CHAPTER 4

DATA AND RESEARCH METHODOLOGY

The present chapter deals with the description of the data and the research methodology incorporated to cover the research objectives. This study comprises of primary and secondary data sets. Section 4.1 and its subsections below represent secondary data set and related methodology followed by the description of the primary data set.

4.1 SECONDARY DATA SAMPLE

The study uses a sample of different market indicators of Nifty 50 stocks. These indicators include

- Daily total returns of the index and securities,
- Daily transaction volume of the index and securities,
- Market capitalization of securities,
- Daily high and low values of the index
- Daily closing prices of Nifty 50 index options

Along with this the daily risk-free rate of return of the T-bill index is also used. The data set has been taken for a period 2006-2013. This period is particularly significant in recent times as India has seen the market swings from boom to recession to recovery within the span of these seven years, which makes this an interesting period for studying the investor psychology and behavior. The data types used for empirical analysis in the study are cross sectional and time-series data.

4.1.1 COLLECTION OF SECONDARY DATA

The data on Nifty 50 stocks is obtained from the Centre for Monitoring Indian Economy (CMIE) Prowess database. Nifty 50 is a leading index for large companies on NSE. It is a well diversified index of 50 stocks that account for 22 sectors of Indian economy. It is the 16th largest stock exchange in the world by market capitalization. The trading volume on NSE has also grown tremendously over the years. In India, it is the largest index by the number of trades and daily
turnover for both equities and derivatives trading, making it one of the most liquid stock exchanges in the world. The dataset of Nifty50 options and Treasury-bill index is obtained from the official NSE website.\(^1\)

### 4.1.2 STATISTICAL TECHNIQUES USED FOR SECONDARY DATA ANALYSIS

The presence and impact of each bias are investigated with the help of specific techniques as suggested in the literature. Some of the techniques are common, but others are specific to each bias. These techniques are listed in table (4.1.2).

**Table 4.1.2:** Techniques used to study the aforementioned biases

<table>
<thead>
<tr>
<th>Technique</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit Root Test</td>
<td>All four biases</td>
</tr>
<tr>
<td>Linear regression</td>
<td>Herding</td>
</tr>
<tr>
<td>Time series regression</td>
<td>Optimism (pessimism)</td>
</tr>
<tr>
<td>Pricing kernel technique</td>
<td>Optimism (pessimism)</td>
</tr>
<tr>
<td>GJR GARCH</td>
<td>Optimism (pessimism)</td>
</tr>
<tr>
<td>Vector Autoregression</td>
<td>Overconfidence and disposition effect</td>
</tr>
<tr>
<td>Impulse Response Function</td>
<td>Overconfidence and disposition effect</td>
</tr>
</tbody>
</table>

The brief description of each technique is mentioned below.

i. **Unit root test:** It is applied to check the stationarity of time series data. It is used as a preliminary test before proceeding to any regression model. A series is said to be stationary if its mean and variance does not vary systematically over time. Any series which does not follow this convention is said to be nonstationary [67]. Stationarity of the series is a necessary condition to make any meaningful inference from the data. Otherwise, the tests spurious results may be obtained which are of little practical value. To elaborate this test, consider a random walk model (RWM) as:

$$Y_t = \rho Y_{t-1} + \epsilon_t \quad -1 \leq \rho \leq 1 \quad (4.1.2.1)$$

Here if \(\rho = 1\), then this process becomes a RWM (without drift) which is a nonstationary stochastic process and this condition is called a unit root problem. This problem is detected with

---

\(^1\) Official NSE website: nseindia.co.in
the help of a variety of tests for e.g. Dickey-Fuller (DF) Test, Augmented Dickey-Fuller (ADF) test and Phillips-Perron (PP) unit root tests.

**ii. Linear regression analysis:** It is concerned with the investigation of linear dependence of one variable (dependent variable) on one or more variables (independent or explanatory variables). Here, the dependent variable is stochastic while the independent variable is non-stochastic, such that the average value of the dependent variable is estimated on the basis of a fixed value of independent variable [67]. A simple univariate linear regression model is of the form:

\[
Y_t = \alpha + \beta X_t + \varepsilon_t
\]  
\[(4.1.2.2)\]

The equation shown above represents a two variable linear regression model. In the analysis, multiple regression model has been used.

Where, \(Y_t\) = is the dependent variable at time ‘t’

\(X_t\) = is the independent variable at time ‘t’

\(\beta\) = measures the impact of \(X\) on \(Y\)

\(\alpha\) = constant

\(\varepsilon_t\) = error term, accounts for all the unexplanatory factors that impact \(Y\) but are uncorrelated with the dependent variable.

Here the estimation of \(\beta\) takes place with the help of ordinary least squares method (OLS) as it is intuitively appealing and mathematically simpler than other methods of estimation like Maximum Likelihood (ML). The data set fits into the OLS model and it satisfies all the assumptions of the classical linear regression model (CLRM) and R-squared measures the Goodness of fit of the model.

