CHAPTER 4

RESEARCH METHODOLOGY

This chapter deals with the research methodology applied for fulfilling the objectives of the study. To analyse the performance of mutual funds secondary data has been used and to study investors’ behaviour, primary data has been used. Attributes taken for analysing the mutual fund’s performance, data used and its collection, time period of the study, population, sample, sampling frame and the tools and techniques employed for analysing the data have been discussed in the present chapter in detail.

4.1 ATTRIBUTES

The past literature has shown that the performance of mutual funds is related to their characteristics, defined as their attributes. Most of the studies as discussed in chapter 3 have investigated the impact of two or three attributes of mutual funds on their performance. Based on the literature review the researcher has considered a wide range of attributes of mutual funds viz., load status, expense ratio, minimum initial investment, risk, age of the scheme, asset size, asset ratio and past performance of the mutual fund schemes. Thus, the relationship of various attributes of the mutual fund schemes have been analysed with their performance in terms of efficiency. Specific variables identified on the basis of literature review have been explained below. Further, these attributes along with their expected relation with the performance have been summarised in table 4.1.

i._LOAD STATUS

Load status has been taken as a dummy variable in many past studies done by Ippolito (1989), Droms & Walker (1994), Philpot et al.(1998), where its value was taken as unity for mutual fund schemes with load fee and 0 otherwise [87], [56], [125]. In the present study, load status has been coded as 1 if the scheme charges any load fee and is coded as 0 if it is a no load mutual fund scheme. (Annexure B).

ii. EXPENSE RATIO

The expense ratio is per unit cost incurred in managing the mutual fund i.e., it is the total amount of annual expenses incurred by the mutual fund divided by their assets under
management. These annual expenses include the management fee and operating expenses as the registrar and transfer agent fee, audit fee, custodian fee, and marketing and distribution fee. The expense ratio of a mutual fund is disclosed yearly. Philpot J. (1998) had analysed 27 bond funds by taking the average of yearly expense ratio under the period studied [125]. In the present study, expense ratio has been obtained by the average of six years expense ratios of mutual fund schemes i.e., April, 2006 to March, 2012. (Annexure B).

iii. MINIMUM INITIAL INVESTMENT

Minimum initial investment is the minimum amount needed initially by the investors to invest in a mutual fund scheme. It has been considered by many researchers as McMullen & Strong, 1998 and Galagedera & Silvapulle, 2002, to assess the efficiency of the mutual funds [114], [66]. (Annexure B).

iv. RISK

For studying the risk and performance relationship of mutual funds, mostly two risk measures as standard deviation ($\sigma$) and beta ($\beta$) have been used in the past researches. Some authors as Sharpe (1966), McMullen and Strong (1998), Anderson et al. (2004), Karlsson and Persson (2005) and Chen at el. (2011) took the standard deviation of mutual fund’s annual returns as a measure of risk [142], [114], [6], [97] and [40]. On the other hand, in some studies by Dellva & Olson (1998), Chang’s (2004) and Tsolas (2011), $\beta$ has been considered as the measure of risk [51], [37] and [156]. Some other authors as Lin and Chen (2008) and Soongswang & Sanohdontree (2011) have taken both standard deviation and $\beta$ for studying the risk return relationship in mutual fund schemes [107] and [150]. In the present study, standard deviation ($\sigma$) has been taken as a risk measure in Logistic Regression and beta ($\beta$) has been used as its measures in DEA.

Standard Deviation or Total Risk of Portfolio: Standard deviation is a statistic to measure the variation in individual returns from the average expected return over a certain period. It represents the total risk of the portfolio. This total risk consists of two components namely diversifiable and non diversifiable risk. Diversifiable or unsystematic risk represents the portion of a portfolio risk that can be eliminated by holding enough stocks i.e., through diversification. This risk results from uncontrollable events unique to an industry or a company. Non diversifiable or systematic risk is external to an industry and is attributed to
broad forces that impact all investments. The total risk or standard deviation is represented by sigma ($\sigma$) and is defined as the square root of the mean of the squares of deviation of individual returns taken from the average return. It has been calculated as:

$$\sigma_p = \sqrt{\frac{1}{n}\sum (R_{pt} - \bar{R}_p)^2}$$ ..........................(i)

Where,

$\sigma_p$ is the total risk or standard deviation of the mutual fund schemes (portfolio)

$R_{pt}$ is the return of the sample mutual fund schemes

$\bar{R}_p$ is the average return of the mutual fund schemes

$n$ is the number of period of the study

The standard deviation ($\sigma$) of all the sample mutual fund schemes has been calculated on the yearly returns (Annexure C).

**Systematic Risk or Beta ($\beta$):** Systematic risk is that component of total portfolio risk which is not controlled through the process of diversification. It is non diversifiable risk and shows how the price of a security responds to market forces. The more responsive the price of a security is to changes in the market, the higher will be its beta. It is calculated by relating the returns on a security with the returns for the market. Market return is measured by the average return of a large sample of stocks such as Sensex or BSE 100 index. The beta for overall market is equal to 1 and other betas are viewed in relation to this value. Mutual fund can be as volatile, less volatile or more volatile. For example, if a mutual fund has below market beta of 0.84, then the fund has 84 percent of the volatility of the market. That is relative to the market index it would decline by the 84 percent of the drop in down markets and would capture only 84 percent of the gain in the index in up markets. In order to obtain beta ($\beta$) of the portfolio, CAPM version of the market model has been applied. The estimation form of CAPM is:

$$R_{pt} = \alpha_p + \beta_p R_{mt} + \epsilon_p$$ ..........................(ii)

Where,

$R_{pt}$ is the return on the mutual fund schemes for time $t$

$R_{mt}$ is the return on the market index for time $t$
\( \alpha_p \) represents the coefficient term

\( \beta_p \) beta coefficient, the measure of sensitivity

\( \hat{\epsilon}_p \) is the error term

Beta (\( \beta \)) is a relative measure unlike absolute measure (\( \sigma \)). Higher value of beta indicates a high sensitivity of fund returns against market returns and lower value indicates a low sensitivity. The parameters in the above model have been estimated by using standard regression methodology (Annexure B). The regression equation also provides the value of coefficient of determination, \( R^2 \) which states the extent of diversification of mutual fund scheme portfolio against the market portfolio. That is, it shows the extent of variation in the mutual fund scheme return that has been explained by the market factor thereby signifying the diversification level of the scheme which indicates the percentage of a security’s risk which cannot be eliminated through further diversification.

