CHAPTER-3
DATA BASE AND RESEARCH METHODOLOGY

The present chapter discusses the database and methodology that has been employed in the present dissertation. It explains in detail the objectives, sample period, sources of data, sample design and research methodology which includes data transformation procedure, statistical tests and software used for data analysis.

3.1 OBJECTIVES OF THE STUDY

The overall objective of the study is to test the Calendar anomalies in Indian stock market. Specific objectives of the study are:

(1) To examine the presence of Week-End/Day-of-the-Week effect in Indian stock market;

(2) To test the existence of January effect/Turn-of-the-Year effect in Indian stock market;

(3) To investigate the Monthly/Turn-of-the-Month effect in Indian stock market;

(4) To examine the existence of Holiday effect in Indian stock market.

3.2 SAMPLE PERIOD OF THE STUDY

The study is based upon the period of fourteen years and three months i.e., from January 1, 1992 through March 31, 2006 for examination of companies listed on Bombay Stock Exchange (BSE). The study further considers the period of eleven years and four months i.e. from November 3, 1994 to March 31, 2006 for companies listed on National Stock Exchange (NSE) with a view to examine the presence of seasonality in Indian stock market. Besides being the most recent period, major changes have been brought in the structure and functioning of the Indian stock market during these years.

Financial market reforms have brought gradual improvement in the functioning of the Indian stock market, that are aimed at enhancing competition, transparency and
efficiency in the Indian financial markets (Sah and Omkarnath 2007). In the wake of the scam of 1992 and the information, communication, and entertainment (ICE) meltdown of 2001, major regulatory activities have taken place during this period. For example, screen based trading was introduced in the year 1994 on the NSE and in 1995 on BSE, circuit filters were introduced by the NSE in 1995, establishment of National Securities Clearing Corporation (NSCC) and National Depository Limited (NSDL) in 1996, abolition of badla transaction and introduction of Rolling Settlement in December 1999, introduction of index based futures in June, 2000 and index option in June, 2001 (Kaur 2004) to improve the capital market efficiency. The scope and activities of stock market have expanded rapidly due to the liberalization drive since 1991. Therefore, the need of the hour is to study the market by considering the impact of all these reforms on Indian stock market.

The study period for companies listed on BSE is bifurcated as:

1. January 1, 1992 to March 31, 2006 (Aggregate time period),
2. January 1, 1992 to December 31, 1995 (Initial years of Post-Liberalization and Globalization era),
3. January 1, 1996 to June 21, 2001 (Pre-Rolling Settlement period),

The study period for NSE is bifurcated as:

1. November 3, 1994 to March 31, 2006 (Aggregate time period),
2. January 1, 1994 to December 31, 1995 (Initial years of Post-Liberalization and Globalization era),
3. January 1, 1996 to June 21, 2001 (Pre-Rolling Settlement period),

The aggregated period is divided into different sub-periods as in first sub-period, the impact of globalization and liberalization era on Indian stock market has been examined.

The second sub-periods considers the efficiency of Indian stock market before the introduction of Rolling Settlement in India.
The third sub-period undertakes the impact of introduction of compulsory Rolling Settlement on Indian stock market.

3.3 SOURCES OF DATA

The main sources of data are given below:

1. CD ROM on daily official list: Price and volume data for the years 1981-2001 of BSE;
2. Financial Dailies:
   (i) The Economic Times, India
   (ii) Financial Express, India
3. Journals / Magazines:
   (i) Business India
   (ii) Capital Market
   (iii) Dalal Street
   (iv) Economic and Political Weekly
   (v) Fortune (Asia Edition)
4. CAPITALINE DATABASE (the online database maintained by the Capital Market Publishers India Private Limited);
5. CAPITACHART DATABASE (the online database maintained by the Capital Market Publishers India Private Limited);
6. PROWESS DATABASE (the online database maintained by the Center for Monitoring of Indian Economy- CMIE);
7. Websites:
   (i) Website of Bombay Stock Exchange: www.bseindia.com
   (ii) Website of National Stock Exchange: www.nseindia.com

3.4 SAMPLE DESIGN OF THE STUDY

The selection of the companies for the purpose of analysis has been done by considering the continuous trading of these companies during the study period and the availability of data. If any company has infrequent trading then that company has not been taken for analysis. The present study has selected the sample of companies which
are listed in the period from January 1, 1992 through December 31, 1995 for BSE and November 3, 1994 through December 31, 1995 for NSE.