**iii. Time series regression:** Time series regression models use historical trends to predict a future response (also known as the autoregressive dynamics). The basic time series model is the autoregressive (AR) process and is represented as:

\[
Y_t = \beta Y_{t-1} + \varepsilon_t
\]  
\[(4.1.2.3)\]
This is a first order autoregressive process, AR (1), where the value of series Y at time t is depended on its first lag value \(Y_{t-1}\) and a random term, \(\varepsilon_t\). Further, if the value of Y at time t depends on its previous two time periods, it becomes an AR (2). Here the number of lags can be more than 1, so if the value of Y at time t depends on its p lags, it becomes an AR (p) process [67]. The lag length or the value of p is decided with the help of Akaike information criterion (AIC) or Schwarz criterion (SC).

Other time series processes are moving average (MA), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA). Time series regression helps in understanding and predicting the behavior of dynamic systems from experimental or observational data. It is generally used for modeling and forecasting of economic and financial series.

**iv. GJR GARCH technique:** The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models are used to capture the volatility clustering phenomenon in financial time series. Volatility clustering occurs when financial series (e.g. stock prices) show wide swings for an extended period followed by a period of tranquility. This phenomenon reveals the impact of news on financial series. The GARCH model has some variants like Glosten, Jagannathan, and Runkle GARCH (GJR GARCH) that capture the differential impact of the news. It is represented as:

\[
\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 - \beta \sigma_{t-1}^2
\]

Here, ‘I’ is the interaction dummy

I=1, when \(\varepsilon_t < 0\) and 0 otherwise. In the presence of good news, \(\varepsilon_{t-1}^2 > 0\) and the total impact is equal to \(\alpha_i\). In presence of bad news, \(\varepsilon_{t-1}^2 < 0\) and the impact becomes \(\alpha_i + \gamma_k\).

**v. Pricing kernel:** In dynamic exchange economy equilibrium, the price of a security \(P_t\) at date t with single liquidating payoff \(Y(C_t)\) at maturity date T is given as:

\[
P_t = E_t[Y(C_T) M_{t,T}]
\]
Where, $M_{t,T}$ is the marginal rate of substitution between dates $t$ and $T$, and $\delta^{T-t}$ is the rate of time preference. Here the current price of a security equates to its expected discounted future payoff and discounting is done by using stochastic discount factor or $M_{t,T}$. It is also known as the pricing kernel.

**vi. Vector autoregression (VAR) models:** VAR model is a part of time series forecasting and is developed by [148]. The VAR method considers several endogenous as well as exogenous variables together. However, each endogenous variable is explained by its past (lag) values along with the lag values of other endogenous variables in the model. It is estimated by the following equations. Here, no exogenous variable is used in VAR equations.

\[
\begin{align*}
Y &= \alpha \sum_{j=0}^{k} A_j X_{t-j} + \varepsilon_t \\
X &= \alpha \sum_{j=0}^{k} A_j Y_{t-j} + \varepsilon_t
\end{align*}
\]

Where, $Y$ and $X$ are endogenous variables and $\varepsilon$’s are the stochastic error terms called impulses or innovations or shocks in VAR terminology. The determination of lag lengths ($k$) is crucial in this estimation as too many lag lengths could introduce Multicollinearity and too few could cause specification errors. Here, ‘$k$’ is decided with the help of Akaike or Schwarz Information Criterion and the model with lowest values for these criteria is chosen [67]. The estimation of parameters ($\beta$ and $\gamma$) in this case is through the OLS method.

**vii. Impulse response function (IRF):** It is the response of one variable to an impulse change in another variable [114]. For an AR (1) equation:

\[
x_t = \alpha x_{t-1} + \varepsilon_t
\]

The IRF is the path the $x$ follows if it is kicked by a single unit shock $\varepsilon$, i.e. $\varepsilon_{t-1}=0$, $\varepsilon_{t}=1$, $\varepsilon_{t+j}=0$; where $\varepsilon$ is the residual and $j$ is the number of periods (Cochrane, 2005). This function is
important because it provides an additional insight into the cause and effect behavior in time series models like VAR.

4.1.3 BIAS-WISE DESCRIPTION OF THE DATA SET AND METHODOLOGY

This section deals with bias-wise description of secondary data and research methodology. The methodology for biases herding and optimism is discussed first as they have been studied independently. It is followed by the methodology for overconfidence and the disposition effect, that have been studied independently as well as in conjunction with each other.

i. TO DETERMINE THE PRESENCE AND IMPACT OF HERDING IN THE INDIAN EQUITY MARKET

Data description:
The data set consists of daily returns of each constituent stock of Nifty50 index as well as the daily total returns of the index itself.