In precise percentage terms, \( R^2 \) indicates how a security’s performance variation paralleled the market over the same time period. Lower figure indicate less correlation with the market, and hence lower significance of beta and vice-versa. A completely diversified security would be perfectly correlated with the market, and would have an R-squared of 100 percent. In other words, the security’s results have so perfectly replicated the market’s results, that the security is 100 percent as diversified as the security’s related index. For example, a security with R-squared of 80 percent is only 80 percent as diversified as security’s related index i.e., 80 percent of the security’s risk is market related, and the other 20 percent is attributable to the security’s unique characteristics. Otherwise a high \( R^2 \) indicates that fund is well diversified.

v. **RISK ADJUSTED RETURN**

Risk adjusted return of the mutual fund schemes under study has been computed by Sharpe Ratio (\( S_p \)). Sharpe Ratio (Sharpe, 1966), is the most commonly used risk adjusted performance measure and has been employed in many researches done by Lhabitant (1995), Droms & Walker (1994, 1996) and Philpot J. (1998) [103], [56], [58] and [125]. It is the spread between fund’s unadjusted total return and the risk free rate, divided by the standard deviation of the fund’s unadjusted total return. This ratio is also called as the reward to
variability ratio. Sharpe ratios for the sample mutual fund schemes have been computed by using the following equation (Annexure C):

\[ S_p = \frac{(R_p - R_f)}{\sigma_p} \] ...........................(iii)

Where,

\( S_p \) stands for the Sharpe ratio of the mutual fund schemes for the duration April, 2006 to March, 2012.

\( R_p \) is the average yearly return on the mutual fund scheme\(^{12} \) from April, 2006 to March, 2012

\( R_f \) is the average risk free rate of return (91 days T-Bills) from April, 2006 to March, 2012

\( \sigma_p \) stands for the total risk or the standard deviation of the yearly returns of portfolio

vi. **JENSEN’S ALPHA**

This measure of mutual fund’s performance was given by Michael C. Jensen in 1968 and is a regression of excess fund return with excess market return. It analyses the portfolio

\(^{12} \) \( R_p \), the average return on the portfolio has been calculate by the following method-

**Step 1:** Monthly rate of return for all the sample mutual fund schemes from April, 2006 to March, 2012 has been computed as \( r_t = \frac{(NAV_t + D_t)}{NAV_{t-1}} - 1 \) Where, \( NAV_t \) is the NAV of the scheme at month end \( t \), \( NAV_{t-1} \) is the NAV of the scheme at month end \( t-1 \) and \( D_t \) are the ex-dividend of the scheme in period \( t \). The NAVs are adjusted for dividends assuming dividends are reinvested at the ex-dividend rate. Rate of return was calculated in this manner on monthly basis for 72 months from April, 2006 to March, 2012. According to Association of Investment Management and Research Performance Presentation standards (AIMR-PPS), monthly returns are geometrically linked to produce more accurate annual return calculations. (Reilly and Brown, 2003)

**Step 2:** Yearly returns for the period under study have been computed as

\[ R_t = \left[ (1+r_1)(1+r_2)\cdots\left(1+r_{12}\right)\right] - 1 \]

\( R_t \) is the return for year \( t \), where, \( t = 2007, 2008, 2009, 2010, 2011, 2012 \)

\( r_1, r_2, \ldots, r_{12} \) are the monthly rate of return for the specific year under study.

**Step 3:** average return of the mutual fund scheme has been computed as

\[ R_p = \text{Average of yearly returns (} R_1, R_2, R_3, R_4, R_5, R_6 ) \]
manager’s predictive ability to achieve higher returns than expected for the given riskiness. The Jensen’s model has been expressed below:

\[ R_{pt} - R_{ft} = \alpha_p + \beta_p (R_{mt} - R_{ft}) + \epsilon_{pt} \] ...................................(iv)

Where,

- \( R_{pt} \) = return on mutual fund scheme for the year t
- \( R_{ft} \) = risk free return for the year t
- \( R_{mt} \) = return on the market portfolio in year t
- \( \alpha_p \) = Alpha, the intercept of the regression equation and measures the stock selection capability
- \( \beta_p \) = Beta of the portfolio
- \( \epsilon_{pt} \) = Error term

The intercept Alpha, \( \alpha_p \) provides Jensen’s measure of performance and has been used by many researchers worldwide as, Lhabitiant (1995), Dellva et al., (1998), Barinov, (2003), Peng (2004) and Christensen (2005) [103], [51], [11], [123] and [45]. Annexure C provides the alpha values with t-stat for all the mutual fund schemes taken for the study.

vii. AGE OF THE MUTUAL FUND SCHEMES

Malhotra & McLeod (1997) took the age of the mutual funds as the number of years since the formation of the fund [110]. Authors stated that the fund age may be non-linearly related to the expense ratio of the fund and therefore, they took the natural logarithm of age in their regression model. Redman & Gullett (2007) and Belgacem & Hellara (2011) hypothesised a non-linear relationship between fund age and returns and took the log of fund age in their regression model [133] and [14]. For computing this attribute, first the age has been calculated in years since the inception date of the mutual fund scheme till 31st March, 2012 and then natural logarithm of the fund’s age has been taken (Annexure D).

viii. ASSET SIZE

Asset size of a mutual fund is the total market value of all the securities held in its portfolio and represents the size of the mutual fund scheme. As described by the Association
of Mutual Funds of India, it is the Asset under Management of the mutual funds. Philpot J. (1998) Dowen & Mann (2004), Redman & Gullett (2007) and Belgacem & Hellara (2011) have taken natural log of the mutual fund’s assets for studying the relation between fund’s return performance and their asset size [125], [55], [133] and [14]. In the present study this attribute has been computed by taking the natural logarithm of the mutual fund’s assets as on 31st march, 2012. (Annexure D).

ix. **ASSET RATIO**

Asset ratio of the mutual fund has been calculated as shown below-

\[
\text{Asset Ratio} = \frac{\text{Total Assets as on 31 March, 2012}}{\text{Total Assets as on 31 March, 2011}}
\]

Asset ratio above 1 indicates a positive asset flow and if it is less than one then it indicates a negative asset flow. (Annexure D). Similar method was adopted by the author Chehade R. T. (1998) for finding the asset ratio [39].

x. **PAST PERFORMANCE**

Past performance of the mutual funds has been measured by Sharpe Ratio \( S_p \) which has been computed by applying equation (ii) above as (Annexure E),

\[
S_p = \frac{R_p - R_f}{\sigma_p}
\]

Here,

- \( R_p \) is the average yearly return on the mutual fund scheme from April, 2006 to March, 2011
- \( R_f \) is the average risk free rate (91 days T-Bills) of return from April, 2006 to March, 2011
- \( \sigma_p \) stands for the total risk or the standard deviation of the yearly returns of portfolio
### Table 4.1: Attributes of Mutual Funds Performance