However, there are some empirical studies which have employed non-dividend adjusted trading returns and close-to-close share prices (Mittal 1994; Poshakwale 1996; Anshuman and Goswami 2000; Gupta and Aggarwal 2004; Nath and Dalvi 2004; Mangala and Mittal 2005; Sharma and Singh 2006). Moreover, Lakonishok and Smidt (1988) and Fishe et al. (1993) have concluded that any dividend bias which occurs from not employing dividend adjusted returns, is relatively small and will not have impact on the statistical significances of any results (Lakonishok and Smidt 1988).


To investigate the Day-of-the-Week, January/Turn-of-the-Year effect and Monthly/ Turn-of-the-Month effect, six hundred and sixty two companies listed on BSE and two hundred and three companies selected for NSE have been investigated (See appendices). But, it is not possible to investigate the Holiday effect in these companies, because it becomes cumbersome process to detect Week-Day, Pre-Holiday, Intra-Holiday and Post-Holiday for the same sample size. Thus, to avoid this issue, daily closing prices of Sensex has been taken to investigate Holiday effect in Indian stock market with reference to BSE.

3.5 RESEARCH METHODOLOGY

In order to analyze the data and interpret the results, a number of statistical techniques have been used. The present chapter takes into account the basic framework and application of these techniques that has been followed in this study. This section contains the following sub-sections which contribute towards the Research Methodology. Section 3.5.1 describes the statistical tests used in the study;
Section 3.5.2 explains the hypotheses that have been tested in the present study;

Section 3.5.3 briefly explains the limitations of the study.

3.5.1 Statistical tests:

In order to examine the EMH and the presence of Calendar anomalies in Indian stock market, following statistical tests have been applied.

(3.5.1.1) Descriptive statistics (Mean, Standard Deviation, Skewness and Kurtosis);

(3.5.1.2) Augmented Dickey Fuller (ADF) test and Philips-Perron (PP) test.

These can be explained as:

3.5.1.1 Descriptive Statistics

The summary statistics have been computed using Mean, Standard Deviation, Skewness and Kurtosis to develop the research design for the analysis of Indian securities market data. The descriptive statistics includes the following:

- **Mean**: It is the average value of the series obtained by adding up the series and dividing by the number of observations.

- **Maximum and Minimum**: It shows the maximum and minimum values in the series in the current sample.

- **Standard Deviation**: Standard deviation is a measure of the variability or dispersion of a statistical population. A low standard deviation indicates that the data points tend to be very close to the mean whereas high standard deviation indicates that the data are spread out over a large range of values.

\[
S = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N-1}}
\]  

(1)

Where \( N \) is the number of observations in the sample undertaken and \( \bar{Y} \) is the mean of the series. For the analysis purpose, the present study has employed the following equation:
\[ Y = \log P_t - \log P_{t-1} \quad (2) \]

- **Skewness:** Skewness tells the deviation of the distribution from symmetry. The distribution is asymmetrical (positively or negatively skewed), if the skewness is different from zero while normal distributions are perfectly symmetrical. It can be computed as:

\[
S_k = \frac{n \sum_{t=1}^{n} (Y - \bar{Y})^3}{(n-1)(n-2)\sigma^3} 
\]  

(3)

An appropriate approach is to compute the ratio of coefficient of skewness to the standard error, Sk/SE. If \(-2 < Sk/SE < 2\), then it can be concluded that the skewness of the distribution is not significant and it may be concluded that the distribution of return is symmetrical. Otherwise, if this ratio lies outside the range \(\pm 2\), then the distribution of returns is not symmetrical, and hence the empirical distribution may be negatively skewed or positively skewed.

- **Kurtosis:** Kurtosis is the degree of peakedness of a distribution, usually taken relative to a normal distribution. Thus, Kurtosis measures the extent to which a distribution is more peaked or flat-topped than the normal curve. The distribution is either flatter or more peaked than normal, if the kurtosis is different from zero. If the value of the kurtosis is three then it means the distribution follows the standard normal distribution. Kurtosis is computed as follows:

\[
K_u = \frac{n(n+1) \sum_{t=1}^{n} (Y - \bar{Y})^4 - 3 \sum_{t=1}^{n} (Y - \bar{Y})^2 \sum_{t=1}^{n} (Y - \bar{Y})^2(n-1)}{(n-1)(n-2)(n-2)\sigma^4} 
\]  

(4)

Where,

\(\sigma^4\) is the standard deviations raised to the power four,

and \(n\) is the sample size.

