$$

where $c$ and $p$ are parameters and is assumed to be white noise. If $-1<p<1$, then $y$ is a stationary series while if $p=1$, $y$ is a non-stationary series. If the absolute value of $p$ is greater than one, the series is explosive. Therefore, the hypothesis of a stationary series is involves whether the absolute value of $p$ is strictly less than one. The test is carried out by estimating an equation with $y_{t-1}$ subtracted from both sides of the equation:

$$\Delta y_t = c + \gamma y_{t-1} + \epsilon_t$$

The DF test is valid only if the series is an AR(1) process. If the series is correlated at higher order lags, the assumption of white noise disturbances is violated. The ADF controls for higher order correlation by adding lagged difference terms of the dependent variable to the right-hand side of the regression:

$$\Delta y_t = c + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \ldots + \delta_p \Delta y_{t-p} + \epsilon_t$$
This augmented specification is then tested for in this regression.

\[ H_0: \gamma = 0 \]
\[ H_1: \gamma < 0 \]

The empirical research work done earlier has used this tool in their work (Kaur, 2004; Abdalla, 2012; Bordoloi & Shankar, 2008 and Karmakar, 2007).

2.5.1.1.2 Phillips–Perron test

In statistics, the Phillips–Perron test (named after Peter C. B. Phillips and Pierre Perron) is a unit root test. That is, it is used in time series analysis to test the null hypothesis that a time series is integrated of order 1. It builds on the Dickey–Fuller test of the null hypothesis \( \delta = 0 \) in

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

where \( \Delta \) is the first difference operator. Like the augmented Dickey–Fuller test, the Phillips–Perron test addresses the issue that the process generating data for \( y_t \) might have a higher order of autocorrelation than is admitted in the test equation – making \( y_{t-1} \) endogenous and thus invalidating the Dickey–Fuller t-test. Whilst the augmented Dickey–Fuller test addresses this issue by introducing lags of \( \Delta y_t \) as regressors in the test equation, the Phillips–Perron test makes a non-parametric correction to the t-test statistic. The test is robust with respect to unspecified autocorrelation and heteroscedasticity in the disturbance process of the test equation. (Kaur, 2004; Bordoloi & Shankar, 2008 and Karmakar, 2007) used the same for their research work.

2.5.1.1.3 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test

In econometrics, Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used for testing a null hypothesis that an observable time series is stationary around a deterministic trend. Such models were proposed in 1982 by Alok Bhargava. The series is expressed as the sum of deterministic trend, random walk, and stationary error, and the test is the Lagrange multiplier test of the hypothesis that the random walk has zero variance. KPSS type tests are intended to complement unit root tests, such as the Dickey–Fuller tests. By testing both the unit root hypothesis and the stationarity hypothesis, one can distinguish series that appear to be stationary, series that appear to have a unit root, and series for which the data (or the tests) are not
sufficiently informative to be sure whether they are stationary or integrated. (Gupta and Basu, 2007), earlier literature mentioned the use of KPSS.

2.5.1.2 Autocorrelations and ACF (k)

Autocorrelation is one of the statistical tools used for measuring the dependence of successive terms in a given time series. Hence it has been widely used to measure dependence in successive share price changes. Autocorrelation has been the basic tool used to test the weak form of EMH. The autocorrelation function ACF(k) for the time series \( Y_t \) and the \( k \) lagged series \( Y_{t-k} \) is defined (stephen A. DeLurgio, 1998 p- 67) as:

\[
ACF(K) = \frac{\sum^{n}_{t=1}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum^{n}_{t=1}(Y_t - \bar{Y})^2}
\]

Where \( Y \) is the overall mean of the series with \( n \) observations. The SE of ACF (k) is given by:

\[
Se_{ACF(K)} = \frac{1}{\sqrt{n-k}}
\]

When \( n \) is sufficiently large (\( n > 50 \)), the approximate value of the standard error of ACF (k) is given by:

\[
Se_{ACF(K)} = \frac{1}{\sqrt{n}}
\]