<table>
<thead>
<tr>
<th>ATTRIBUTES</th>
<th>SYMBOL USED</th>
<th>DESCRIPTION</th>
<th>EXPECTED RELATION WITH PERFORMANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load Status</td>
<td>LOAD</td>
<td>A dummy variable coded 1 if the fund scheme charges any load, 0 otherwise</td>
<td>Positive</td>
</tr>
<tr>
<td>Expense Ratio</td>
<td>EXPENSE</td>
<td>The average of yearly expense ratios from April, 2006 to March, 2012</td>
<td>Positive</td>
</tr>
<tr>
<td>Min. Initial Investment</td>
<td>MINII</td>
<td>Minimum amount in Rs. Needed initially for investment</td>
<td>Positive</td>
</tr>
<tr>
<td>Risk</td>
<td>RISK</td>
<td>Values of standard Deviation (σ) and beta (β) of the mutual fund</td>
<td>Positive</td>
</tr>
<tr>
<td>Risk Adjusted Return</td>
<td>SHARPE</td>
<td>Sharpe ratio for the mutual fund schemes from April, 2006 to March, 2012</td>
<td>Positive</td>
</tr>
<tr>
<td>Jensen’s Alpha</td>
<td>ALPHA</td>
<td>Value of coefficient alpha, α in equation, (ii)</td>
<td>Positive</td>
</tr>
<tr>
<td>Age of the scheme</td>
<td>LAGE</td>
<td>Natural logarithm of the age (yrs.) of fund schemes till 31st March, 2012</td>
<td>Positive</td>
</tr>
<tr>
<td>Asset Size</td>
<td>ASSETS</td>
<td>Natural logarithm of fund’s assets (AUM) as on 31st March, 2012</td>
<td>Positive</td>
</tr>
<tr>
<td>Asset Ratio</td>
<td>ASSETR</td>
<td>Ratio of total assets on 31 March, 2012 and 31 March, 2011</td>
<td>Positive</td>
</tr>
<tr>
<td>Past Performance</td>
<td>LSHARPE</td>
<td>Holding period Sharpe measure lagged one period (April, 2006 to March, 2012)</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Thus, LOAD, EXPENSE, MINII, RISK, SHARPE, ALPHA, LAGE, ASSETS, ASSETR, and LSHARPE are the attributes of mutual funds analysed to study their performance in terms of efficiency. For this purpose, hypotheses have been framed and have been discussed in the next section.

### 4.2 HYPOTHESES

The first objective of the research is to study the performance of mutual funds in India. To analyse the first objective, the researcher has formulated following hypothesis:

- $H_0$: The sample mutual fund schemes do not perform efficiently.
- $H_1$: The sample mutual fund schemes perform efficiently.
Hypothesis $H_0$ has been investigated by employing Data Envelopment analysis (DEA) in which attributes LOAD, EXPENSE, RISK ($\beta$) and MINII have been taken as input variables and SHARPE and ALPHA have been taken as output variables. DEA methodology has been discussed further in later section in detail.

Second objective of the research is to study the performance of mutual funds with respect to different performance attributes. For fulfilling this objective, attributes considered are LAGE, ASSETS, ASSETR, LSHARPE and RISK ($\sigma$). Five hypotheses have been formulated for studying mutual fund’s performance with respect to their attributes as-

$H_{0a}$: Age of the mutual fund schemes is not related to their efficiency.

$H_{1a}$: Age of the mutual fund schemes is related to their efficiency.

$H_{0b}$: Asset Size of the mutual fund schemes is not related to their efficiency.

$H_{1b}$: Asset Size of the mutual fund schemes is related to their efficiency.

$H_{0c}$: Asset Ratio of the mutual fund schemes is not related to their efficiency.

$H_{1c}$: Asset Ratio of the mutual fund schemes is related to their efficiency.

$H_{0d}$: Past performance of the mutual fund schemes is not related to their efficiency.

$H_{1d}$: Past performance of the mutual fund schemes is related to their efficiency.

$H_{0e}$: Risk of the mutual fund schemes is not related to their efficiency.

$H_{1e}$: Risk of the mutual fund schemes is related to their efficiency.

To examine these hypotheses, population, sample, data collection and techniques employed for analysing the data have been discussed below in further sections.

4.3. TIME PERIOD AND POPULATION

To study the performance of Indian mutual fund industry, a time period of six years (April, 2006 to March, 2012) has been taken. Hence, all the open ended mutual fund schemes that were operational on 1 April, 2006 are the population for the study. According to AMFI, there were 463 open ended mutual fund schemes as on March 31, 2006, which have been
classified into growth schemes, income schemes, balanced schemes, liquid schemes and ELSS schemes. All the 463 mutual fund schemes comprise the population for study.

Further, to study the investors’ behaviour, primary data is considered. The present study aims at studying the perception of individual investors towards the mutual funds with respect to other investment options and towards the various attributes of the mutual funds. This study explores the reasons that stop investors from investing in mutual fund schemes. Also, various steps or measures that may be taken by the companies and policy makers for motivating investors to start investment in mutual fund schemes have been looked at.

Therefore, all the ‘mutual fund investors (MFI)’, i.e., those who invest in the mutual funds and ‘non mutual fund investors (NMFI)’, i.e., those who do not invest in mutual funds are the universe or population for the study. The data required from the investors is very sensitive and indicative in nature as it comprises of the information regarding their savings and investments. Most of the investors are hesitant in discussing their investment and related issues. Also, they may find it unsecure to provide this kind of data on mails and phone calls. Hence, only possible way to collect this data was the personal interaction with the investors. Therefore, due to the limitation of time and resources, the scope of study related to the investors has been limited to the investors in Delhi NCR region. Sample selection from this population has been discussed further.

4.4. SAMPLE

In this section, the sample of secondary and primary data taken for the study and their sampling frame has been explained.

4.4.1 SECONDARY DATA SAMPLING

For selecting the sample mutual fund schemes, following criteria have been considered:

a. Only those mutual fund schemes that were launched before March, 2006 have been considered. The reason is, present study analyses the performance of mutual funds in India and therefore, a long term period of six years has been considered. Mutual fund schemes launched after March, 2006 have the operation period of less than six years and hence have been excluded from the study.
b. Only open-ended mutual fund schemes have been considered because they possess several advantages over close-ended mutual funds. As, Open ended funds are sold and redeemed everyday on an ongoing basis, the investors are free to buy and sell any number of open ended mutual fund’s unit at any point of time, these funds are available for subscription all throughout the year and do not carry any fixed maturity. Also, open ended mutual fund schemes offer high level of liquidity to the investors. Also, many empirical studies by Sharpe (1966), Jensen (1968), Henriksson (1984), Cai et al. (1997), Otten and Balms (2002) and Bauer et al. (2006) have considered open ended mutual funds [142], [91], [83]. [26]. [118] and [13]. In addition to this, data for very few close ended mutual funds schemes i.e., 15 was available. This comprises a very small number as compared to the total sample. Therefore, these schemes have been excluded from the study.

c. According to AMFI, there were 463 open ended mutual fund schemes as on March 31, 2006 consisting of 190 growth schemes, 139 income schemes, 34 balanced schemes, 45 liquid/money market schemes, 29 gilt schemes and 26 ELSS. A number of these schemes have been redeemed during the said period or have closed down due to poor performance. Also, some schemes have been merged with the existing one. All such schemes which have been redeemed, closed or merged have been excluded from the study.

d. This study encloses two veins for mutual fund performance. First, the performance of mutual funds with various inputs and outputs has been tested on the basis of their efficiency and inefficiency. Second, relationship of the performance of mutual funds with their attributes has been analysed. Therefore, apart from NAV, data for other attributes as asset size, expense ratio, load status, minimum initial investment, age of the mutual fund schemes and so on is also required. Therefore, only those mutual fund schemes for which complete data set is available for the period of study i.e., from April 1, 2006 to March 31, 2012, have been considered. For example, if for any mutual fund scheme, data for NAV and asset size is available but the data for expense ratio is not assessable, then that scheme has been excluded from the sample.

e. Also, those schemes which invest some percentage of its corpus in equities and have primary or secondary objective as capital appreciation through investments in equities have been considered for the study. Liquid/money market funds and gilt funds invests exclusively in safe instruments like commercial papers, treasury bills, government securities, etc., and therefore these schemes have been excluded from the sample. Also, Debt funds generally
invest in fixed income securities such as corporate debentures, government securities, bonds and money market instruments. These funds cannot be compared to a diversified or equity mutual fund as the dominant objective of debt funds is capital preservation rather than the appreciation [20]. Further, it would be inappropriate to employ risk adjusted measures of Sharpe and Treynor ratios for performance evaluation of debt funds because these measures are based on the excess return over risk free rate for a given unit of risk. Bond portfolios comprise of mainly government securities and therefore, these are inherently more or less risk free [54]. Hence, debt funds have also been excluded from the sample. However, monthly income plans (MIP) has been considered for the study as these are a part of income schemes and have a small exposure to equities. The portfolio of a typical MIP comprises an exposure of 10-15 percent to equities and 80-85 percent to debt and money market instruments.