Methodology

A. Presence of herding on the market as a whole

The study follows the methodology given by [36]. A regression model is run to find out the effect of market stress on individual return dispersion.

\[
CSSD = \beta_0 + \beta_1 D_U + \epsilon
\]  

(1)

CSSD has been used as a measure of individual return dispersion. The dummy variables in regression equation (1) are used as explanatory variables to differentiate the periods of market stress from normal periods, taking into consideration that market stress occurs when aggregate returns lie in the upper or lower tail of return distribution.

\[D_L = 1 \text{ if, on day } t, \text{ the market return } (R_{m, i}) \text{ lies in the lower tail of return distribution and 0 otherwise.}\]

\[D_U = 1 \text{ if, on day } t, \text{ the market return } (R_{m, i}) \text{ lies in the upper tail of return distribution and 0 otherwise.}\]
The upper and lower tails are determined at 66 percent \((R_m \pm \sigma)\), 95 percent \((R_m \pm 2\sigma)\), 99 percent \((R_m \pm 3\sigma)\).

\[
\sqrt{\frac{\sum_{i=1}^{N} R_{i,t}^2}{N}}
\]

(2)

Where:

\(R_{i,t}\) = return of stock ‘i’ at time \(t\)

\(R_{m,t}\) = cross sectional average return of \(N\) stocks of the sample at time \(t\)

It was argued that in cases of extreme market stress, investors go with the consensus rather than following their own beliefs so as to seek certainty and conformity. They want to avoid the anxiety of making incorrect decisions under the conditions of uncertainty which comes with market stress and leads to herding. In the presence of herding, the investors’ decisions would be based solely on market movements, so that the individual asset returns would be similar to overall market returns. Hence the value of CSSD increases at a decreasing rate. In the presence of severe herding it may lead to decrease in dispersion.

B. **Nonlinearity of herding pattern**

Nonlinearity between dispersion and market return was checked using curve estimate (Fig.1).

![Fig. 4.1.3: Curve estimate: Relation between CSAD and market returns](image-url)
[33] suggest that in the presence of moderate to severe herding, the return dispersions will decrease (or increase) at decreasing rates. According to them, this relationship should be negative and nonlinear in the situation of herding. Therefore, a second test is conducted to investigate the presence of the nonlinear relationship between dispersion and market returns. [33] give cross sectional absolute deviation (CSAD) as the measure of dispersion.

They consider a general quadratic equation to test this behavior:

$$CSAD_t = \alpha + \gamma_1 R_{m,t} + \gamma_2 R_{m,t}^2$$  \hspace{1cm} (4)

Where the presence of a negative and significant $\gamma_2$ indicates herd behavior. The stationarity of CSAD series is checked by unit root tests.

**C. Presence of herding in bull and bear phase of the market respectively**

Considering that the stock behavior may be asymmetric in up and down market phases, the generalized relationship mentioned above can be bifurcated into following;

$$CSAD_{t}^{UP} = \alpha + \gamma_1^{UP} |R_{m,t}^{UP}| + \gamma_2^{UP} (R_{m,t}^{UP})^2 + \epsilon_t$$  \hspace{1cm} (5)

$$CSAD_{t}^{DOWN} = \alpha + \gamma_1^{DOWN} |R_{m,t}^{DOWN}| + \gamma_2^{DOWN} (R_{m,t}^{DOWN})^2 + \epsilon_t$$  \hspace{1cm} (6)

Where:

$|R_{m,t}^{UP}|$ and $|R_{m,t}^{DOWN}|$= the absolute values of the average overall sample return when the market is up (or down).

Similar to the previous case, here also negative and significant $\gamma_2^{UP}$ and $\gamma_2^{DOWN}$ captures herding behavior.

**ii. TO INVESTIGATE THE PRESENCE AND IMPACT OF EXCESSIVE OPTIMISM (PESSIMISM) IN THE INDIAN EQUITY MARKET**

**Data description**
The sample consists of daily total returns of the index, closing prices of Nifty 50 index options that are of four types: in-the-money put options; out-of-the-money put options; in-the-money call options and out-of-the-money call options, and daily risk-free rate of return of the T-bill index\(^2\).

**Methodology**

This section explains the methodology to investigate the presence and impact of optimism (pessimism) bias in the Indian equity market. It has been divided into two subsections. The first section deals with capturing the presence of optimism (pessimism) bias. In the second section, the relationship of this bias with stock market indicators is considered using time series regression.

**A. Capturing optimism (pessimism)**

The major concern in this investigation is to find an appropriate measure for optimism (pessimism). Going by the definition, excessive optimism is the tendency of representative investor to overestimate mean returns and pessimism is the tendency to underestimate the same \([70], [12]\). Therefore the assessment of optimism (pessimism) requires two types of estimates of expected returns: first is the objective or rational investor’s expected returns that is not prone to optimism (pessimism) and second is the representative investor’s expected returns that is prone to optimism (pessimism). Here the rational investor conforms to standard finance theories and tries to maximize expected utility. Such an investor would expect a higher return for taking higher risks. In contrast, the representative investor in the study mimics the behavior of a real life individual investor whose decision making is biased depending upon her sentiment.