To test whether ACF (k) is significantly different from zero, the following distribution of \( t \) is used:

\[
T = \frac{ACF(k)}{Se_{ACF(K)}}
\]

As it is true for random walks, trends are also characterized by extremely high autocorrelation. For both random walk series and series with trends, autocorrelation ACF (k) are very high and decline slowly as the lag value (k) increases. At the same time the ACF (k) of the first difference series (price changes or returns) are statistically insignificant when the series is a random walk series. A random walk series drifts up and down over time. In some situations it may be difficult to judge whether a trend or drift is occurring. Hence to determine whether a series has significant trend or whether it is a random walk, the t-test is applied on the series of first differences. (Bordoloi & Shankar, 2008).
2.5.1.3 Heteroscedasticity

One of the most important issues before applying the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) methodology is to first examine the residuals for evidence of heteroscedasticity. To test for the presence of heteroscedasticity in residuals of KSE index return series, the Lagrange Multiplier (LM) test for ARCH effects proposed by Engle (1982) is applied. In summary, the test procedure is performed by first obtaining the residuals $\epsilon_t$ from the ordinary least squares regression of the conditional mean equation which might be an autoregressive (AR) process, moving average (MA) process or a combination of AR and MA processes; (ARMA) process. For example, in ARMA (1,1) process the conditional mean equation will be as:

$$ r_t = \theta r_{t-1} + \theta_1 \epsilon_{t-1} \quad \text{XI} $$

after obtaining the residuals $\epsilon_t$, the next step is regressing the squared residuals on a constant and q lags as in the following equation:

$$ e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + ... + \alpha_q e_{t-q}^2 + v_t \quad \text{XII} $$

The null hypothesis that there is no ARCH effect up to order q can be formulated as:

$$ H_0 : \alpha_1 = \alpha_2 = ... = \alpha_q = 0 $$

against the alternative:

$$ H_1 : \alpha_i > 0 $$

for at least one $i = 1, 2, ..., q$

The test statistic for the joint significance of the q-lagged squared residuals is the number of observations times the R-squared ($TR^2$) from the regression. $TR^2$ is evaluated against $\chi^2(q)$ distribution. This is asymptotically locally most powerful test.

2.5.1.3.1 The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Model

In this model, the conditional variance is represented as a linear function of its own lags. The simplest model specification is the GARCH (1,1) model:

**Mean Equation**

$$ r_t = \mu + \epsilon_t \quad \text{XIII} $$

**Variance Equation**

$$ \sigma_t^2 = \omega + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \text{XIV} $$
where $\omega > 0$ and $\alpha_1 \geq 0$ and $\beta_1 \geq 0$, and

$r_t = \text{return of the asset at time } t$

$\mu = \text{average return}$

$\varepsilon_t = \text{residual returns, defined as:}$

$\varepsilon_t = \sigma_t z_t$

Where $z_t$ is standardized residual returns (i.e. iid random variable with zero mean and variance 1), and $\sigma^2_t$ is conditional variance. For GARCH (1,1), the constraints $\alpha \geq 0$ and $\beta_1 \geq 0$ are needed to ensure $\sigma^2_t$ is strictly positive. In this model, the mean equation is written as a function of constant with an error term. Since $\sigma^2_t$ is the one–period ahead forecast variance based on past information, it is called the conditional variance. The conditional variance equation specified as a function of three terms:

- A constant term : $\omega$
- News about volatility from the previous period, measured as the lag of the squared residual from the mean equation: $\varepsilon^2_{t-1}$ (the ARCH term)
- Last period forecast variance: $\sigma^2_{t-1}$ (the GARCH term)

The conditional variance equation models the time varying nature of volatility of the residuals generated from the mean equation. This specification is often interpreted in a financial context, where an agent or trader predicts this period’s variance by forming a weighted average of a long term average (the constant), the forecast variance from last period (the GARCH term), and information about volatility observed in the previous period (the ARCH term). If the asset return was unexpectedly large in either the upward or the downward direction, then the trader will increase the estimate of the variance for the next period. The general specification of GARCH is, GARCH (p, q) is as:

$$\sigma^2_t = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon^2_{t-1} + \sum_{j=1}^{p} \beta_i \sigma^2_{t-1}$$

where, $p$ is the number of lagged $\sigma^2$ terms and $q$ is the number of lagged $\varepsilon^2$ term