By taking the above said criteria in account, out of 463 open ended mutual fund schemes as on March 31, 2006, 119 schemes fits into the sample. Therefore, for studying the performance of mutual funds, 119 mutual fund schemes has been analysed. These are the open ended schemes from Growth, Income, Balanced and ELSS investment styles. (Annexure A). Table 4.2 provides the sampling frame for the mutual fund schemes studied.

Table 4.2: Sampling Frame for the mutual fund schemes

<table>
<thead>
<tr>
<th>Investment Style</th>
<th>Number of Mutual Fund Schemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equity/Growth Funds</td>
<td>48</td>
</tr>
<tr>
<td>Income Funds</td>
<td>30</td>
</tr>
<tr>
<td>Balanced Funds</td>
<td>23</td>
</tr>
<tr>
<td>Equity Linked Saving Schemes (ELSS)</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>119</td>
</tr>
</tbody>
</table>

4.4.2 PRIMARY DATA SAMPLING

For studying the behaviour of individual investors towards mutual funds and other investment options, primary data is required. Stratified random sampling has been used for selecting the respondents. Initially four major categories have been formed out of the total population as Government sector, Private sector, Business and Professionals. Further, each major category has been subdivided.
Government sector has been divided into Central government, State governments and public sector units and from each 50 respondents were selected through random sampling. For doing this, five departments each of central and state government and PSU were selected randomly. The employees list of central government, state government and public sector undertakings offices in Delhi National Capital Region (NCR) has been taken out from the Government of India (GOI) Web Directory13, Department of Public Enterprises14, Delhi helpline15, yellow pages directory16 and Wikipedia17. In this way 150 respondents from government employee section were selected.

From the private sector, four industries i.e. Information Technology (IT) Industry, Banking and Insurance Industry, Higher Education Industry and Health Industry, have been selected randomly as these industries cover most of the employees from private sector. A list of 20 IT companies in Delhi NCR region was taken out from the Dataquest18, one of the most reliable data set for IT Industry. Using stratified random sampling, 50 employees from these companies were selected for the study. List of private sector banks and insurance companies was availed from the yellow pages19 and Delhi helpline20. 50 employees were selected using stratified random sampling. List of universities and colleges was obtained from the website of University Grants Commission21 from which 50 employees were selected. In the same way, 50 employees from the list of private hospitals, availed from the websites of Delhi helpline22.

13 http://goidirectory.nic.in/search_result.php
14 http://dpe.nic.in/
16 http://delhi.clickindia.com/general/governmentoffices.html
18 http://www.dqindia.com/
earch=life%20insurance%20companies
21 http://www.ugc.ac.in/
and just dial\textsuperscript{23} were selected. Thus, by following the above mentioned process, 200 respondents from private sector were selected.

From Business category, five businesses have been picked randomly as Retailers, Wholesalers, Tour and Travel, Real Estate and Home Interiors. During the initial stage of research, it was found that the general, less educated investors are ignorant and dependent on others in making fund investments decision. Rather they are easily lured and motivated lot to get their investment made in any recommended mutual fund. Hence to make this study meaningful, the focus was then shifted towards educated and informed investors. Therefore, in this category, 50 respondents from retail, wholesale, tour and travel, real estate and home decor industry were contacted. The same kind of approach was adopted by Singh J. (2004) for conducting the primary survey regarding investors’ attitude for mutual funds. In the same way from the professional category, Private Doctors, coaching institutes, Chartered accountants (CA), Company secretaries (CS), lawyers and advocates and freelance writers have been approached. From every category hundred respondents have been targeted making the sample size to 500.

Thus, total 500 investors from major cities of NCR region as Gurgaon, Faridabad, Delhi, Meerut, Ghaziabad, Noida and Greater Noida were targeted. One major reason for selecting the NCR region is the nature of data required and the time and resource constraints. This data requires the information regarding investments by respondents and their behaviour towards mutual funds. In general, investors are hesitant in discussing the issues relating to their saving and investments. Therefore, most of the respondents were approached personally so that they can be convinced for giving the data for the research. Hence, because of the kind of approach adopted to collect the data with the time and resource constraint, the NCR region was selected for the study. Figure 4.1 below describes the sampling frame of the primary data collected for the study in detail.

\textsuperscript{23} http://www.justdial.com/Delhi/private-hospitals-%3Cnear%3E-noida
Figure 4.1: Sampling Frame for collecting the Primary Data

Total Number of Respondents: 500
4.5 COLLECTION OF DATA

4.5.1 COLLECTION OF SECONDARY DATA

For analysing the performance of mutual funds, secondary data has been collected from the various published sources for six years, i.e., April 1, 2006 to March 31, 2012. Method and process of collecting the data has been discussed below in detail.

Data for all the attributes discussed above i.e., SHARPE, ALPHA, LSHARPE, ASSETS, ASSETR, EXPENSE, LOAD, RISK (β & σ), LAGE and MINII, has been gathered for 119 sample mutual fund schemes. Some of the data as expense ratio is being declared on yearly basis. Therefore, for maintaining the unanimity in the data for study, yearly data has been used. For getting the yearly return, month end Net Asset Value (NAV) of the sample mutual fund schemes has been collected for the period of study i.e., from April 1, 2006 to March 31, 2012 from the Alpha data base of Centre for Monitoring Indian Economy (CMIE) and the website of Association of Mutual Funds of India (AMFI).

For calculating RISK (beta, β) Bombay Stock Exchange (BSE) Sensex Index has been considered as a surrogate for market portfolio. BSE Sensex has been a widely accepted market proxy amongst the practitioners and researchers in India as Chander (1999) and Kaur A. (2011) [32] & [98]. BSE Sensex, the 30-shares sensitive index, consists of the most heavily traded stocks in the BSE, the country’s premier stock market. It came into existence in 1986 and provides data over a long period of time, 1979 onwards. It is scientifically designed and is based on the globally accepted construction and review methodology. Since September 1, 2003, the index is calculated based on ‘Free-float market capitalisation’ methodology. This methodology of index construction is regarded as an industry best practice globally and is used by all major index providers like MSCI, FTSE, STOXX, S &P and Dow and Jones. For getting the yearly return on the market, index values have been taken from the official website of Bombay Stock Exchange.

