CHAPTER 4
DATA DESCRIPTION AND THEORETICAL FRAMEWORK

This research study explores the linkage between Foreign Institutional Investor (FII) investments and Indian stock market. To capture the linkage, the entire research study has been sub-divided into four major parts based on the research questions and objectives framed. This research uses the data series of different macroeconomic and stock market variables for this purpose. The type of data and its frequencies vary from one framework to another to meet the ultimate purpose of each framework undertaken in the study. The following are the description of data series used in the four major framework of the research study.

4.1 RELATIONSHIP BETWEEN FII INVESTMENTS AND INDIAN STOCK MARKET

This part of the research uses the daily data to identify the relationship between FII investment flows and Indian stock market. FIIs net equity investment, closing value of NSE Nifty index and closing value of CNX Nifty 500 are considered to evaluate the relationship. Table 4.1 describes the data features of this section of research study.
Table 4.1 Data description for Relationship Study

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Unit of Measurement</th>
<th>Source cited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net FII equity investment (FII)</td>
<td>Daily</td>
<td>Rs. (Indian Rupee) in crore</td>
<td><a href="http://www.sebi.gov.in">www.sebi.gov.in</a></td>
</tr>
<tr>
<td>NSE Nifty index (Nifty)</td>
<td>Daily</td>
<td>Index closing point (base year:1995)</td>
<td><a href="http://www.nseindia.com">www.nseindia.com</a></td>
</tr>
<tr>
<td>NSE CNX Nifty 500 index (Nifty 500)</td>
<td>Daily</td>
<td>Index point (base period: 1995)</td>
<td><a href="http://www.nseindia.com">www.nseindia.com</a></td>
</tr>
</tbody>
</table>

4.1.1 Rationale behind the selection of Stock Exchange and Indices

4.1.1.1 National Stock Exchange

NSE of India was set up with the objective of providing a fully automated screen-based trading system with national spread. The Exchange has brought great transparency, speed & efficiency apart from creating market integrity. The NSE uses state-of-art information technology to deliver an efficient and transparent trading, clearing and settlement mechanism. As of June 2014, NSE has a market capitalization of Rs. 88,54,702 crore. According to World Federation of Exchange Market Highlights 2013, NSE is the largest stock exchange in the world in the number of Electronic Order Book (EOB) trades in 2013. As per SEBI annual report, NSE accounts for Rs. 28,08,488 crore of out of Rs. 33,41,337 crore total turnover which is about 84.1% of total turnover value in 2013-14. Moreover FII’s activity in NSE is much more significant (22.9% of NSE’s cash market turnover) than BSE (only 4.8% of BSE’s cash market turnover).
4.1.1.2 Nifty index

The benchmark CNX Nifty is a well diversified 50 stock index which represents 23 sectors of the economy. The CNX Nifty Index represents about 66.85% of the free float market capitalization of the stocks listed on NSE as on June 30, 2014. In 2013-14, CNX Nifty had 62.31% of total market capitalization of all listed companies in NSE.

4.1.1.3 Nifty CNX 500

NIFTY CNX 500 is a broad based index (represents 93% of the total turnover on NSE) used to capture the FIIs relationship with stocks beyond the 50 large market capitalized blue-chip companies listed in stock exchange. This helps the researcher to know whether FIIs are having the same level of relationship with NIFTY 50 and CNX NIFTY 500.

4.1.2 Theoretical Framework for the Study

4.1.2.1 Normal distribution of data

According to Central Limit Theorem, the term ‘normal distribution’ refers to a particular way in which observations will tend to pile up around a specific value rather than be spread evenly across a range of values. This is most required and applicable for the continuous data. Graphically, normal distribution is best described by a bell-shaped curve. The simplest way to assess the normality of the data set is to plot the data and look at the frequency distribution histogram.

Though several normality tests are available, the study uses the Jarque-Bera (JB) test to check whether the selected variables are normally distributed. The JB test of normality is an asymptotic or large-sample test. This test first computes the Skewness and Kurtosis measures of the Ordinary
Least Square (OLS) residuals and the test-statistic used in the JB test is as follows:

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

(4.1)

where, $n = \text{Sample size}$, $S = \text{Skewness coefficient}$, and $K = \text{Kurtosis coefficient}$. Normally distributed variables have $S = 0$ and $K = 3$.

### 4.1.2.2 Unit root test

Unit Root test is basically a test for stationarity or non-stationarity of a time series data. Whenever two or more time-series data are used in regression analysis, one can frequently find a high $R^2$ even without any expressive relationship among the selected variables. In a few time-series variables, regression of one on other variable(s) shows a significant relationship between them though one expects no relationship. This occurrence demonstrates the problem of spurious regression (Gujarati, 2004). Thus, the regression of one non-stationary time series variable on another may produce a spurious regression. Therefore, Unit root test is conducted on all data-series variables considered for the study to ensure stationarity property of the time series.