According to standard asset pricing theories, equilibrium prices are set by investors holding correct beliefs. Keeping this assumption in mind, the first step is to calculate the objective probability density function (PDF) which depicts the correct beliefs of a rational investor. The second part takes into account a realistic investor who sets the equilibrium prices and is prone to bias. This involves the calculation of a probability density function (PDF) of a representative investor which should reflect the behavioral biases in the investor population. For this purpose,

\(^2\) At-the-money options are not reported as the trading volume of these options are extremely low as compared to the other two options to provide any significant results.
the objective PDF is converted into a representative PDF by incorporating a sentiment measure. This measure is calculated using the pricing kernel approach as suggested by [12]. The study uses both an empirical pricing kernel and a theoretical pricing kernel approach. The difference between the empirical pricing kernel and the theoretical pricing kernel gives the sentiment measure.

The detailed methodology of each step is discussed below:

**Step 1: Empirical pricing kernel**

It is a stochastic discount factor (SDF) which is the inter-temporal marginal rate of substitution defined as the discounted ratio of marginal utilities in two successive periods [28]. It is used in the pricing of assets in a risk neutral environment, under the dynamic equilibrium model. According to financial theories, the price of any asset is equal to the expected present value of future payoffs where discounting is done using the risk-free rate and then expectation are taken with respect to the marginal rate of substitution weighted probability density function of the payoffs [11]. It is denoted by $M_{t,T}$, where $(T-t)$ is fixed and equal to three months throughout this paper. These three months are the expiration period of Nifty fifty index options which are European in nature.

$$M_{t,T} = \frac{S_T}{S_t}$$

(7)

Where:

$r_f$: daily risk free rate of return taken as daily T-bill index value

$q$: risk neutral density

$p$: objective density

$S$: Nifty 50 index

**Calculation of the objective density ($p$) and the risk neutral density ($q$):** Call and put option prices of the Nifty50 index are used to estimate risk neutral density ($q$) and daily returns of the Nifty 50 index are used for calculating the objective price density ($p$). The GJR GARCH
model is used to capture the index dynamics of the Nifty 50.\textsuperscript{3} The effectiveness of this model has been verified with the help of the Ljung Box test. GJR GARCH is represented as:

$$\varepsilon_t = \sigma_t z_t, \quad z_t$$ is the standardized historical innovation and \(I_{t-1}\) is a dummy variable which takes the value 1 when there is a bad news \((\varepsilon_{t-1}<0)\) and 0 otherwise. In the presence of bad news, the model accounts for the leverage effect if \(\gamma>0\). This suggests that bad news \((\varepsilon_{t-1}<0)\) increases the volatility more than good news \((\varepsilon_{t-1} \geq 0)\). Similarly the GJR GARCH model is applied to out of the money Nifty50 options.

The estimates of these two GJR-GARCH models are used for simulating 20000 trajectories of Nifty50 returns and option prices. Monte Carlo simulation is used for this purpose. The probability density functions \((p \text{ and } q)\) are subsequently obtained from these trajectories by using the kernel density function.\textsuperscript{4}

**Step 2: Theoretical pricing kernel**

Here the stochastic discount factor (SDF) incorporates a sentiment measure giving it a behavioral attribute. In constant relative risk aversion (CRRA) the SDF is calculated as given by [135].

$$M(\theta_0, \theta_1, S)$$

Where \(M\) is the pricing kernel

\(\theta_0\): discount factor measuring degree of impatience or the time discount factor

\(\theta_1\): degree of risk aversion

\(S\): proxy value of the market portfolio, taken as the daily returns of Nifty50 index.

---

\textsuperscript{3} GJR GARCH is an asymmetric GARCH model and it is preferred over other GARCH models as it captures the leverage effect. The asymmetry term is based on Glosten, Jagannathan, and Runkle (1993) (GJR).

\textsuperscript{4} GJR GARCH model, Monte Carlo Simulation and Kernel density function have been applied with the help of MATLAB.
Here $\theta_0$ and $\theta_1$ are calculated with the help of asset pricing equation given by [28]$^5$ given by

$$\text{rf: log gross risk free rate}$$

$$E[\log(S_t)] \quad \text{Objective time t conditional mean}$$

$$\nu[\log(S_t)] \quad \text{Objective time t conditional variance}$$

Equation (3) can also be represented as

$$\text{[135] captures the impact of sentiment ($\Lambda_t$) on log SDF such that}$$

$$\text{Step 3: Calculation of sentiment ($\Lambda_t$):}$$

The empirical pricing kernel provides the projected estimates for Nifty50 by incorporating historical values for the period 2006 to 2011. On the other hand, the CRRA pricing kernel provides the estimate of sentiment from the results of empirical analysis. The sentiment measure is then taken as the difference between the empirical pricing kernel and the CRRA pricing kernel.

$$\Lambda_t = \log(M_{t,T}) - \log(M_{t,T}(\theta)) \quad (13)$$

Where, $\Lambda_t$ provides an estimate of the sentiment function. Given the equation (12), $\Lambda_t$ is a scaled log-change measure which transforms the objective return pdf $p$ into a representative investors’ return pdf, $pr$.