2.5.1.3.2 The Exponential GARCH (E-GARCH) Model

This model captures asymmetric responses of the time-varying variance to shocks and, at the same time, ensures that the variance is always positive. It was developed by Nelson (1991) with the following specification

$$\ln(\sigma^2_t) = \omega + \beta_1 \ln(\sigma^2_{t-1}) + \alpha_1 \left( \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \frac{2}{\sqrt{\pi}} \right) - \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$
where $\gamma$ is the asymmetric response parameter or leverage parameter. The sign of $\gamma$ is expected to be positive in most empirical cases so that a negative shock increases future volatility or uncertainty while a positive shock eases the effect on future uncertainty. In macroeconomic analysis, financial markets and corporate finance, a negative shock usually implies bad news, leading to a more uncertain future. Consequently, for example, shareholders would require a higher expected return to compensate for bearing increased risk in their investment. Above Equation is an E-GARCH (1,1) model. Higher order E-GARCH models can be specified in a similar way; E-GARCH (p, q) is as follows:

$$
\ln(\sigma_t^2) = \omega + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}^2) + \sum_{i=1}^{q} \alpha_i \left( \frac{\varepsilon_{t-i}}{\sigma_{t-i}} - \frac{2}{\sqrt{\pi}} \right) - \gamma_t \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \tag{XVII}
$$

2.5.1.3.3 The Threshold GARCH (T-GARCH) Model

Another volatility model commonly used to handle leverage effects is the threshold GARCH (or T-GARCH) model. In the T-GARCH (1,1) version of the model, the specification of the conditional variance is:

$$
\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma d_{t-1} \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{XVIII}
$$

Where $d_{t-1}$ is a dummy variable, that is:

$$
d_{t-1} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} \leq 0 \text{ bad news} \\ 0 & \text{if } \varepsilon_{t-1} > 0 \text{ good news} \end{cases}
$$

the coefficient $\gamma$ is known as the asymmetry or leverage term. When $\gamma = 0$, the model collapses to the standard GARCH forms. Otherwise, when the shock is positive (i.e., good news) the effect on volatility is $\alpha_1$, but when the news is negative (i.e., bad news) the effect on volatility is $\alpha_1 + \gamma$. Hence, if $\gamma$ is significant and positive, negative shocks have a larger effect on $\sigma_t^2$ than positive shocks. In the general specification of this model, T-GARCH (p,q), the conditional variance equation is specified as follows:

$$
\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i + \beta_i d_{t-1}) \varepsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{XIX}
$$

$\alpha_i, \gamma_i$ and $\beta_j$ are non-negative parameters satisfying conditions similar to those of GARCH models. (French, Schwert and Stambaugh 1987; Akgiray, 1989; Ballie and DeGennaro, 1990; Lamoureux and Lastrapes, 1990; Corhay and Tourani, 1994; Geyer, 1994; and Sakata and White, 1998) used ARCH, GARCH and their extension models in their research work.
2.5.2 Statistical tools used to analyze the stock price movements to show that any trend or movements in the market are interdependent and to understand the weak form efficiency of the stock market.

2.5.2.1 Run test analysis

Run test is a non-parametric test. This test considers the sign of the price changes and not the values as such. Statistical tests based on theory of runs do not consider the absolute values but consider only their directions. This test does require the specification of the probability distribution. A run is defined as a sequence of price changes of same sign, preceded or followed by price changes of different signs. In case of stock indices or stock prices there are three possible types of price changes, they are: increase or decrease or no change in prices. This implies we can have three types of runs, positive runs, negative runs and no change runs. Under the hypothesis that the successive price changes are independent and the sample proportions of positive, negative and no change runs are unbiased estimates of the population proportions, the expected number of runs can be computed by using the following formula proposed by Wallis and Roberts (1956).