The data series is said to be ‘stationary’ if its mean and variance are constant over time and the value of covariance between the two time periods depends only on the distance or lag between the two time periods and not the actual time at which the covariance is computed. In other words, a time series is said to be stationary if its mean, variance and covariance are time invariant. If the mean, variance and covariance of the data series are varying with time, then the data series is said to be non-stationary.
Though there are many unit root tests available, the presence of any unit root in the data series is checked by employing the Augmented Dickey-Fuller (ADF) test. Time series data should be stationary for analysing i.e. statistical properties of the series should be constant for putting the data to use. SadanandaPrusty (2009) stated, if a series in actual form is stationary, it is called I(0), denoting “integrated of Zero”. If a series needs differencing ‘d’ times to become stationarity, it is called integrated of order ‘d’, denoted by I(d).

Augmented Dickey-Fuller (ADF) test has been developed by the statisticians David Dickey and Wayne Fuller. In Dickey Fuller (DF) test, it was assumed that the error terms \((U_t)\) were uncorrelated. However, in some of the time series analysis it is common that error terms \((U_t)\) are correlated. If the error terms \((U_t)\) are correlated, the ADF test is the suitable test to carry out. Augmented Dickey-Fuller (ADF) test controls for higher order correlation by adding lagged difference terms of the dependent variable to the right-hand side of the regression.

The Augment Dickey Fuller specification used here is as follows:

\[
\Delta Y_t = b_0 + \beta Y_{t-1} + \mu_1 \Delta Y_{t-1} + \mu_2 \Delta Y_{t-2} + \ldots + \mu_p \Delta Y_{t-p} + u_t
\]  \hspace{1cm} (4.2)

where, \(Y_t\) represents time series to be tested, \(b_0\) is the intercept term, \(\beta\) is the coefficient of interest in the unit root test, \(\mu_p\) is the parameter of the augmented lagged first difference of \(Y_t\) to represent the \(p^{th}\) order auto regressive process and \(u_t\) is the white noise error term.

### 4.1.2.3 Linear correlation

The primary purpose of correlation analysis is to measure the strength or degree of association between two variables. The correlation coefficient \((r)\) measures the strength of linear association.
where, X and Y are the random variables. The researcher interpreted the correlation coefficients by referring ‘how to interpret a correlation coefficient r’ (Rumsey 2011)

Methodology adopted in Correlation analysis:

The linear correlation test has been applied for the variables viz., daily net FII investments, Nifty and CNX 500 for the entire study period. To identify the degree of comovement between FIIs’ investment and Nifty during various phases of the stock market, the study period has been divided in to six phases as per the market trend and global events. The Figure 4.1 shows the classification of nifty movements into various phases.
Phase 1 (April 2001 to May 2003): Market found volatile but remained within range. No consistent upward or downward movement.

Phase 2 (June 2003 to March 2008): This is one of the long-run bull phases of the stock market. During this period, strong economic growth and huge foreign investments observed.

Phase 3 (April 2008 to April 2009): This is a bear phase in which stock witnessed a sharp fall owing to global financial crisis and bankruptcy of many major financial institutions like Lehman brothers.

Phase 4 (May 2009 to November 2010): During this phase a market revived from the global financial crisis and moves upward towards the fair valuation. This is another bull-run in the stock market.

Phase 5 (December 2010 to June 2012): Doubts over US recovery, persistent distress in debt burdened euro zone and falling rupee, collectively turns the market to the bear run.

Phase 6 (July 2012 to March 2014): Upward trend in the market noticed as the growth seen in economies across the globe. Yet another bull run experienced.

4.1.2.4 Granger causality test

A time series $X_t$ Granger-Causes another time series $Y_t$ if series $Y_t$ can be predicted with better accuracy by using past values of $X_t$ rather than by not doing so, other information is being identical. The pre-condition for applying Granger causality test is to ascertain the stationarity of the variables. In this study, the Granger Causality test is undertaken to assess whether there is any potential predictability power of FII investment flows for stock market indices or stock market indices for FII investment flows. The researcher uses
pairwise Granger causality test to capture the degree and direction of causality between FII investment flows and stock market indices.