There are three types of kurtosis in a distribution:
1. A curve having a high peak than the normal curve is called lepto-kurtic, having too much concentration of the items near the centre. Here, K>3.

2. A curve having a low peak (or flat topped) than the normal curve is platy-kurtic, having less concentration of items near the centre. Here K<3

3. A Curve having normal peak or normal curve itself is meso-kurtic, having equal distribution of items around the central value. Here K=3.

An Appropriate approach is to compute the ratio of coefficient of kurtosis to standard error, Ku/SE. The kurtosis of the distribution is not significant and the distribution of returns is meso-kurtic, if -2<Ku/SE<2.

- **The Jarque-Bera (JB) test:** The JB test of normality is an asymptotic or large sample test (i.e., size N) and first computes the skewness (Sₖ) and kurtosis (k) of the measures of the OLS residuals and uses the following statistics to test the null hypothesis of the normal distribution:

\[ JB = N \left( \frac{S^2}{6} + \frac{(K-3)^2}{24} \right) \]  

(5)

3.5.1.2 Augmented Dickey Fuller (ADF) test and Philips-Perron (PP) test

According to Dickey and Fuller (1979), a time series is stationary if its mean, variance and auto-covariance (at various lags) are similar but hardly makes any difference about the point of measurement. Thus, they are time invariant, whereas on the other hand a non-stationary time series will detect the presence of time-varying mean or a time-varying variance or both. Whereas a stochastic process is purely random or white noise process if it has zero mean, constant variance σ² and is serially uncorrelated.

In any time series analysis the test of stationarity plays an important role because in the presence of non-stationary series the standard estimation procedures are not applicable. Thus, the analysis has been started with testing of stationarity i.e., unit root testing. Generally, it is assumed that time series are stationary indicate the presence of mean and variance are constant and its covariance is time invariant (Leybourne et al.
2003). If a time series is integrated of order 1 [i.e., it is I(1)], its first differences are I(0), i.e. stationary. In the same way, if the time series I(2), its second difference is I(0).

However, if a time series is not stationary then there are methods to make them stationary. The formal procedures to detect unit roots in any given time series is graphical analysis, Autocorrelation (ACF) and Partial Autocorrelation (PACF). For the purpose of the study, ADF and PP tests have been applied to determine unit roots that consist of regressing the first difference of the series against a constant, the series lagged one period and the differenced series at n lag lengths (Pindyck and Rubinfeld 1998). The model used in equation 5 is as follows:

\[
Y_t = \alpha + \sum_{i=1}^{m} \beta_i \Delta Y_{t-i} + \Delta Y_{t-1} + \varepsilon_t
\]  

(6)

Where \(r_t\) = return for the period t; defined as \((\log P_t - \log P_{t-1})\) is the trend variable by taking values 1, 2 and so on. Here, \(P_t\)=Stock price at time t; \(P_{t-1}\)=Stock price at time t-1. I is the order of lagged returns which varies from 1 to n and \(r_{t-1}\) is the one period lagged value of the variable \(r\).

3.5.2.1 Unit Root Testing

The Unit Root Testing is conducted by adding the lagged values of the dependent variable i.e., return \(\Delta Y_t\)

\[
Y_t = \log P_t - \log P_{t-1}
\]  

(7)

It also represents average growth of share price between two periods and thus, can be justified as the measure of return in Indian share market. It consists of estimating the following regression equation in case of Day-of-the-Week effect. It is to be checked that if there is any Day-of-the-Week effect in both high frequency as well as the close to close return series. For testing the Day-of-the-Week effect, dummy variables have been used. The assigned values of 1, 2, 3 and 5 for Tuesday, Wednesday, Thursday and Friday respectively (leaving out 1 day for robustness of the regression results) as dummy variable values and designed the equation as below to test the Day-of-the-Week effect:
\[ \Delta \log P_t = \beta_1 + \varphi \log P_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta \log P_{t-i} + \partial_1 D_{\text{Tues}} + \partial_2 D_{\text{Wed}} + \partial_3 D_{\text{Thurs}} + \partial_4 D_{\text{Fri}} + \varepsilon_t \quad (8) \]