\[ M = \frac{N(N+1) - \sum i^2 n_i^2}{N} \]

Where

\( M \) = Expected number of runs, \( n_i \) = Number of price changes of each sign (i=1,2,3)\( N \) = Total number of price changes. The standard error of expected number of runs of all signs is given by:

\[ \alpha = \left[ \frac{\sum_{i=1}^{3} n_i^2 \left( \sum_{i=1}^{3} n_i^2 + N(N+1) \right) - 2N \sum_{i=1}^{3} n_i^3 - N^3}{N^3(N+1)} \right] \]

When \( N \) is sufficiently large, the sampling distribution of expected number of runs of all types is approximately normally distributed with mean \( M \) and standard error \( \alpha \). The difference between the actual number of runs and expected number of runs is expressed by standard normal variable \( Z \). Where \( R \) is the total number of observed runs of all signs. (Sharma and Mahendru, 2009; Sekar and Arasu, 2007; and Aggarwal, 2012) included runs test in their study.

2.5.2.2 Kolmogorov–Smirnov test

In statistics, the Kolmogorov–Smirnov test (K–S test) is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to
compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test). The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution (in the two-sample case) or that the sample is drawn from the reference distribution (in the one-sample case). In each case, the distributions considered under the null hypothesis are continuous distributions but are otherwise unrestricted. The two-sample KS test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples. The empirical distribution function $F_n$ for $n$ iid observations $X_i$ is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i} \leq x$$

where $I_{X_i} \leq x$ is the indicator function, equal to 1 if $I_{X_i} \leq x$ and equal to 0 otherwise. Earlier, Poshakwale (1996); and Aggarwal (2012); used K-S test for their research.

### 2.5.2.3 Variance Ratio Test

The variance ratio test (Lo and MacKinlay 1988) is used to measure the randomness of markets pre and post liberalization. The test is based on one of the properties of the random walk process, specifically that the variance of the random walk increments must be a linear function of a time interval, say $q$. The variance ratio is computed by dividing the variance of returns estimated from longer intervals by the variance of returns estimated from shorter intervals, (for the same measurement period), and then normalizing this value to one by dividing it by the ratio of the longer interval to the shorter interval. For independent identically distributed returns $r_t$, $q$ the variance $Var(r_t, q)$ must be equal to $q$ times the variance of $r_t, 1$. A variance ratio that is greater than one suggests that the returns series is positively serially correlated or that the shorter interval returns trend within the duration of the longer interval. A variance ratio that is less than one suggests that the return series is negatively serially correlated or that the shorter interval returns tend toward mean reversion within the duration of the longer interval. Earlier research work used this model (Mishra et al., 2011; Madhusoodanan, 1998; and Sekar & Arasu, 2007).
The variance ratio is calculated as:

\[ VR (q) = \frac{\text{Var}(r_{t+q})}{\text{qVar}(r_t)} \]  

We need to calculate the variance of the longer and the shorter horizons using the following formulae:

\[ \sigma^2(q) = \frac{1}{m-q} \sum_{k=q}^{n-q} (P_k - P_{k-q} - q\mu)^2 \]  

\[ \sigma^2(1) = \frac{1}{n-1} \sum_{k=q}^{n-q} (P_k - P_{k-1} - \mu)^2 \]

2.5.3 Statistical tools used in identifying the day-of-the-week effect and month-of-the-year effect in India stock market.