The optimum lag length for variables is found by vector auto regression (VAR) lag order selection method. There are six criteria namely log likelihood value (LL), sequential modified likelihood ratio (LR) test statistic, final prediction error (F & E), Akaike information criteria (AIC), Schwarz information criterion (SC) and Hannan-Quin information criterion (HQ) for choosing the optimal lag length. Apart from LR statistic, all are minimizing functions of lag length and the choice of optimum lag length is the minimum of the individual criterion function and is symbolized as *. The optimum lag length for this model is selected based on AIC. The Granger causality test is performed as follows and the following equations are framed for causality testing.

\[
\Delta FII_t = \alpha_1 + \beta_{11}\Delta Nifty_{t-1} + \beta_{12}\Delta Nifty_{t-2} + \beta_{1n}\Delta Nifty_{t-n} + \gamma_{11}\Delta FII_{t-1} \\
+ \gamma_{12}\Delta FII_{t-2} + \ldots + \gamma_{1n}\Delta FII_{t-n} + u_{1t}
\]

(4.4)

\[
\Delta Nifty_t = \alpha_2 + \beta_{21}\Delta FII_{t-1} + \beta_{22}\Delta FII_{t-2} + \beta_{2n}\Delta FII_{t-n} + \gamma_{21}\Delta Nifty_{t-1} + \\
+ \gamma_{22}\Delta Nifty_{t-2} + \ldots + \gamma_{2n}\Delta Nifty_{t-n} + u_{2t}
\]

(4.5)

where, \(\Delta FII_t\) is the first difference at time ‘t’ of FII, \(\Delta Nifty_t\) is the first difference at time ‘t’ of Nifty, ‘\(\alpha\)’ is the constant, ‘\(n\)’ is a positive integer, \(\beta_j\) and \(\gamma_j\) are parameters and \(u_t\) is an error term.

4.1.2.5 Variance decomposition

The Vector Auto-Regression (VAR) is mostly used for the purpose of forecasting interconnected time series variables and for examining the dynamic impact of random disturbances on the system of variables. Variance decomposition provides a system for testing 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 i.e. a shock to the \( i^{th} \) variable will directly affect that variable of course, but it will also be transmitted to all other variables in the system (Chris Brooks, 2008). It also decides on how much time ahead the forecast error variance of a particular variable is explained by shocks to each exogenous variable. In reality, it is noted that own shocks describe most of the forecast error variance. In this study, bivariate vector auto-regression (BVAR) model has been applied on the stationary time series of FII and Nifty.

### 4.1.2.6 Impulse response function

The significance of impulse response function is that it gives more insight in finding out the relationship and impact between two or more time series variables. i.e. it adds value to the results given by F-statistics and Variance decomposition. Impulse responses capture the responsiveness of the endogenous variables in the VAR system when a shock (innovation) is given to the error term of own variable and other variables. In impulse responses, a unit shock is given to each variable and captures its effect on the VAR system. The following VAR system models have been framed to capture the responsiveness of FII and Nifty when the shock is given to the error terms.

\[
\text{FII} = b_1 + \beta_2 \times \text{Nifty}_{t-i} + \beta_3 \times \text{FII}_{t-i} + U_1 \tag{4.6}
\]

\[
\text{Nifty} = b_4 + \beta_5 \times \text{FII}_{t-i} + \beta_6 \times \text{Nifty}_{t-i} + U_2 \tag{4.7}
\]

where, \( b_1 \) and \( b_4 \) are the intercepts, \( \beta_2, \beta_3, \beta_5 \) and \( \beta_6 \) are the beta coefficients; ‘\( t-i \)’ is a lag period (\( i = 1 \) to \( n \)); \( U_1 \) and \( U_2 \) are error terms.

The idea behind this model is any change (shock) in \( U_1 \) will bring change to FII. This will alter Nifty and also FII during the next period.
4.2 DETERMINANTS OF FII INVESTMENTS

The researcher after a careful thought and going through previous research identified various key macroeconomic and stock market variables as determinants of FII investment flows. The monthly data for those variables are collected from various sources. The data descriptions for the variables used in this framework are presented in the Table 4.2.

Table 4.2 Data description for FII determinants Study

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Unit of Measurement</th>
<th>Source cited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net FII equity Investment (FII)</td>
<td>Monthly</td>
<td>Indian Rupees (Rs.) Crore</td>
<td><a href="http://www.sebi.gov.in">www.sebi.gov.in</a></td>
</tr>
<tr>
<td>Growth rate of Index for Industrial Production of India (INDIIP)*</td>
<td>Monthly</td>
<td>Percentage (Index series: 1993-94 &amp; 2004-2005)</td>
<td><a href="http://www.rbi.org.in">www.rbi.org.in</a></td>
</tr>
<tr>
<td>Returns in Nifty index (RN)#</td>
<td>Monthly</td>
<td>Percentage (from Index, Base 1995 = 1000)</td>
<td>Calculated by author based on Nifty index data in <a href="http://www.nseindia.com">www.nseindia.com</a></td>
</tr>
<tr>
<td>Volatility in Nifty index (VN)</td>
<td>Monthly</td>
<td>Percentage (from Index, Base 1995 = 1000)</td>
<td>Calculated by author based on Nifty index data in <a href="http://www.nseindia.com">www.nseindia.com</a></td>
</tr>
</tbody>
</table>
Table 4.2 (Continued)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Unit of Measurement</th>
<th>Source cited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth rate of Index for Industrial production of USA (USIIP)</td>
<td>Monthly</td>
<td>Percentage (Index series: 2007 = 100)</td>
<td><a href="http://www.research.stlouisfed.org">www.research.stlouisfed.org</a></td>
</tr>
</tbody>
</table>