Where,
\[
D_i = \begin{cases} 
1: \text{i}^{\text{th}} \text{ day} & \text{; } i = \text{Tues}, \text{Wed}, \text{Thurs}, \text{Fri} \\
0: \text{otherwise} 
\end{cases}
\]

The following equation will be estimated in case of January effect.

\[ \Delta \log P = \beta_1 + \varphi \log P_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta \log P_{t-i} + \partial_{D \text{Feb}} D_{\text{Feb}} + \partial_{D \text{Mar}} D_{\text{Mar}} + \partial_{D \text{Apr}} D_{\text{Apr}} + \partial_{D \text{May}} D_{\text{May}} + \partial_{D \text{Jun}} D_{\text{Jun}} + \partial_{D \text{Jul}} D_{\text{Jul}} + \partial_{D \text{Aug}} D_{\text{Aug}} + \partial_{D \text{Sep}} D_{\text{Sep}} + \partial_{D \text{Oct}} D_{\text{Oct}} + \partial_{D \text{Nov}} D_{\text{Nov}} + \partial_{D \text{Dec}} D_{\text{Dec}} + \varepsilon_t \quad (9) \]

Where,
\[
D_i = \begin{cases} 
1: \text{i}^{\text{th}} \text{ month} & \text{; } i = \text{Feb}, \text{Mar}, \text{Apr}, \text{May}, \text{June}, \text{July}, \text{Aug}, \text{Sept}, \text{Oct}, \text{Nov}, \text{Dec} \\
0: \text{otherwise} 
\end{cases}
\]

The following equation will be estimated in case of Holiday effect

\[ \Delta \log P_t = \beta_1 + \varphi Y_{t-1} + \sum_{i=1}^{m} \alpha_i Y_{t-i} + \partial_1 D_{\text{WD}} + \partial_2 D_{\text{PD}} + \partial_3 D_{\text{ID}} + \partial_4 D_{\text{POD}} + \varepsilon_t \quad (10) \]

Where,
\[
D_i = \begin{cases} 
1: \text{i}^{\text{th}} \text{ holiday} & \text{; } i = \text{WD, ID, POD} \\
0: \text{otherwise} 
\end{cases}
\]

Where, \( R_t \) represents the daily returns at time \( t \), \( D_i \) in equation (12) is a dummy variable that represents each day (i.e., Monday, Tuesday,…….., Friday) for Day-of-the-Week effect, each month (i.e., January, February, …………, December) in the January/Turn-of-the-Year effect and the day before (i.e., Week Day, Pre-Holiday, Intra-Holiday, Post-Holiday) a holiday in the Holiday effect respectively. It takes the value of one to mark the various calendar conditions and zero otherwise.

If average of return is statistically significant then we conclude that the structural break is significant and the dummy representing the structural break is significantly causing seasonality.
3.5.2 Research Hypotheses

The various hypotheses tested in the present study are:

(1) **To examine the daily returns for Week-End/Day-of-the-Week effect:** For this purpose, two hypotheses have been drawn:

Null hypothesis (H₀)

\[ H₀ : \bar{r}_1 = \bar{r}_2 = \ldots = \bar{r}_5 \]  \hspace{1cm} (11)

Here, \( \bar{r}_1, \bar{r}_2 \) represent the average returns of different trading days of the week. The null hypothesis implies that there is no significant difference in the mean trading returns across different trading days in a week.

Alternative hypothesis (H₁)

\[ H₁ : \bar{r}_1 \neq \bar{r}_2 \neq \ldots \neq \bar{r}_5 \]  \hspace{1cm} (12)

This implies that there is significant difference in the mean returns across the trading days in a week. Descriptive statistics have been applied to study the distribution pattern of the average daily returns across the week.

(2) **To examine monthly returns for January effect:** For this purpose, returns are calculated in the following way:

\[ \bar{r}_{mt} = \frac{P_{mt} - P_{mt-1}}{P_{mt-1}} \]  \hspace{1cm} (13)

Where:

\[ \bar{r}_{mt} = \text{Monthly return on stock prices for the period ‘t’} \]

\[ P_{mt} = \text{Closing price of the stock for the month ‘t’} \]

\[ P_{mt-1} = \text{closing price of the stock for the preceding month.} \]

To test the January effect, the mean monthly returns have been compared across different months in a year. The following null hypothesis has been tested against the alternative hypothesis:
Here, $\bar{r}_{m1}$, $\bar{r}_{m2}$ represent mean returns for different months of a year. The null hypothesis implies that there is no significant difference in mean monthly returns across different months of a year.