2.5.3.1 Dummy Variable Regression

To test the presence of seasonality in stock returns of the series, dummy variable regression model is used. This technique is used to quantify qualitative aspects such as race, gender, religion and after that one can include as another explanatory variable in the regression model. The variable which takes only two values is called dummy variable. They are also called categorical, indicator or binary variables in literature. While 1 indicates the presence of an attribute and 0 indicates absence of an attribute. There are mainly two types of model namely ANOVA and ANCOVA. The study uses ANOVA model. Analysis of variance (ANOVA) model is that model where the dependent variable is quantitative in nature and all the independent variables are categorical in nature. To examine the weekend effect and days of the week effect, the following dummy variable regression model is specified as follows:

\[ \text{Returns} = \alpha + \beta_1 \text{Monday} + \beta_2 \text{Tuesday} + \beta_3 \text{Wednesday} + \beta_4 \text{Thursday} + \mu \]

The variables Monday, Tuesday, Wednesday and Thursday are defined as:

Monday = 1 if trading day is Monday; 0 otherwise
Tuesday = 1 if trading day is Tuesday; 0 otherwise,
Wednesday = 1 if the trading day is Wednesday; 0 otherwise
Thursday = 1 if the trading day is Thursday; 0 otherwise

\[ \alpha \] represents the return of the benchmark category which is Friday in our study. Similarly, to find whether there are monthly effects in NIFTY returns, we used ANOVA model specified below as:

\[ \text{Returns} = \alpha + \beta_1 \text{June} + \beta_2 \text{July} + \beta_3 \text{Aug} + \beta_4 \text{Sep} + \beta_5 \text{Oct} + \beta_6 \text{Nov} + \beta_7 \text{Dec} + \beta_8 \text{Jan} + \beta_9 \text{Feb} + \beta_{10} \text{Mar} + \beta_{11} \text{Apr} + \mu \]
where \( Y \) = Monthly returns of NIFTY

- \( D_1 = 1 \) if the month is June; 0 otherwise
- \( D_2 = 1 \) if the month is July; 0 otherwise
- \( D_3 = 1 \) if the month is August; 0 otherwise
- \( D_4 = 1 \) if the month is September; 0 otherwise
- \( D_5 = 1 \) if the month is October; 0 otherwise
- \( D_6 = 1 \) if the month is November; 0 otherwise
- \( D_7 = 1 \) if the month is December; 0 otherwise
- \( D_8 = 1 \) if the month is January; 0 otherwise
- \( D_9 = 1 \) if the month is February; 0 otherwise
- \( D_{10} = 1 \) if the month is March; 0 otherwise
- \( D_{11} = 1 \) if the month is April; 0 otherwise

\( \alpha \) represents the mean return on the May month where as \( \beta_1 \) to \( \beta_{11} \) indicates the shift in mean returns across months. Statistically significant values of \( \beta \)'s imply significant shifts in mean monthly returns, thus confirming the existence of the month of the year effect. The problem with this approach is that disturbance error term may have autocorrelation. Besides this, residual may contain ARCH effect. Therefore, autocorrelation and ARCH effect is required to be tested in residual. Aggarwal and Tondon (1994); Swami (2011); and Sah(2010) used dummy regression model in their research work.

### 2.5.3.2 GARCH Models with Dummy Regression

And, in order to address the problem of heteroscedasticity, variances of the errors to be time dependent to include a conditional heteroscedasticity that captures time variation of variance in stock returns. Thus, error terms now have a mean of zero and a time changing variance of \( h_t^2 \) that is, \( \varepsilon \sim (0, h_t^2) \). To do so, the Generalized Autoregressive Conditional Heteroscedastic (GARCH) model, proposed initially by Engle (1982) and further developed by Bollerslev (1986) is used. In presenting GARCH models, there are two distinct equations, the first for the conditional mean and the second one for the conditional variance. In this model, the conditional variance is represented as a linear function of a long term mean of the variance, its own lags and the previous realized variance. Sah (2010); Sewraj et al. (2010); and Keong et al. (2010) used GARCH to check the seasonality in their work. The general specification of GARCH, GARCH \((p, q)\) is given by:
- Mean equation: \[ r_t = \mu + \varepsilon_t \] \text{XXVIII}
- Variance equation:
  \[ \sigma_t^2 = \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \] \text{XXIX}

where \( w > 0 \). And \( \alpha_t \geq 0 \) and \( \beta_t \geq 0 \), and: \( r_t \) = return of the asset at time \( t \), \( \mu \) = average return, \( \varepsilon_t \) = residual returns, defined as: \( \varepsilon_t = \sigma_t z_t \). Where \( z_t \) is standardized residual returns (i.e. iid random variable with zero mean and variance 1), and \( \sigma_t^2 \) is conditional variance.