Yield on 91-day Treasury Bills (T-bills) has been used as a surrogate for risk free rate of return as has been done by most of the researchers’ world over. The documented and referred literature by Lee & Rahman (1990), Kao et al. (1998), Chander (2000), Gupta (2001),

---

24 Free float market capitalisation is used as weight age to determine the overall change in value. Free float refers to the overall market capitalisation minus market capitalisation of promoter holding. Promoter holdings are equity shares that are available for trading and are not available in the market to invest. (Khedekar, 2007).

25 (http://www.bseindia.com)
Tripathy (2004), Muthappan & Damodaran (2006), Kaur A. (2011) etc, is in favour of return on 91-day treasury bill [105], [96], [33], [79], [155], [116] and [98]. Yield on 91-day treasury bills (T-91), has been collected from the RBI Bulletin. The data for ASSETS, ASSETR, EXPENSE, LOAD, LAGE, LSHARPE and MINII has been collected from various online sources as website of mutual fund companies, AMFI and value research online. Table 4.3 provides the detail of the databases referred and the data collected:

<table>
<thead>
<tr>
<th>Data bases</th>
<th>Data taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha data base of CMIE</td>
<td>NAV, asset size, expense ratio, minimum initial investment and inception date</td>
</tr>
<tr>
<td>Website of AMFI</td>
<td>NAV, Asset size and inception date</td>
</tr>
<tr>
<td>Website of Bombay Stock Exchange</td>
<td>Return on market index (BSE Sensex)</td>
</tr>
<tr>
<td>RBI Bulletin</td>
<td>Yield on 91 day T-bills</td>
</tr>
<tr>
<td>Websites of individual mutual funds</td>
<td>Asset size, expense ratio and inception date</td>
</tr>
<tr>
<td>Value Research</td>
<td>Expense ratio and Load status</td>
</tr>
</tbody>
</table>

**Table-4.3: Databases used**

4.5.2 COLLECTION OF PRIMARY DATA

Collection of primary data has been done through questionnaire to analyse the behaviour of investors towards mutual funds and other investment options. After discussion with experts, it was found that two separate questionnaires should be used for analysing the investors’ behaviour. One for Mutual fund investors (MFI) regarding detailed questions about their investment behaviour in mutual funds while the other for non-mutual fund investors (NMFI) regarding the questions about their general investment behaviour and reasons for not investing in mutual funds. Therefore, two separate questionnaires were constructed for MFI and NMFI. These questionnaires cover different aspects of the behaviour of investors towards mutual funds and other investment options. It is being judged on the basis of various parameters as return, risk, liquidity, tax saving, procedural understanding, diversification and so on. Through these questionnaires, the investment behaviour of investors like how they take investment decision, what type of mutual fund schemes they select and the attributes they prefer has been investigated. Also, the reasons for not investing in mutual funds and the

---

26 http://www.rbi.org.in
27 http://valueresearchonline.com
measures/actions that can be taken by mutual fund companies and policy makers for motivating the investors to invest in mutual funds has been explored.

**PILOT STUDY:**

Pilot study of both questionnaires i.e., for MFI and NMFI has been done and some respondents were approached for filling up the questionnaires. On the basis of feedback received some modifications in the language of the questions and in the manner of asking questions regarding the savings and investments of respondents were done. In this way, two modified questionnaires for MFI and NMFI were prepared and the pilot survey was done. The questionnaires for MFI and NMFI have been presented in Annexure F. For pilot survey, these questionnaires were administered on 80 respondents, 40 were MFI and another 40 were NMFI and the validity and reliability was tested.

**VALIDITY:**

Validity refers to the degree to which our test or other measuring device is truly measuring what we intended it to measure. In other words, validity is the extent to which the questionnaire provides adequate coverage of the investigative questions guiding the study. That is, it represents the degree to which the content of the items adequately represents the universe of all relevant items under study. For its determination, a panel of persons judges how well the instrument meets the standards. Therefore, the validity for both the questionnaires was checked by sending them to expert’s panel and were found valid enough to conduct the study.

**RELIABILITY:**

Reliability refers to the consistency of a measure. Reliability test has been done using Cronbach alpha technique. Cronbach’s alpha (Cronbach, 1951) is a coefficient of reliability. It is a measure of internal consistency, that is, how closely related a set of items are as a group. A "high" value of alpha is often used as evidence that the items measure an underlying construct. Cronbach's alpha is written as a function of the number of test items and the average inter-correlation among the items. The formula for the same is given below:

\[
\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N - 1) \cdot \bar{c}}
\]
Here N is the number of items, c-bar is the average inter-item covariance among the items and v-bar equals the average variance. Cronbach’s alpha ranges from 0 to 1.00, with values close to 1.00 indicating high consistency. Relation between Cronbach’s alpha coefficient and internal consistency is as follows:\footnote{George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon. Kline, P. (1999). The handbook of psychological testing (2nd ed.). London: Routledge}

\[
\begin{align*}
\alpha \geq 0.9 & \quad \text{Excellent;} \\
0.8 \leq \alpha < 0.9 & \quad \text{Good;} \\
0.7 \leq \alpha < 0.8 & \quad \text{Acceptable;} \\
0.6 \leq \alpha < 0.7 & \quad \text{Questionable;} \\
0.5 \leq \alpha < 0.6 & \quad \text{Poor;} \\
\alpha < 0.5 & \quad \text{Unacceptable}
\end{align*}
\]

The value of Cronbach’s Alpha coefficient came out as 0.72 for the questionnaire of MFI and 0.77 for the questionnaire of NMFI. Therefore, both the questionnaires are found to be valid and reliable enough to be filled up by the respondents. 500 copies of final questionnaire were mailed/ distributed to 250 MFI and 250 NMFI in NCR region. In all 463 responses were received out of which 440 (88 percent, 218 MFI and 222 NMFI) were found to be usable and have been considered for the study.

4.6 DATA ANALYSIS

This section presents the tools and techniques used to evaluate the performance of mutual funds in terms of efficiency and its relation with their attributes. Also, methodology used for analysing the investors’ behaviour towards mutual funds and other investment options has been discussed.

4.6.1 PERFORMANCE OF MUTUAL FUNDS

Following models have been used for evaluating the performance of mutual funds:
4.6.1(a) DATA ENVELOPMENT ANALYSIS (DEA)

Data Envelopment Analysis (DEA) was first introduced by Charnes, Cooper and Rhodes in 1978 for measuring the productive efficiency of management or any other Decision Making Unit (DMU). It is a non parametric linear programming method and provides a comprehensive analysis of relative efficiency. Their work expanded the single input/output efficiency measure, originally developed by Farrel, to a multiple input/output efficiency measure [64]. It is the ratio of the weighted sum of outputs to the weighted sum of inputs, subject to the constraint, the efficiency of all the units is less than or equal to unity.

Basically, the methodology compares the input and output data of DMU to the data of other similar DMUs which is the measure of efficiency. Efficiency is calculated in two steps. First, an empirical production frontier is identified mathematically by the ratio of weighted sum of outputs to weighted sum of inputs and is considered the best practice frontier – a benchmark for all DMUs in the analysis. Second, efficiency for each individual DMU is calculated as its distance from the frontier. It represents the amount DMU can decrease its inputs while maintaining the current output level (input oriented DEA model) or the amount it can increase its outputs without consuming more inputs (output oriented DEA model). DMUs with efficiency score of unity, act as the benchmark for all inefficient DMUs. The necessary elements for DEA analysis as DMUs, inputs and outputs are depicted in figure 4.2.