(*): 1993-94 series and 2004-05 series of INDIIP and INDWPI are connected using the linking factor provided by the CSO (Central Statistical Office). The linking factor for IIP general index and WPI all commodities are 2.111 and 1.8726 respectively.

(##): Monthly data for Nifty and S&P 500 is calculated by averaging daily closing value of Nifty and S&P 500 for that month.

Apart from the above major frameworks of the study, the researcher used normality test to check the data distribution and correlation analysis to know the degree of association between the variables used in the study. These are done in the earlier stage of the study to get the overall familiarity about the variables used in relation to FII investment flows.
Even though there are several variables which may have their influence on FII flows, the researcher has confined himself only to the selected macroeconomic and stock-market variables based on the extensive reviews gone through. The independent variables used in the study are categorized into Host country variables and Home country variables.

4.2.1 Determining Variables

4.2.1.1 Host country variables

a. Macro-economic variables

i. Index for Industrial Production (IIP) as a proxy for economic growth

ii. Wholesale Price Index (WPI) representing inflation level

iii. Exchange rate of Indian Rupee against USD (ER$)

b. Stock-market variables

i. Returns on Nifty Index (RN)

\[ RN = \ln \left( \frac{P_t}{P_{t-1}} \right) \]

where, \( P_t \) = Average closing price of Nifty at time-period ‘t’

\( P_{t-1} \) = One time-period lagged average closing price of Nifty

ii. Volatility in Nifty Index (VN)

\[ VN = \text{STDEV} \ (12 \ \text{period RN}) \]

where, STDEV is Standard Deviation and RN is Returns on Nifty index
4.2.1.2 Home country variables

a. Macro-economic variables

i. Index of Industrial Production (IIP) as a proxy for economic growth of USA

ii. Producers Price Index (PPI) representing inflation level of USA

b. Stock-market variables

i. Returns on S&P 500 Index (RSP)

\[
RSP = \ln \left( \frac{S&P\ 500_t}{S&P\ 500_{t-1}} \right)
\]

where, \( S&P\ 500_t \) = Average closing price of S&P 500 in time-period ‘t’

\( S&P\ 500_{t-1} \) = One time-period lagged average closing price of S&P 500

ii. Volatility in S&P 500 index (VSP)

\[
VSP = \text{STDEV} \left( 12 \text{ period RSP} \right)
\]

where, STDEV is Standard Deviation and RSP is Returns on S&P 500 index

4.2.2 Priori Expectations of Selected Exogenous Variables

\textbf{INDIIP} > 0: \text{ An increase in the industrial production certainly indicates the positive economic growth of a country. So, it is expected that FII flows increase with the higher Index for Industrial Production (IIP) in India.}

\textbf{INDWPI} < 0: \text{ When inflation in the domestic country increases, the purchasing power of the funds invested declines. Hence, investors will withdraw from the domestic market.}
ER$ < 0: It is expected to have a negative relationship between FII flows and Exchange rate. This is because higher the exchange rate reflects higher the depreciation of INR against USD. The depreciation (appreciation) of INR against USD lowers (improves) the value of FII investments in dollar terms.

RN > 0: An increase in Nifty Return will attract more foreign investments. So it is believed that FII flows and Returns on Nifty will have positive relationship.

VN < 0: The higher level of volatility in Indian stock market brings more risk to the investment made. This makes the risk averse investors who want to invest in India for diversifying their investment risk to rethink on their stance. Thus, volatility in the stock market is expected to reduce the FII inflows in India.

USIIP < 0: A vibrant growth in foreign economy deters the investment flow out of the country. Thus, greater the USIIP, lower may be the FII flows.

USPPI > 0: Higher inflation in home country erodes the return earned by the investment made there. Hence, it is assumed that higher inflation in USA brings more FII flows to India.

RSP < 0: It is expected that bullish trend in S&P 500 will have negative relationship with FII flows in India. This is because: better return on stocks in USA deters investors from investing in Indian stocks.