2.5.4 Statistical tools used to investigate and compare the stock returns and volatility behavior of the Indian stock market as compared to International stock markets.

2.5.4.1 Correlation Test
The correlation coefficient of two variables in a data sample is their covariance divided by the product of their individual standard deviations. It is a normalized measurement of how the two are linearly related. Formally, the sample correlation coefficient is defined by the following formula, where \( s_x \) and \( s_y \) are the sample standard deviations, and \( s_{xy} \) is the sample covariance, (Pearson; 1930).

\[ r_{xy} = \frac{s_{xy}}{sxsy} \] \text{XXX}

Ranpura (2011); Aktan et al. (2009); and Mukherjee (2011) applied correlation test for their research work.

2.5.4.2 Co-integration
If 2 or more than 2 variables are integrated of the same order where \( d > 0 \) and there exists a stationary linear combinations of these variables, the variables are said to be co-integrated.

2.5.4.2.1 Engle-Granger Test for Co-integration in a Bivariate Process
Let \( x \) and \( y \) are I(1)

\[ Y_t = \alpha + \beta x_t + \mu_t \] \text{XXXI}

If \( u(t) \) is stationary, \( x \) and \( y \) are co-integrated, the regression is co-integration equation. If \( u(t) \) is non-stationary, \( x \) and \( y \) are not co-integrated and regression is spurious. Subha and Nambi (2010); and Paramati et al. (2012); used Engle-Granger Test for research work.
### 2.5.4.2.2 Johansen Co-integration Test

In statistics, the **Johansen test**, named after Søren Johansen, is a procedure for testing co-integration of several time series. This test permits more than one co-integrating relationship so is more generally applicable than the Engle–Granger test which is based on the Dickey–Fuller (or the augmented) test for unit roots in the residuals from a single (estimated) co-integrating relationship.

\[
\Delta Y_t = \theta_1 (y_t - \beta x_t)_{t-1} + \sum_{i=1}^{p} \gamma_{1i} \Delta Y_{t-i} + \sum_{j=1}^{p} \delta_{1j} \Delta x_{t-j} + \epsilon_{1t} \quad XXXII
\]

\[
\Delta x_t = \theta_2 (y_t - \beta x_t)_{t-1} + \sum_{i=1}^{p} \gamma_{2i} \Delta Y_{t-i} + \sum_{j=1}^{p} \delta_{2j} \Delta x_{t-j} + \epsilon_{2t} \quad XXXIII
\]

Bose (2005); and Paramati et al. (2012); applied Johansen test in research.

### 2.5.4.3 Granger causality test

A technique for determining whether one time series is useful in forecasting another. Ordinarily, regressions reflect "mere" correlations, but Clive Granger, who won a Nobel Prize in Economics, argued that there is an interpretation of a set of tests as revealing something about causality. If two variables, x and y are correlated, it is possible that: X is caused by y, Y is caused by x, Both x and y are caused by some other variable z. Causality cannot be inferred from contemporaneous correlations. Granger Causality is based on the simple logic that effect cannot precede cause. We may have following situations:

- No causality between x and y
- One way causality
- Two-way causality or feedback causality

Granger causality measures precedence and information content but does not by itself indicate causality in the more common use of the term. Let y and x be stationary time series. To test the null hypothesis that x does not Granger-cause y, one first finds the proper lagged values of y to include in a univariate auto regression of y:

\[
Y_t = \alpha + \sum_{i=1}^{l} \alpha_i Y_{t-i} + \sum_{j=1}^{l} \beta_j x_{t-j} + \epsilon_t \quad XXXIV
\]

\[
X_t = \omega + \sum_{i=1}^{l} \gamma_i Y_{t-1} + \sum_{j=1}^{l} \theta_j Y_{t-j} + \epsilon_t \quad XXXV
\]