$^5$ $^1$According to the equation given by Campbell et al. (1997) $\text{rf}$ is negatively related to the degree of impatience ($\theta_0$) as it takes a high real rate of interest to make people overcome their impatience and save rather than consume.

With respect to the degree of risk aversion ($\theta_1$) it is stated that a risk averse investor has a concave utility function. The higher the risk aversion, the more concave the utility function and the higher the desire of investor to maintain a smooth investment flow across time. Thus it takes a bigger change in $\text{rf}$ to induce a given change in her investment pattern. Therefore $\text{rf}$ is positively related to the degree of risk aversion.
Step 4: Conversion of objective pdf (p) into representative pdf (pr) using sentiment function ($\Lambda_t$)

Barone-Adesi et al. (2012) provide the relationship between $pr$, $p$ and $\Lambda_t$.

$$\frac{p_{pr}}{Q_t} \theta_0, t, p$$

(14)

Where, $\theta_{0,t,p}$ is the scaling factor of $\theta_0$, such that the representative investors’ pdf (pr) integrates to unity. Here $\log \left( \frac{pr}{p} \right)$ represents the error in probability density associated with a specific return. This means that if an investor underestimates the probability of x percent market return of 5 percent, then the log change measure at x percent return will be -5 percent.

Step 5: Calculation of optimism bias

As mentioned earlier in this section, optimism bias occurs when a representative investor overestimates mean returns. [12] give a measure to capture optimism

$$\frac{p_{pr}}{Q_t} \theta_0, t, p$$

(15)

Where,

$E_{i}^{pr} = \text{conditional expectation at date } t \text{ under the representative investors’ pdf (pr)}.$

$E_{i}^{p} = \text{conditional expectation at date } t \text{ under the objective pdf (p)}.$

$S_t = \text{Nifty 50 index returns at date } t.$

In case of optimism this measure will be positive. However, its negative sign means that the representative investor underestimates mean returns i.e. she is pessimistic.

B. Impact of optimism (pessimism) on risk premium and volatility

This section deals with time series properties of optimism (pessimism) and related market indicators (risk premium and volatility).

Risk premium: It is the difference between expected return and the risk free rate of return ($rf$). This risk free rate is incorporated into the risk neutral pdf (q). In the present study, there are two risk premiums: the objective risk premium and the representative investors’ risk premium.

Objective risk premium: $E_{i}^{p} - E_{i}^{pr}$

(16)
Representative investors’ risk premium:  

\[ \frac{\text{Representative investors’ risk premium}}{\text{Risk premium (objective or representative investor) at date } t} \]  

Time series regression is used to check the impact of optimism on risk premium.  

\[ t_j \text{ optimism} \text{ RP}_t \text{ j} \beta \text{ Optimism}_t \text{ } \lambda \text{ measures the impact of optimism or pessimism on risk premium.} \]

Where,  

\[ RP_t \text{ Risk premium (objective or representative investor) at date } t. \]

\[ j \text{ number of lags of risk premium} \]

\[ \beta_j \text{ measures the impact of ’} j \text{’ lags on the risk premium} \]

\[ \text{Optimism}_t \text{ optimism or pessimism in a day is calculated by equation (15) } \]

\[ \lambda \text{ measures the impact of optimism or pessimism on risk premium.} \]

Here the appropriate number of lags (j) has been identified with the help of Akaike information criteria (i.e. j =2).

\[ \text{Volatility: The impact of volatility on optimism is investigated using time series regression.} \]

\[ \text{Where,} \]

\[ \text{Optimism}_{t-1} = \text{one day lagged value of Optimism (Pessimism)} \]

\[ S_{t-1} = \text{previous day lagged value of the Nifty50 index return} \]

\[ \text{Lagvol}_t = \text{past volatility is measured using high and low values of the previous day} \]

\[ \beta' = \text{measures the impact of the most recent lag on Optimism (Pessimism)} \]

\[ \delta = \text{measures the impact of the previous day lag of Nifty50 index return on optimism} \]

\[ \gamma = \text{measures the impact of Lagvol on optimism (pessimism).} \]

**iii. TO INVESTIGATE THE PRESENCE AND IMPACT OF OVERCONFIDENCE AND THE DISPOSITION EFFECT IN THE INDIAN EQUITY MARKET**
Data description

The study uses a mobile sample of Nifty50 stocks. The sample consists of total returns & transaction volume for each constituent stock, and the index. This is collected for a period of 7 years, starting from 1st April, 2006 to 31st March, 2013. The returns and trading volume are taken on a daily basis. The market capitalization of each stock is also procured at the beginning of the period.