VSP > 0: The frequent and higher level of volatility in the home country’s stock market creates more risk and uncertainty on the investment returns. This forces the foreign investors to diversify their investment. It is presumed that higher volatility in USA stock market brings more FII investments to India.
4.2.3 Autoregressive Distribution Lag – Bounds test for Cointegration

To discover macroeconomic and stock market determinants of FII investments in India, ARDL model using bounds test approach is employed. The ARDL model was initially developed by Pesaran & Shin (1999) and it was advanced by Pesaran et al (2001). ARDL model is built on Unrestricted Error Correction Model (UECM) and enjoys numerous benefits over the other cointegration techniques. First, it simultaneously estimates the short-run and long-run components of model after removing autocorrelation problems. Second, it provides unbiased estimates of the long-run model and valid t-statistic even when some of the regressors are endogenous (Harris & Sollis 2003). Third, ARDL can be applied to small sample size study (Pesaran et al 2001). Fourth, it is not necessary for the variables to be integrated in the same order.

Even though, all the variables are not essentially integrated at the same order (I₀ or I₁), the variables integrated of order I₂ or more will ruin the ARDL estimation as the computed Wald or F-statistic under bounds test approach are based on the assumption that variables are either I₀ or I₁. Thus, employing unit root test is still necessary in ARDL bounds test approach to ensure none of the variables are I₂ or integrated at higher order (Shahbaz 2009).

The ARDL estimation equation (8) has ten determinants of FII investments including its own lags⁵. Both short-run coefficients and long-run coefficients of all the determinants are taken as exogenous variables for

---

⁵ Number of lags is determined based on VAR lag order selection criteria.
framing the equation. The ARDL estimation model used in the present study is as follows;

\[
\Delta FII_t = \beta_0 + \sum_{i=0}^{n} \delta_1 \Delta FII_{t-i} + \sum_{i=0}^{n} \delta_2 \Delta INDIIP_{t-i} + \sum_{i=0}^{n} \delta_3 \Delta INDWPI_{t-i} + \sum_{i=0}^{n} \delta_4 \Delta ER_{t-i} + \sum_{i=0}^{n} \delta_5 \Delta RN_{t-i} + \sum_{i=0}^{n} \delta_6 \Delta VN_{t-i} + \sum_{i=0}^{n} \delta_7 \Delta USIIP_{t-i} + \sum_{i=0}^{n} \delta_8 \Delta USPPI_{t-i} + \sum_{i=0}^{n} \delta_9 \Delta RSP_{t-i} + \beta_1 FII_{t-i} + \beta_2 INDIIP_{t-i} + \beta_3 INDWPI_{t-i} + \beta_4 ER_{t-i} + \beta_5 RN_{t-i} + \beta_6 VN_{t-i} + \beta_7 USIIP_{t-i} + \beta_8 USPPI_{t-i} + \beta_9 RSP_{t-i} + \beta_{10} VSP_{t-i} + U_t \quad (4.8)
\]

where, \(\beta_0\) is an intercept; \(\Delta\) is the first order difference; ‘t’ is the time dimension; ‘U_t’ is a white noise error term; \(\delta_1,…,\delta_{10}\) are the short-run coefficients; \(\beta_1, \ldots, \beta_{10}\) are the long-run coefficients.

The first and foremost step in the ARDL bounds test approach is to estimate the above Equation (4.8) by ordinary least squares to test for presence of long run relationship among the variables by carrying out F-test for the joint significance of the coefficients of the variables as stated below;

Null Hypothesis (\(H_0\)): \((\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8, \delta_9, \delta_{10}) = 0\)

Alternative Hypothesis (\(H_1\)): \((\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8, \delta_9, \delta_{10}) \neq 0\)

Critical value bounds for F-statistic with two sets are created by Pesaran et al (2001). If the calculated F-statistic is below the lower bound critical value, the null hypothesis of no cointegration cannot be rejected. Whereas, if the calculated F-statistic is above the upper bound critical value; the null hypothesis is rejected, meaning that there is a long-run cointegration relationship between the variables in the ARDL model. In case, if the calculated F-statistic falls between the upper bound and lower bound, the result of existence of any cointegration relationship is unsure.
After estimating Equation (4.8) to test long-run relationship, the next step is to estimate the conditional ARDL long-run model for FII as mentioned below:

\[
FII_t = \beta_0 + \sum_{i=0}^{n} \delta_1 \Delta FII_{t-i} + \sum_{i=0}^{n} \delta_2 \Delta \text{INDIIP}_{t-i} + \sum_{i=0}^{n} \delta_3 \Delta \text{INDWPI}_{t-i} + \sum_{i=0}^{n} \delta_4 \Delta \text{ER$}_t + \sum_{i=0}^{n} \delta_5 \Delta \text{RN}_{t-i} + \sum_{i=0}^{n} \delta_6 \Delta \text{VN}_{t-i} + \sum_{i=0}^{n} \delta_7 \Delta \text{USIIP}_{t-i} + \sum_{i=0}^{n} \delta_8 \Delta \text{USPPI}_{t-i} + \sum_{i=0}^{n} \delta_9 \Delta \text{RSP}_{t-i} + \sum_{i=0}^{n} \delta_{10} \Delta \text{VSP}_{t-i} + U_t \quad (4.9)
\]