![DEA Analysis – Necessary Elements](image)

**Figure 4.2: DEA Analysis – Necessary Elements**

*Source: Chehade, Ramez T (1998) - Data Envelopment Analysis*
Given these elements i.e., inputs and outputs, the DEA analysis produces several types of results. These results have been explained below. Further, figure 4.3 provides a graphical representation of these results through a one input, one output example.

- **Empirical Production Frontier:** It represents the “best practice DMUs” as explained earlier.
- **Efficient Projections (Targets):** These are the point of the virtual DMU on the frontier.
- **Efficiency Reference Set (Peer Group):** When talking about improving the performance of an inefficient DMU by moving it to a point on the efficient frontier, a virtual DMU is simultaneously formed as a weighted combination of some efficient DMUs. For each inefficient DMU, the set of the suitable efficient units is called its reference set, or the set of its efficient peers.

Source: Chehade, Ramez T (1998) - Data Envelopment Analysis

From figure 4.3, the best practices DMUs are b, c, d and e

Peer Groups for DMU a: Input minimization: (b, c); Output maximisation: (d, e)

Efficiency of DMU a: Input minimisation: \( \frac{X_1}{X_1^*} \); Output maximisation: \( \frac{Y_1}{Y_1^*} \)

Arrows \( A_1 \) and \( A_2 \) indicate the targets for DMU a.
i. TERMINOLOGY USED IN DEA

- Returns to Scale (RTS):
  It deals with the response in output levels due to change in inputs by some amount, i.e., \( t \geq 0 \). If all the outputs change by the same amount as the change in input \( t \), it exhibits Constant Returns to Scale (CRS). If the increase in outputs is higher or lesser than the scale of inputs, it exhibits Variable Returns to Scale (VRS).

- Increasing Returns to Scale (IRS):
  If output increases by more than the scale of inputs, then it exhibits the increasing returns to scale.

- Decreasing Returns to Scale (DRS):
  When the increase in outputs is lesser than the increase in input, decreasing returns to scale exists.

- Input Oriented Model:
  This model projects the inefficient DMU towards the frontier with a proportional decrease in all of its inputs while maintaining the same level of outputs. In figure 4.3, the projection \( A_1 \) between DMU b and DMU c represents the input oriented.

- Output Oriented Model:
  This model projects the inefficient DMU towards the frontier with a proportional increase in all of its outputs while maintaining the same level of inputs. In figure 4.3, the projection \( A_2 \) between DMU e and DMU d, represents the output oriented.

The output variables as SHARPE (Sharpe ratio) and ALPHA (Jensen’s alpha) are beyond the control of mutual fund scheme’s manager and therefore, output oriented model of DEA is insignificant for such studies [66]. In the present study, output variables are Sharpe ratio and Jensen’s Alpha and hence, input orientation version of the DEA model named BCC hereafter, is employed in the study.

ii. HANDLING NEGATIVE DATA

Negative data has been handled by the translation invariance property inherited in the BCC model of the DEA technology. Translation Invariance means that the efficiency scores of the DMUs will be invariant to the translation of inputs and outputs by a scalar. The input oriented BCC model is translation invariant in the output. It states that the efficiency of a particular DMU will be unchanged to the translation of one or more of the outputs of all
DMUs by a scalar quantity. In other words, in an input oriented BCC model, if there are negative values in the output, then the efficiency of DMUs will not be affected by translating output values by a constant amount to remove the negatives. Figure 4.4 represents the graphical description of the translation invariance. For inefficient DMU a, the efficiency is $X_1^*/X_1$ before output translation. Translating this output by a value N, results in the efficiency of $NX_1^*/NX_1$, which simplifies to the original efficiency score as $X_1^*/X_1$.

![Figure 4.4: Input Oriented BCC Model Translation Invariance](image)

Simak et al. (1997) and Chehade (1998) handled the negative data while employing DEA by the invariance property [147] [39]. In the present study, the output variable SHARPE contains some negative values. These values have been converted into positive one by employing translation variance.

iii. MATHEMATICAL TREATMENT

Banks, Charnes and Cooper developed Variable Returns to Scale or BCC model, named after the first letters in their last name. The model considers n DMUs ($j = 1, 2, \ldots, n$), all having similar inputs and outputs, yet each consumes varying amounts of m different
inputs \((X_{ij}, X_{2j}, \ldots, X_{mj})\) to produce \(s\) different outputs \((Y_{1j}, Y_{2j}, \ldots, Y_{sj})\). Additionally, for \(DMU_j\), \(X\) represents the \(m \times n\) input matrix and \(Y\) represents the \(s \times n\) output matrix.

The linear programming for the envelopment form of the input oriented BCC model is as follows:

\[
\min \quad z_0 = \Theta - \epsilon_1 s - \epsilon_1 e \\
\Theta, \lambda, s, e \\
st \quad Y\lambda - s = Y_0 \\
\Theta X_0 - X\lambda - e = 0 \\
\rightarrow \quad 1 \lambda = 1 \\
\lambda, s, e \geq 0
\]

\(\Theta\), is the proportional reduction applied to the inputs of the DMU under evaluation, to improve its efficiency. This results in a radial movement towards the envelopment surface.

\(\epsilon\), is non-Archimedean, an infinitesimal constant in the primal objective function. It allows for a minimisation over \(\Theta\) to pre-empt the optimization involving input slacks, \(s\) and output slacks, \(e\). Optimization is accomplished in two steps. First, there is a reduction in inputs by the optimal \(\Theta^*\). Second, there is a movement to the frontier via the slack variables.

The vector \(\lambda\), indicates the relevancy of each efficient DMU in a particular reference set or peer group for the DMU under evaluation. The degree to which the inefficient DMU is compared to the efficient DMU on the frontier, used to construct the virtual DMU, explained earlier is represented by the magnitude of \(\lambda\).

The linear programming for the multiplier or dual form of the input oriented BCC model is as follows:

\[
\max \quad w_0 = \mu^T Y_0 + \mu_0 \\
\mu, \nu \\
st \quad \nu^T X_0 = 1 \\
\rightarrow \quad \mu^T Y - \nu^T X + u_0 \leq 0 \\
\rightarrow \quad - \mu^T \leq - \epsilon_1 \\
u_0 \quad \text{free}
\]

\[
\begin{align*}
\Theta, \lambda, s, e \\
\lambda, s, e \geq 0 \\
\epsilon_1 \geq 0 \\
\end{align*}
\]
ν and μ, are the vectors of input and output weights, called multipliers, representing the relative values of inputs and outputs assigned for each DMU to maximize its efficiency.

Variable u₀ is an indicator of scale efficiency. Negative u₀ indicates increasing return to scale, IRS and a positive one represents decreasing return to scale, DRS. Zero value of u₀ indicates that the DMU is operating at a constant return to scale, CRS.