Methodology

To investigate the presence of overconfidence and the disposition effect in the Indian equity market, a vector autoregression (VAR) is applied and its validity is verified with the help of impulse response functions (IRF’s). Here the investor’s overconfidence is detected with the help of VAR on market-wide transaction volume and market returns. Further, VAR is also applied to security wide transaction volume, security returns and market return to investigate and segregate the impact of the disposition effect and overconfidence as suggested by [160].

A. Market-wide VAR to investigate investor’s overconfidence

Here the endogenous variables are log market turnover and daily market return of Nifty50 index, and the exogenous variable is the daily index volatility calculated using daily high and low value of market index.

\[
\begin{align*}
\text{LogT} & = \log \text{ value of trading volume of market index} \\
\text{Rm} & = \text{daily return of market index} \\
\text{Vol} & = \text{daily volatility of market calculated using daily high and low values of the index.} \\
k & = \text{number of lags (equal to 10) decided on the basis of Akaike Information Criteria}
\end{align*}
\]

Where:

LogT = log value of trading volume of market index
Rm = daily return of market index
Vol = daily volatility of market calculated using daily high and low values of the index.

k= number of lags (equal to 10) decided on the basis of Akaike Information Criteria
B. Security-wide VAR to investigate and segregate the impact of the disposition effect and overconfidence.

The existing literature suggests that the positive portfolio returns make the investors overconfident, which lead to an increase in trading volume of the overall market [160] and [150]. However, for an investor affected by the disposition effect, an increase in return of a particular stock will influence her to trade that particular stock and avoid trading in other stocks at that point of time. This creates an increase in transaction volume for an individual stock. This asymmetry in the mechanism makes it possible for the researchers to measure the effect of these biases and is captured with by applying security-wide VAR.

\[
\begin{align*}
\text{LogT} & = \log \text{value of number of shares traded for security i} \\
\text{Rm} & = \text{daily return of market index} \\
\text{Ri} & = \text{daily return security i} \\
\text{Ivol} & = \text{Idiosyncratic volatility of firm ‘i’ on day t, calculated using CAPM.} \\
\end{align*}
\]

Here \( k = 10 \) based on Akaike Information Criteria.

Here the positive value of \( \gamma_j \) captures the impact of the disposition effect and positive value of \( \lambda_j \) captures the impact of overconfidence.

C. Impulse Response Function (IRF)

IRF is applied to illustrate how the endogenous variables relate to each other over time. It traces the effect of one standard deviation shock in one residual to current and future values of the endogenous variables through the dynamic structure of VAR. For instance, in market wide VAR
(refer equation 20) the IRF captures the impact of changes in one residual say $\varepsilon_t$, on current and future values of $\log T$ and $R_m$. Similar case happens for security wide VAR, which has three endogenous variables, individual securities trading volume, security return and market return. In the present study, the IRF is applied for 7 periods.

The next section describes the data and explains the methodology of the primary survey, which is conducted to investigate the most prominent bias in Indian context.

**4.2 PRIMARY DATA: SURVEY DESIGN AND SAMPLE COMPOSITION**

To determine which bias is most pronounced in Indian context

Primary data has been collected through survey based technique for the present research. As per the objective of the study, only a specific segment of the population is applicable. Therefore the data has been collected “subjectively, but from a relevant segment of population” [68], [43], [126]. In this case, the target respondents are the people from investing class i.e. the people having financial savings and the capacity to invest in various financial segments [126]. Further, the respondents of the Delhi-NCR region were selected for the study. This region has been selected for the reason that per-capita income of Delhi is three times the national average making it the highest in the country. Along with Mumbai, it also accounts for 65 percent of the equity trading volume in the country. This makes the average individual in this region financially eligible to invest in stock markets. The next step is to narrow down the investor population in this region. Since, an official list of investors is not available, the study follows the approach given by previous researchers like [68], [43], [126]. The sample composition is decided on the basis of combination of judgment and snowball sampling [126]. The criteria for selecting the respondents of the survey are as follows.

i. The respondent should be a resident of the Delhi-NCR region.

ii. The respondent must invest in the Indian equity market

---


iii. The respondent should belong to middle income level or higher (i.e. annual income equal to 2 lakhs or above)\(^8\) to ensure that she is financially capable of investing in the stock market.

Around 500 people were approached for participating in the survey. The survey was administered online as well as on one-to-one basis. 410 responses were received out of which 9 responses were incomplete in some way or the other making the final number of responses to be 401.

**4.2.1 RESPONDENT PROFILE:** (Refer figure 4.2.1)

![Figure 4.2.1: Data composition of investor sample](image)

The summary statistics show that 64 percent of the respondents fall in the age group of 41-60 years (with 31.20 percent in the age group IV 1-50 years and 33.20 percent in 51-60 years). The sample contains 64.6 percent male and 35.4 percent female. More than half of respondents are post-graduate (53.4 percent) with 55.9 percent having annual income between 6-11 lakhs.