In the last step, by estimating an Error Correction Term (ECT) connected with the long-run estimates, we attain the short-run dynamic factors. The short-run dynamics of ARDL model is stated as follows;

\[
\Delta FII_t = \beta_0 + \sum_{i=0}^{n} \delta_1 \Delta FII_{t-i} + \sum_{i=0}^{n} \delta_2 \Delta \text{INDIIP}_{t-i} + \sum_{i=0}^{n} \delta_3 \Delta \text{INDWPI}_{t-i} + \sum_{i=0}^{n} \delta_4 \Delta \text{ER$}_t + \sum_{i=0}^{n} \delta_5 \Delta \text{RN}_{t-i} + \sum_{i=0}^{n} \delta_6 \Delta \text{VN}_{t-i} + \sum_{i=0}^{n} \delta_7 \Delta \text{USIIP}_{t-i} + \sum_{i=0}^{n} \delta_8 \Delta \text{USPPI}_{t-i} + \sum_{i=0}^{n} \delta_9 \Delta \text{RSP}_{t-i} + \sum_{i=0}^{n} \delta_{10} \Delta \text{VSP}_{t-i} + \varphi \text{ECT}_{t-i} + U_t \quad (4.10)
\]

where, \( \delta_1, \delta_2, \ldots, \delta_9 \) are the short-run coefficients of the model’s convergence to equilibrium, \( \varphi \) is the speed of adjustment factor and ECT is the error correction term.

### 4.3 FII INVESTMENTS AND INDIAN STOCK MARKET VOLATILITY

This section of research aims to explore whether FII investment creates Stock Market Volatility in India. The daily data on FII investment and NSE’s Nifty log return are used for this part of study. Table 4.3 explains the data description for the study.
In order to study the FII activities in Indian stock market, daily net FII investments are considered and the study covers a long span of 13 years from April 2001 to March 2014. After carrying out the descriptive statistics, the unit root test using Augmented Dickey Fuller (ADF) test to verify the stationarity of data is conducted.

### 4.3.1 ARIMA Structure

ARIMA framework is used to find out the best fit model for daily Nifty log returns. ARIMA or Box-Jenkins models are generalizations of the simple AR model. It brings together three tools for modelling namely, Autoregressive (AR) term, Integration order term and Moving Average (MA) term. The identification phase of ARIMA model helps in deciding the kind of ARIMA model to be used. If the autocorrelation function dies off efficiently at an orderly level and the partial autocorrelations were zero after one interval (lag), then a first-order autoregressive model is suitable.

### 4.3.2 Model Selection

The characteristics of the error terms (residuals) in the dependent variable need to be verified for deciding the suitable model to measure the

---

**Table 4.3 Data description for FII and Indian Stock Market Volatility**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Unit of Measurement</th>
<th>Source cited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net FII equity investment (FII)</td>
<td>Daily</td>
<td>Rs. (Indian Rupee) in crore</td>
<td><a href="http://www.sebi.gov.in">www.sebi.gov.in</a></td>
</tr>
<tr>
<td>NSE Nifty Log Returns</td>
<td>Daily</td>
<td>Index closing point (Base year:1995)</td>
<td><a href="http://www.nseindia.com">www.nseindia.com</a></td>
</tr>
</tbody>
</table>
impact of FII investment on volatility in Nifty. If the time series dependent variable exhibits the characteristics such as volatility clustering (heteroscedasticity) or leverage effect or Leptokurtosis, then it leads to violations of homoscedasticity as well as autocorrelation assumptions of Ordinary Least Squares (OLS). The consequence of heteroscedasticity is that though the estimates of parameters are unbiased, it leads to large Standard Errors (SE) and the precision is affected. The application of linear models does not hold good in such circumstances as the linear models are unable to explain the special characteristics.

The simplest GARCH (1,1) specification for mean equation with exogenous variable(s) and error term is:

\[ Y_t = bX_t + U_t \] (4.11)

where, ‘b’ is the coefficient; ‘X_t’ is the independent variable and ‘U_t’ is the residual or error term.