$H_0 : \bar{r}_{m1} = \bar{r}_{m2} = \ldots = \bar{r}_{m12}$

(14)

This implies that the mean returns for the month of January are significantly greater than the average returns of the remaining eleven months of the year.

(3) To examine monthly returns for Monthly/Turn-of-the-Month effect: The trading day approach has been used to investigate the Turn-of-the-Month pattern in the stock returns. The following null hypothesis has been tested against the alternate hypothesis suggested by Ariel (1987).

$H_0 : \text{Mean daily returns in the first half of the trading month (-1 to 8) are equal to the mean daily returns in the second half of the trading month (9 to -2)}$ against the alternate hypothesis;

$H_1 : \text{Mean daily returns in the first half of the trading month (-1 to 8) are more than the mean daily returns in the second half of the trading month (9 to -2)}$.

The first and the second half of each month consist of -1, 1, 2, …., 8 trading days and 9, -9, …., -2 trading days respectively. Here, 1 represents the first trading day of a month and -1 represents the last trading day of the previous month. The days which do not fall in the said interval are ignored. The last trading day of the previous month has been included in the following month as suggested by Ariel (1987). However, first fifteen and last fifteen trading days of the months of all the subsequent year have been taken for the analysis of Monthly effect.

(4) To examine monthly returns for Turn-of-the-Year effect: Turn-of-the-year effect has been studied after taking the mean returns for five days, ten days, fifteen days and twenty days prior to and after the new year (Mangala and Mittal 2004).
The following null hypothesis has been tested against the alternate hypothesis:

\( H_0 \): Mean returns for the first five trading days in January are equal to the mean returns on other trading days against the alternate hypothesis;

\( H_1 \): Mean returns for the first five trading days in January are more than the mean returns on other trading days.

(5) **To examine Holiday effect:** Following Arumugam (2001) the trading days have been classified into four categories such as Week-Day, Pre-Holiday, Intra-Holiday and Post-Holiday.

(a) **Weekdays:** Week days are those trading days which have both at least one preceding as well as one succeeding day as trading days.

(b) **Pre-Holidays:** Pre-Holidays are those trading days which have at least one preceding day as trading day, but at least one succeeding day as holiday.

(c) **Post-Holidays:** Post-Holidays are those days which have at least one preceding day as holiday, but at least one succeeding day as trading day.

(d) **Intra day:** Intra-Holidays are those days which have at least one preceding as well as at least one succeeding day as holidays.

The following null hypothesis has been tested against the alternate hypothesis:

\( H_0 \): Mean returns on the days before holidays are equal to the mean returns on other trading days against the alternate hypothesis;

\( H_1 \): Mean returns on the days before holidays are more than the mean returns on other trading days.

3.5.3 **Softwares used for analysis:**

For the purpose of analysis, three softwares have been used:

- SPSS 13.0
- Easy-Reg International
- E-Views 5.1
3.5.4 LIMITATIONS OF THE STUDY

The present study is subjected to certain limitations and constraints. Some of its limitations are as under:

1. The present study has examined only equity segment of the stock market and doesn’t examine the debt and derivative segments of the Indian stock market.

2. There are time and resource constraints.

3. The study is based on secondary data collected from a number of sources. The data has been collected from authentic sources like the databases i.e., Prowess and Capitaline, but even then the validity of results is subject to the accuracy of the databases.

4. For the purpose of analysis, different statistical techniques have been used. Each statistical technique or tool has its own limitations in terms of assumptions. Thus, the present study might face the limitations posed by the statistical tools used.

5. In order to examine the seasonality in Indian stock markets, present study has employed daily closing prices whereas use of high frequency data could provide more authenticated results.

6. The present study has assumed that traders do not consider the transaction cost.

7. The study undertakes only closing prices of securities whereas open, high and low prices can also be used to test the efficiency of Indian securities market.

8. The companies that are not listed have not been taken for the analysis in order to examine the efficiency in Indian stock market.