To achieve the technical efficiency, the following conditions are satisfied:

- $\Theta^*_0 = 1$
- $w^*_0 = z^*_0 = 1$ (primal and dual objective functions)
- slack variables = 0

The linear programme is solved for all the DMUs under study and an efficiency score for each is produced. The efficient DMUs, satisfying the above criteria determine the envelopment surface. Projection of inefficient DMUs on to the efficient frontier or envelopment surface is achieved by reducing their inputs by a value $(1-\Theta)$ and accommodating for any of the slack variables.

iv. INPUT AND OUTPUT VARIABLES

The inputs and outputs for DEA model are defined in this section. Selecting the appropriate input and output variables is one of the important steps in DEA process. Inputs and outputs of the DEA are chosen as per the objective of the study. One of the objectives of present study is to analyse the performance of mutual funds. Therefore, the outputs are the benefits derived from the investment in the mutual fund schemes in terms of their performance.

Sharpe Ratio (SHARPE) and Jensen’s Alpha (ALPHA) is taken as the output variables for the study. Also, inputs are the resources expanded by the investors while investing in mutual funds. These are the charges as load fee (LOAD), expense ratio (EXPENSE), risk in terms of β (RISK) and minimum initial investment (MINII) needed by the investors. All these variables have been described in detail earlier in section 4.1.

Therefore, null hypothesis $H_0$ that the mutual fund schemes in India do not perform efficiently has been tested through DEA analysis.
4.6.1(b) LOGISTIC REGRESSION MODEL

Performance of mutual funds has been analysed in two stages. In first stage, performance of mutual funds in terms of their efficiency has been analysed through DEA technique. In the second stage, reason for the variation in efficiency of the mutual funds has been explored through logistic regression model. Through this model, relationship between mutual fund’s attributes and their efficiency has been analysed.

Logistic regression deals with the crucial limitation of linear regression as the later one cannot deal with dependent variables of dichotomous or categorical nature. That is in cases where dependent variable is of dichotomous nature as mutual fund schemes are efficient or not, logistic regression is used. It determines the impact of multiple independent variables presented simultaneously to predict membership of one or other of the two dependent variables categories. Since the dependent variable is dichotomous, we cannot predict a numerical value for it using logistic regression and hence the usual regression least squares deviations criteria for best fit approach of minimizing error around the line of best fit is inappropriate. Instead, logistic regression employs binomial probability theory inheriting only two values to predict. First, the probability (p) is 1 and second it is 0. That is, event or person belongs to one group rather than the other.

Logistic regression forms a best fitting equation or function using the maximum likelihood method that maximizes probability of classifying the observed data into the appropriate category given the regression coefficients. Like ordinary regression, logistic regression provides a coefficient ‘b’, which measures each independent variable’s partial contribution to variations in the dependent variable. To correctly predict the category of outcome for individual cases, a model i.e. an equation is created including all predictor variables (independent variable) useful in predicting the response variable (dependent variable).

Logistic regression works upon some assumptions as described below:

a. It does not assume a linear relationship between the dependent and independent variables.

b. The categories or groups must be mutually exclusive and exhaustive. That means, a case can only be in one group and every case must be a member of one of the groups.
c. The dependent variable must be a dichotomy i.e., with two categories.

d. The independent variables need not be interval, normally distributed, linearly related, or of equal variance within each group.

i  LOGISTIC REGRESSION EQUATION

Logistic regression gives each predictor (independent variable) a coefficient ‘b’ which measures its independent contribution to variations in the dependent variable with only two values, 0 or 1. Therefore, from a pool of independent variables and coefficients, numerical value of dependent variables are not predicted as in case of linear regression, but rather the probability, $p$ that it is 1 or 0 (belonging to one group rather than the other) is being found out. If $p$ is expressed as a linear function of investment, its value may be predicted as greater than 1 that cannot be true, as probabilities can only take values between 0 and 1.

Additionally, because logistic regression has only two Y values as ‘in the category or not in the category’, a straight line best fit as in linear regression is not possible to draw. Therefore, instead of using a least-squared deviations criterion for the best fit as in linear regression method, it uses a maximum likelihood method, which maximises the probability of getting the observed results given the fitted regression coefficients. Hence, logistic regression involves fitting an equation of the form to the data:

$$\text{Logit}(p) = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \latex
ii. **TERMINOLOGIES USED IN RESULTS**

a. **EXP (B):**

This is ln (odds ratio) and is given in the ‘Variables in the Equation’ table. It predicts the change in dependent variable due to change in independent variable and is calculated by using the regression coefficient of the independent variable as the exponent or \( \exp \).

b. **Likelihood Ratio Test:**

This tests the difference between -2LL for the full model with independent variable and -2LL for the initial chi-square in the full model. Here, LL is the log likelihood and is always negative. Moreover, log likelihood is the basis for tests of a logistic model.

iii. **VARIABLE SPECIFICATIONS AND EQUATION**

Two stage analyses provide explanation of the variations in the efficiency score of sample mutual fund schemes obtained in the DEA runs. According to Galagedera & Silvapulle (2002), efficiency of mutual fund schemes is affected by their attributes as operational characteristics i.e., experience, scale of operation and level of investors’ confidence [66]. In the present study, dependent variable is the efficiency score of all the mutual fund schemes. Independent variables are mutual fund’s operational characteristics as experience represented by their age (LAGE), scale of operation signified by their asset size (ASSETS) and level of investor confidence specified by asset ratio (ASSETR), their past performance (LSHARPE) and total risk (RISK). Past performance and risk of the mutual fund schemes has been measured by the Sharpe Ratio (from April, 2006 to March, 2011) and standard deviation respectively. All these variables have been defined earlier in section 4.1. Therefore, to meet the second objective of the research i.e., to study the performance of mutual funds with respect to different performance attributes, null hypotheses \( H_{0a}, H_{0b}, H_{0c}, H_{0d} \) and \( H_{0e} \) that age, asset size, asset ratio, past performance and risk of the mutual fund schemes is not related to their efficiency has been tested by employing Logistic Regression.
From the variables discussed above, Logistic Regression equation used for the analysis is given below:

\[
\text{Logit}(p) = a + b_1 \text{LAGE} + b_2 \text{ASSETS} + b_3 \text{ASSETR} + b_4 \text{LSHARPE} + b_5 \text{RISK} \ldots \ldots(viii)
\]

4.6.2 INVESTORS’ BEHAVIOUR

To analyze the primary data for studying the investors’ behaviour, statistical techniques as ANOVA and Factor Analysis have been employed. All these analysis have been performed with the help of SPSS. These techniques have been explained below in brief:

4.6.2 (a) ANALYSIS OF VARIANCE (ANOVA)

ANOVA - Analysis of Variance is used when the differences between the means of several populations or groups need to be investigated simultaneously. It is a generalization of Student's t-test which compares means of two groups. The one-way Analysis of Variance, ANOVA is used when there is one categorical independent variable and one continuous dependent variable. The independent variable can consist of any number of groups called factors. In the 2 group case, ANOVA becomes equivalent to a 2-tailed T test as the t-test, often used in similar experiments, is appropriate for situations where there are only two groups of one independent variable. Therefore, when there is a categorical independent variable and a continuous dependent variable and there are more than two groups/ population of the independent variable, then the appropriate analysis is the analysis of variance, ANOVA.