Bose (2005); Ranpura (2011); & Paramati et al. (2012); applied Granger Causality Test earlier.
2.5.4.4 Vector autoregressive models

Vector autoregressive models (VARs) were popularized in econometrics by Sims (1980) as a natural generalization of univariate autoregressive models. VARs have often been advocated as an alternative to large-scale simultaneous equations structural models. The simplest case that can be entertained is a bivariate VAR, where there are only two variables, \( y_{1t} \) and \( y_{2t} \), each of whose current values depend on different combinations of the previous \( k \) values of both variables, and error terms,

\[
\begin{align*}
    y_{1t} &= \beta_{10} + \beta_{11} y_{1t-1} + \cdots + \beta_{1k} y_{1t-k} + \alpha_{11} y_{2t-1} + \cdots + \alpha_{1k} y_{2t-k} + u_{1t} \\
    y_{2t} &= \beta_{20} + \beta_{21} y_{2t-1} + \cdots + \beta_{2k} y_{2t-k} + \alpha_{21} y_{1t-1} + \cdots + \alpha_{2k} y_{1t-k} + u_{2t}
\end{align*}
\]

where \( u_{it} \) is a white noise disturbance term with \( \mathbb{E}(u_{it}) = 0 \), \( i = 1, 2 \), \( \mathbb{E}(u_{1t}u_{2t}) = 0 \). As should already be evident, an important feature of the VAR model is its flexibility and the ease of generalization. For example, the model could be extended to encompass moving average errors, which would be a multivariate version of an ARMA model, known as a VARMA. Instead of having only two variables, \( y_{1t} \) and \( y_{2t} \), the system could also be expanded to include \( g \) variables, \( y_{1t}, y_{2t}, y_{3t}, \ldots, y_{gt} \), each of which has an equation. Another useful facet of VAR models is the compactness with which the notation can be expressed. For example, consider the case from above where \( k = 1 \), so that each variable depends only upon the immediately previous values of \( y_{1t} \) and \( y_{2t} \), plus an error term. This could be written as

\[
\begin{align*}
    y_{1t} &= \beta_{10} + \beta_{11} y_{1t-1} + \alpha_{11} y_{2t-1} + u_{1t} \\
    y_{2t} &= \beta_{20} + \beta_{21} y_{2t-1} + \alpha_{21} y_{1t-1} + u_{2t}
\end{align*}
\]

2.5.4.4.1. VARs with exogenous variables

Consider the following specification for a VAR (1) where \( X_t \) is a vector of exogenous variables and \( B \) is a matrix of coefficients:

\[
Y_t = A_0 + A_1 Y_{t-1} + B X_t + e_t
\]

The components of the vector \( X_t \) are known as exogenous variables since their values are determined outside of the VAR system -- in other words, there are no equations in the VAR with any of the components of \( X_t \) as dependent variables. Such a model is sometimes termed a VARX, although it could be viewed as simply a restricted VAR where there are equations for each of the exogenous variables, but with the coefficients on the RHS in those equations restricted to zero. Such a restriction may
be considered desirable if theoretical considerations suggest it, although it is clearly not in the true spirit of VAR modeling, which is not to impose any restrictions on the model but rather to ‘let the data decide’. Paramati et al. (2012); Mukherjee (2011); and Singh and Sharma (2012); applied VAR test in their earlier studies to see the dependence between variables.

2.5.4.5 Variance Decompositions and Impulse Responses

Block $F$-tests and an examination of causality in a VAR will suggest which of the variables in the model have statistically significant impacts on the future values of each of the variables in the system. But $F$-test results will not, by construction, be able to explain the sign of the relationship or how long these effects require to take place. That is, $F$-test results will not reveal whether changes in the value of a given variable have a positive or negative effect on other variables in the system, or how long it would take for the effect of that variable to work through the system. Such information will, however, be given by an examination of the VAR’s impulse responses and variance decompositions. Variance decompositions offer a slightly different method for examining VAR system dynamics. They give the proportion of the movements in the dependent variables that are due to their ‘own’ shocks, versus shocks to the other variables. A shock to the $i$ th variable will directly affect that variable of course, but it will also be transmitted to all of the other variables in the system through the dynamic structure of the VAR. Variance decompositions determine how much of the $s$-step-ahead forecast error variance of a given variable is explained by innovations to each explanatory variable for $s = 1, 2, \ldots$. In practice, it is usually that own series shocks explain most of the (forecast) error variance of the series in a VAR. To some extent, impulse responses and variance decompositions offer very similar information. Singh and Sharma (2012); Aktan et al.(2009); Chattopadhyay and Behera (2006) used variance decompositions in research work done by them.