The variance equation of GARCH (1, 1) specification is:

\[ \sigma^2_t = \omega + \alpha \mu^2_{t-1} + \beta \sigma^2_{t-1} \] (4.12)

where, \( \sigma^2_t \) is the conditional variance of residual at time ‘t’; \( \omega \) is the constant; \( \alpha \) and \( \beta \) are the parameter; \( \mu^2_{t-1} \) is the news about volatility from previous period which is measured as the lag of squared residual derived from the mean equation and it is the ARCH term. \( \sigma^2_{t-1} \) is the previous period’s conditional residual variance and it is known as GARCH term. In the (1,1) terms of the GARCH model, the first ‘1’in the parentheses denotes the presence of first- order autoregressive GARCH term and the second ‘1’ in the parentheses refers the first-order moving average ARCH term.
4.3.3 Diagnostic Test

Finally, the developed model needs to be verified for possible presence of autocorrelation and ARCH effect to ensure the model fit. The good model should be free from these issues. The correlogram of squared residuals demonstrates the Autocorrelation (AC) and Partial autocorrelation (PAC) functions of residuals along with the Ljung-Box Q-statistics for higher order serial correlation. In the absence of serial correlation in residuals, the AC and PAC at all lags should be closer to zero, and all Q-statistics should be insignificant with large p-values. ARCH test is a Lagrange multiplier (LM) test for ARCH in the residuals (Engle 1982). The observations in various financial time series have inspired the inclusion of heteroscedasticity specification. ARCH by its own does not invalidate standard Least Square (LS) inference but ignoring ARCH effects may result in loss of efficiency.

4.4 FORECASTING FII INVESTMENTS IN INDIAN STOCK MARKET

“Prediction is a very difficult art, especially when it involves the future”

- Neils Bohr (Nobel Laureate Physicist).

Foreign Institutional Investor (FII) investments are significant for any emerging economy for various reasons such as facilitating domestic investments without having adverse impact on Balance of Payment (BOP), allows for better pricing and liquidity of the domestic company shares, helps in stabilizing the Exchange rate etc. Moreover, FIIs are believed to have better market knowledge and informational advantage compared to retail investors. Due to these reasons, FIIs are likely to predict the market movements in a better way and thereby earning higher returns for their investment. Thus,
knowing the prospective FII investment in future may render a huge help for policy makers, investors and other stakeholders.

Forecasting is an essential use of time series analysis. Forecasting is the process of making declarations about any happenings or events whose real outcomes have not yet been detected. The art of forecasting into the future is essential and an important exercise to many stakeholders in various diverse fields. For example, Farmers may be interested to know the future rainfall possibility to cultivate the right crop at the right time. Similarly, investors would like to know the future performance of the assets to choose the correct asset for making investment. Thus, forecasting helps various stakeholders by providing precious information that can be used to take decision about the future.

It is a well-known fact that none of the techniques available can predict or forecast the event accurately and this is more factual when the forecasting is done on distant future. The reason for the difficulty in forecasting accurately is, any forecasting into future comes with the margin of error and this margin of error increases with longer duration of forecasting. It is always easy to forecast immediate future than forecasting the distant future. It is understandable that as the time passes, the factors and their likely influence may change and new factors may arise. In forecasting, suppose ‘t’ is the current time index and ‘t + h’ is the forecast period, then the time index ‘t’ is called the’ forecast origin’ and the ‘h’ is the forecast horizon.

This section of the chapter describes the concepts, techniques and terms which are used in forecasting. It comprises approaches in forecasting, forecasting types, modelling techniques and brief narrative of various forecasting models.
4.4.1 Approaches in Forecasting

According to Chris Brooks (2008), Forecasting can be carried on either in a cross-sectional or a time series context.

4.4.1.1 Econometric (Structural) forecasting

It narrates a dependent variable to one or more independent variables. These models generally work better in the long run, as a long run association between variables often arises from no-arbitrage or market efficiency conditions. Long run exchange rate forecasting based on purchasing power parity is an example of such forecasting.

4.4.1.2 Time series forecasting

Involves forecasting the future values of a data series given its previous values and/or previous values of error terms.

4.4.2 Types of Forecasting

Forecasts are made principally because they are useful in making decisions about a future event. In a broader context, there are two different types of forecasts namely, Qualitative forecast and Quantitative forecast.

4.4.2.1 Qualitative forecasting

The qualitative forecasting techniques are beneficial whenever past and reliable data are not available. The qualitative forecasting techniques are built on the expert’s information, views, market research etc. These techniques use the expertise’s views to create a model which forecasts future values. The qualitative techniques are also used where the historical data is accessible but due to some reasons the use of the data cannot be reliable.
4.4.2.2 Quantitative forecasting

The Quantitative forecasting techniques use the availability of historical data to develop forecast models. These techniques use the variables own past data or other data for forecasting the future values of the variables. The quantitative forecast techniques attempt to discover the relationships between dependent variable and independent variable(s) and further use these relationships to predict the values of the dependent variable.

4.4.3 In-sample and Out-sample Forecasts

Chris Brooks (2008) illustrates the difference between in-sample and out-sample forecasts. According to him, in-sample forecasts are generated for the same set of data that was used to estimate the model’s parameters. The general expectation is that the forecasts of a model to be relatively good in-sample. The best way to examine the forecast accuracy is not to use all observations in estimating the model parameters, but rather to hold some observations back. The latter sample, sometimes known as a holdout sample, would be used to construct out of sample forecasts.