---

\(^8\) According to National Council of Applied Economic Research (year 2007-08) the middle income category lies anywhere between 3830 and 22,970 U.S. dollars annually which is equal to 2.8-14.3 lakhs in I.N.R (Shukla and Purusothaman, 2008).
Professionally 26.20 percent respondents are financial experts, 27.2 percent are in Government services or public sector units (excluding banks), 17 percent are in banks (including both public and private), 12.5 percent are in private firms (excluding banks) and the rest belong to other profession (e.g. self employed individuals). It is seen that there is a high preference of respondents to invest in stocks or mutual funds of old companies with high growth (57.60 percent) followed by new companies (27.9 percent). Further, the trading experience of 24.4 percent respondents is 3-5 years, 27.9 percent have an experience of 5-7 years and 31.2 percent have an experience of more than 7 years. When asked about the determinants of investment, 39.9 percent respondents reported that they make their investments based on market movements and 35.9 percent invest when surplus funds are available with them. Finally the frequency of trading of 27.20 percent is 3-12 months, 31.40 percent have a trading frequency of 12-36 months and 18 percent are intraday traders. It is to be noted that the survey only focuses on the behavior of retail investors. The survey of marginal or institutional investors is not included as it demands a much detailed attention and its results can become a part of separate study. Presently, due to the time and resource constraints these investors are not taken into consideration.

4.2.2 SURVEY INSTRUMENT

Descriptive Research was undertaken to investigate behavioral biases in the investors with the help of a structured questionnaire. This questionnaire consisted of 36 items that are divided into three sections. The first section consists of 10 items that provide personal information including details about name, age, gender, education, annual income, trading experience, frequency of trading etc. which helps the researcher in making a demographic profile of the investor. The rest of the items are scenario based questions that make the respondents relate to hypothetical stock market situations. The scenarios are constructed in a manner that their response reflects their underlying behavioral biases. The scenario based questions are divided into two parts, Part A and B. Part A contains a mix of open ended and close ended questions. For part B, a five point Likert scale is used that range from 1 (strongly disagree) to 5 (strongly agree). In all there are 6 items pertaining to optimism (pessimism), 6 items for overconfidence, 5 items for disposition effect, and 7 items for herding and 2 control items to verify the reliability of investor responses.
The item code of each statement along with the bias it captures is mentioned in table 2. The questionnaire is finalized after judges’ validity that includes academic as well as an industry expert. Further, the reliability of the questionnaire is verified with the help of Cronbach’s alpha.

**Description of items in each bias category: refer table 4.2.3**

This part describes the survey instrument in greater detail. To the best of researcher’s knowledge, there are no survey instruments that collectively analyze all the four biases given in the objective of the study. Therefore, we try to tackle each bias individually and develop hypothetical situations based on previous literature.

---

### i. Overconfidence

In order to capture overconfidence, we ask the respondents to estimate the precision of their knowledge to specific stock market questions. This has been achieved with the help of items mentioned by [105]. Here, we first ask them to give an estimate on stock market values like its average return and prospect of gold prices and then ask how sure they are about their response (A4, A5). The certainty about an unrealistic response shows the level of overconfidence in respondents. The other situations where this bias can be measured are given by [3]. We modify these statements to make it relevant for Indian scenario. These items capture the investors’ perception of the accuracy of their knowledge (B1), ability of picking better stocks than others (B2), taking full control and responsibility of their portfolio performance (B3) and their efficiency of trading (B5).

---

### ii. Optimism (pessimism)

For investigating this bias, the investors are asked about their outlook of the Indian equity market in the near future (A1). This is a closed ended item and the responses range from very optimistic to very pessimistic. In another item, they were asked to give a guess estimate of the average return of the Indian equity market for last 15 years. This is an open ended question where the responses are categorized into three groups; realistic estimate, pessimistic estimate and optimistic estimate. Thus, when investors’ guess a better than average estimate for this value they are considered to be optimistic. In contrast, lower than the average estimate shows the presence of pessimism. The respondents’ perspective on the Indian equity market is crosschecked with their outlook on gold prices (A4). Studies show that investment demand for gold increases when there
is an uncertainty in the stock market\(^9\). Gold provides a safe haven in times of risk and thus optimistic estimates of gold prices can be related to less optimistic outlook of the stock market by the investors. Further, the respondents are asked to agree or disagree on statements such as, do they plan to increase their investment in stock market in next quarter (B7) and do they think that if NSE drops by a certain percent it would recover within few days (B16). These statements also provide the optimistic or pessimistic outlook of investors and are in the lines of survey conducted by [3].