Impulse responses trace out the responsiveness of the dependent variables in the VAR to shocks to each of the variables. So, for each variable from each equation separately, a unit shock is applied to the error, and the effects upon the VAR system over time are noted. Thus, if there are $g$ variables in a system, a total of $g^2$ impulse responses could be generated. The way that this is achieved in practice is by
expressing the VAR model as a VMA -- that is, the vector autoregressive model is written as a vector moving average. Provided that the system is stable, the shock should gradually die away. To illustrate how impulse responses operate, consider the following bivariate VAR (1). Aktan et al. (2009); and Chattopadhyay and Behera (2006) used impulse responses earlier.

\[ y_t = A_1 y_{t-1} + u \]  

The VAR can also be written out using the elements of the matrices and vectors as

\[
\begin{bmatrix}
Y_{1t} \\
Y_{2t}
\end{bmatrix} = \begin{bmatrix}
0.5 & 0.3 \\
0.0 & 0.2
\end{bmatrix} \begin{bmatrix}
Y_{t-1} \\
Y_{t-2}
\end{bmatrix} + \begin{bmatrix}
u_{1t} \\
u_{2t}
\end{bmatrix}
\]

2.6 STRUCTURE OF RESEARCH WORK

The structure of research work is as follows:

**Chapter 1: Introduction**

This chapter consists of meaning of Stock market volatility, market indices, opening and closing balances, stock markets movements, day of the week effect, global markets, Monday effect, and Friday effect. It also included review of literature to see the gaps regarding returns and volatility of stock markets, stock markets movements, day-of-the-week effect, and integration of international stock markets, stock market patterns and behavior.

**Chapter 2: Research Methodology**

This chapter discusses the objective of the study, hypothesis of the study, research methodology, research hypothesis, data sources, data collection and procedure and tools used for the purpose of analysis of data.

**Chapter 3: Return and Volatility Relationship in Indian Stock Market**

This chapter is concerned with examining the return and volatility relationship by using Descriptive Statistics of Daily Returns, Unit root test, Autocorrelation and GARCH, T-GARCH and E-GARCH models.

**Chapter 4: Random Walk of Indian Stock Market**

This chapter identifies whether Indian stock market are efficient or not. For this purpose, various tools like by Descriptive Statistics of Daily Returns, Unit Root Test, Kolmogorov-Smirnov Test, Runs Test, Autocorrelation Test and Variance Ratio Test are used.
Chapter 5: Seasonal Anomalies in Indian Stock Market
This chapter is regarding identifying the seasonal anomalies i.e. day-of-the-week effect and month-of-the-year effect in Indian stock market. The tools used for this purpose were Descriptive Statistics of Daily Returns, OLS regression model, GARCH (1, 1) model with dummy variables.

Chapter 6: Interdependence of Indian and International Stock Markets
This chapter identifies to what extent Indian stock market are correlated with International stock markets. For this purpose, eleven stock markets are selected from emerging economies and developed economies. These stock markets are Brazil, USA, UK, Germany, Hong Kong, Australia, Russia, India, Japan, Singapore and China. Different tools like Descriptive Statistics of Daily Returns, Unit Root Test, Correlation test, Granger Causality test, Co-integration test and Vector Autoregressive Test are used.

Chapter 7: Summary and Conclusions
This chapter discusses the findings, conclusions and implications of the research work. This also includes the scope of research work based upon analysis of stock market volatility, day of the week effect, weak efficiency and impact of international markets on Indian stock market.
References