4.4.4. Data and Forecasting Methodology

In this study monthly FII investments and Return on Nifty index (RN) data are used and the time period considered for this forecasting analysis is from July 2012 to March 2014\(^6\). This is because any prior data before this period are assumed to be too old and may not have any significant bearing on the period of forecast. Among the time period for forecast analysis, July 2012 to September 2013 (15 observations) monthly data of FII investments and

---

\(^6\)Chosen as it is the last phase (i.e. phase 6) of the time period considered for the study – as mentioned in figure 4.1 of this chapter.
Return on Nifty index are in-sample period data which are used for estimating the regression equation and based on this estimated output, the FII investments for the next six months which is known as out-of-sample period data i.e. from October 2013 to March 2014 are forecasted. Figure 4.2 depict the in-sample and out-sample periods.

![Figure 4.2 In-sample and Out-sample periods for forecast analysis](image)

The Autoregressive model framed for the forecasting net FII investments is shown in the equation (4.13).

\[
FII_t = \beta_0 + \beta_1 RN + \beta_2 FII(-1) + U_t
\]  

(4.13)

where, \( \beta_0 \) is the constant; \( \beta_1 \) and \( \beta_2 \) are the coefficients; RN is Nifty returns; FII(-1) is the first lag of FII investments and \( U_t \) is the error term.

The model is called as Autoregressive model because it includes lagged values of the dependent variable among its explanatory variables. If the chosen model is correct, the forecast errors will be fewer. The forecasting errors widens usually due to wrong method and level of incorrectness in information used to forecast the future. The Table 4.4 displays the data description of the variables used in this part of the study.
Table 4.4 Data description for Forecasting net FII investments

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>No. of Observations</th>
<th>Unit of Measurement</th>
<th>Source cited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net FII equity Investment (FII)</td>
<td>Monthly</td>
<td>15</td>
<td>Indian Rupees (Rs.) Crore</td>
<td><a href="http://www.sebi.gov.in">www.sebi.gov.in</a></td>
</tr>
<tr>
<td>Return of Nifty index (RN)</td>
<td>Monthly</td>
<td>6</td>
<td>Percentage (from Index, Base 1995 = 1000)</td>
<td><a href="http://www.nseindia.com">www.nseindia.com</a></td>
</tr>
</tbody>
</table>

The forecasted observations are then compared with the actual FII investments to check the accuracy of the forecast model. The forecasting procedure followed in this study is standard procedure which is illustrated in the Figure 4.3


**Figure 4.3 Forecasting Procedure**
The data series considered for forecasting needs to be examined for stationarity. The stationarity property of the data is essential for forecasting. Upon confirming this, the suitable model of forecasting may be chosen. The estimated model should include the necessary dependent and independent variables related to forecasting. There are numerous forecasting methods and techniques available in practice. Forecasting the financial time series is commonly assumed to be a very tough job. Popular time series forecasting models are as follows:

1. Box – Jenkins (ARIMA) approach.
2. Exponential smoothing method
4. ARCH family models (used in the presence of heteroscedasticity)

The selected estimate model is considered to be fit and good, only if the model has no statistical error. The term ‘no statistical error’ denotes that the regression model used for forecasting should have reasonable $R^2$, no serial correlation, no heteroscedasticity and residuals should be normally distributed. In this study, the regression forecast equation estimated has been put to test for all these diagnostic checking before the model is used for forecasting.

After fulfilling all diagnostic tests, the estimated model is ready for forecasting. There are two types of forecasting methods available namely, Static forecast and Dynamic forecast. Static forecast calculates a sequence of one-step ahead forecasts by using actual data instead of forecasted data based on lagged dependent variables. On the other hand, Dynamic forecast computes forecasts for time periods after the first period in the sample by
using the formerly forecasted values of the lagged left-hand variable. Dynamic forecasts are also known as n-step ahead forecasts. Dynamic forecasts are used in this research study.

The forecasting evaluation is required to ensure the reliability of the selected forecast model. One good way to do this is by laying both actual and forecasted values in a graph and observes how closely these two are moving together. If the forecast line captures all the upward and downward movements of the data and is closely moving along with the actual line, then it is considered that the estimated model is good. The ideal forecast model should have minimum forecast error. The forecast error is the gap between the actual data and forecasted data. Statistically, this gap is known as Root Mean Squared Error (RMSE).

Thus, this chapter describes the data used and framework designed for all the stated objectives of this research in detail. These research frameworks outline the procedure followed in analysing the time series data and are the base upon which all the analyses are carried on.