### iii. Herding

Herd behavior exists wherever people have the tendency to mimic the crowd. This bias is determined by detecting whether the respondents make their investment decisions with their own knowledge and understanding or do they rely on the judgment of others (A7 and A8). The items A9 and A10 identify this tendency by asking the respondents to rate the opinion of their peers and market experts in the level of importance in their decision making process. The respondents are also asked if they agree to the fact that a loss in the group is less disappointing than individual loss (B 14 and B15). Respondents who agree to these statements have the tendency to follow the crowd. The herd mentality is also seen when investors buy stocks just because many “buy” orders were placed on that stock (B13).

### iv. The disposition effect

The statements on the disposition effect come from its definition given by [136]. This bias has two sides, where one side deals with the investors’ tendency to keep stocks that have dropped in value and the other side deals with early selling shares whose prices have increased. This definition has been framed in the form of statements on Likert scale and the respondents are asked to agree or disagree to the same (B9 and B12). In a different frame, we present two hypothetical situations in which the stock has first dropped in value (A7) and then the value of that share has increased (A8). In these closed ended questions, the responses are framed in a way

---


that they capture the two sides of the disposition effect as well as the herding tendencies of investors. These situations are similar to the ones given by [22]. Prior literature also shows that this bias can increase the momentum of the stock market [146] and investors prone to this bias increase their trading activity in their winning stocks [160]. This behavior is also captured on the Likert scale (B6).

The summary of items in each bias category along with their references is mentioned in table 4.2.2

<table>
<thead>
<tr>
<th>Bias</th>
<th>Item</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overconfidence</td>
<td>A3, A5, B1, B2, B3, B5</td>
<td>Mangot (2009), Awan et al. (2010)</td>
</tr>
<tr>
<td>Optimism (pessimism)</td>
<td>A1, A2, A4, B7, B16</td>
<td>Chopra (2011), Awan et al. (2010)</td>
</tr>
<tr>
<td>Herd Behavior</td>
<td>A7, A8, A9, A10, B4, B15, B16</td>
<td>Menkhoff et al. (2005), Lutje and Menkhoff (2003), Hon-Snir et al. (2012)</td>
</tr>
<tr>
<td>Disposition Effect</td>
<td>A7, A8, B6, B9, B12</td>
<td>Shefrin and Statman (1985), Bodie et al. (2009), Shumway and Wu (2006), Statman et al. (2006)</td>
</tr>
</tbody>
</table>

4.2.3 STATISTICAL TESTS:

This part describes the statistical techniques used to analyze the responses collected with the help of a questionnaire. These tests aim at fulfilling the objectives mentioned in section 3.

i. CHI SQUARE TEST: Chi square tests are conducted when independent variables are categorical. It reflects the dependence of one variable on another categorical variable. It assists us in determining whether a systematic association exists between the two variables. It is used to compare the proportions between two or more groups simultaneously [103]. In the present study it has been conducted to check whether the responses to scenarios in Part A and B vary with each demographic and investor sophistication variable. It signifies whether investor’s decision making process is independent of these variables. The null hypothesis for the test $H_0$ implies that there is no association between demographic variables and behavioral biases.
H0₂ represents that there is no association between investor sophistication variables and behavioral biases.

ii. **INDEPENDENT SAMPLE t-TEST:** It is used to test whether the mean values of two separate groups differ significantly with each other. In the present case, this test is conducted to check if the underlying bias of respondents varies by demographic and sophistication subgroups. This test gives an idea about the specific characteristics of respondents that are associated with a particular bias. Here the null hypothesis is as follows.

H0₃: There is no difference between the mean responses of two groups.

iii. **ONE SAMPLE t-TEST:** It is a technique for testing the significance of the mean value of a distribution. Here it is applied to each item in Part-A and Part-B separately so as to analyze if the mean response significantly varies from neutral responses. This technique helps in unraveling the underlying bias of respondents in each statement.

iv. **RANKING OF THE MEAN VALUES IN ACCORDANCE OF IMPORTANCE:** It is used to find out which bias is given the highest importance by the respondents. For this purpose, the mean level of importance of each item on the Likert scale in Part-B is calculated. Next, the significance of these values is checked using one sample t-test. Here, the items with insignificant means and control items are ignored from further analysis. The remaining items are then ranked according to their means. The ranking is done in three ways. First, an overall ranking of all the items is conducted. This method takes into account all the biases at once. However, it cannot reveal which bias is the most prominent as the items representing the same bias can have different ranks. The second method involves categorization of items into four groups, each representing a particular bias. Then the ranking of items within each group is done such that each bias has a statement which is given the highest importance by the respondents. For instance, out of the three items that represent overconfidence, one with the highest mean is ranked first followed by the next highest. In the third method, the means of items in each bias category are consolidated to get a single value. Then ranking is done with rank 1 given to the bias with highest mean and rank 4 given to bias with lowest mean. This method helps in identifying which bias is most prominent amongst the investor.