The null hypothesis that is tested with an ANOVA is that there is no significant difference among the means of different groups. Symbolically,

Null hypothesis, \(H_0: \mu_1 = \mu_2 = \mu_3 = \ldots \ldots = \mu_k\)

Alternate hypothesis, \(H_1: \) At least two means are different from each other

1, 2, 3, \ldots \ldots k are the k groups or population of independent variable. In the present study, there are 9 groups (k=9) of independent variable as investment options viz. fixed deposits, insurance, post office savings/NSC, gold/ e-gold, Bonds, PPF, real estate, mutual funds and shares.
\( \mu_1, \mu_2, \mu_3, \ldots \ldots \ldots \mu_k \) are the means of 1\textsuperscript{st}, 2\textsuperscript{nd}, 3\textsuperscript{rd}, \ldots \ldots \ldots k\textsuperscript{th} group respectively.

In order to accept the null hypothesis, all the means must be equal. Even if one mean is not equal to other, the null hypothesis will be rejected. A low p-value indicates that the null hypothesis should be rejected. The analysis is based on the assumption that the data in each group is drawn independently from a normal distribution, and that all group distributions share a common variance. In ANOVA with more than two groups of the independent variable the analysis goes through two steps. First, an F test is carried out to determine if there is any significant difference existing among any of the means. If this F score is statistically significant, then a second step analysis is carried out in which sets of two means are compared at a time in order to determine specifically, where the significance difference lies. In the first step of analysis i.e., F test, the following results are obtained:

**Between Groups Variance, \( \sigma^2_{\text{between}} \)**

Between Groups Variance, \( \sigma^2_{\text{between}} \) represents the explained variance or systematic variance. This variance is due to the independent variable i.e., the difference among k groups. For example, the difference between an investors’ perception for return in Fixed Deposits and mutual funds would represent between groups or explained variance.

**Within Groups Variance, \( \sigma^2_{\text{within}} \)**

The Within Groups variance represents the error variance. This variance is within the particular groups and not due to the independent variable. For example, the difference between one investor’s behaviour and another investor’s behaviour for return in Fixed Deposits would represent error variance.

**D.F. (Degree of Freedom)**

Since the variance between groups comes from many groups, therefore, the degree of freedom associated with variance between groups, \( \sigma^2_{\text{between}} \) is (k-1), where, k is the number of groups in independent variable. Since each variance in each group is associated with the size of the group, n therefore, the degree of freedom associated with variance within groups \( \sigma^2_{\text{within}} \) is k(n-1) where, n is the size of each group. In the present study, number of groups, k = 9 as the investment options and total number of investors, n = 440 (218 as MFI and 222 as NMFI). Therefore, degree of freedom is 3951.

118
**Sum of Squares (SS)**

SS represents the Sum of squared deviations from the mean. Further, SS\textsubscript{Within} captures variability within each group. If all group members had the same score, SS\textsubscript{Within} would equal 0. It is also called SS\textsubscript{Errors} or SS\textsubscript{Residual}, because it reflects variability that cannot be explained by group membership. SS\textsubscript{Between} captures variability between each group. If all groups had the same mean, SS\textsubscript{Between} would equal 0. For this, the term SS\textsubscript{Explained} is also used because it reflects variability that is explained by group membership. Also, SS\textsubscript{Total} = SS\textsubscript{Within} + SS\textsubscript{Between}.

**Mean Squares (MS)**

\[
\text{MS}_{\text{Within}} = \frac{SS_{\text{Within}}}{DF_{\text{Within}}} = \frac{SS_{\text{Within}}}{k(n-1)}
\]

\[
\text{MS}_{\text{Between}} = \frac{SS_{\text{Between}}}{DF_{\text{Between}}} = \frac{SS_{\text{Between}}}{(n-1)}
\]

**F Ratio**

R. A. Fisher developed the F-Test and determined that the difference in \(\sigma^2_{\text{between}}\) and \(\sigma^2_{\text{within}}\) could be expressed as a ratio to be designated as F-value, so that \(F = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}}\).

If the population means are exactly the same, then \(\sigma^2_{\text{between}}\) will be equal to the \(\sigma^2_{\text{within}}\) and the value of F will be equal to 1. If the population means are not equal then their sample means also vary greatly from one another, resulting in a larger value of \(\sigma^2_{\text{between}}\) and hence a larger value of F. \(\sigma^2_{\text{within}}\) is based only on sample variances and not on sample means and hence is not affected by the sample means. Therefore, a large value F indicates the rejection of null hypothesis and supports that there is significant difference between the means of sample.

**F Probability**

F probability indicates the probability of F ratio magnitude, and is denoted by p. A small value of p \(\leq 0.5\) indicates that the F ratio is significant. Therefore, if F calculated is greater than the F table and p value is well below the 0.5 then, the null hypothesis is rejected and it can be concluded that there is a significant difference between the means of groups.

But two very important questions remain. First, which means are significantly different from that of other means and, second what were the actual scores of the group? To answer these pair comparisons questions, Tukey's post-hoc test is performed. In other words, once it has been determined that differences exist among the means, post hoc range tests and
pair wise multiple comparisons can determine which means differ. Post hoc range tests identify *homogeneous subsets* of means that are not different from each other. Pair wise multiple comparisons test the difference between each pair of means and yield a matrix where asterisks indicate significantly different group means at an alpha level of 0.05.

4.6.2 (b) FACTOR ANALYSIS

Factor analysis attempts to identify underlying variables, or factors, that explain the pattern of correlations within a set of observed variables. Sometimes, the number of independent variables used in predicting the response variables is too many. This increases the computational time to get solution, time in data collection, expenditure in data collection and difficulty in making inferences. In such situations, factor analysis is used in data reduction to identify a small number of factors that explain most of the variance that is observed in a much larger number of manifest variables.

TERMINOLOGIES OF FACTOR ANALYSIS

The terminologies used in the study for the interpretation of factor analysis have been explained below:

*i. Correlation Coefficient Matrix:*

It is the matrix of correlation coefficients of the original observations between different pairs of input variables. In the present study, there are 11 input variables as the characteristics of mutual funds that investors look upon before investing in mutual fund schemes as mentioned below:

1. Past performance of mutual fund
2. Current NAV of mutual fund
3. Rating by a research agency/ Newspaper/ Magazine
4. Reputation of the mutual fund company
5. Mutual Fund manager
6. Portfolio of the scheme (% of investment in different co’s)
7. Exit load (fee charged at the time of selling of units)
8. Availability of tax benefits
9. Turnover of the mutual fund scheme (Sales during the period)
10. Asset size/ Total capital of the mutual fund scheme

11. Mutual Fund is Indian or Foreign

**ii. Factor Loadings, \( L_i(j) \):**

This matrix represents the correlation between different combinations of variables and factors. \( L_i(j) \) is the factor loading of the variable \( j \) on the factor \( i \). In the present study, \( j = 1, 2, 3, \ldots, 11 \) and five factors have been obtained in the result and therefore, \( i = 1, 2, 3, 4, 5 \).

**iii. Communality, \( h^2_i \):**

It is the sum of squares of the factor loadings of the variable \( i \) on all the factors as:

\[
h^2_i = \sum_{j=1}^{11} L_{ij}^2; \text{ in the present study, } j = 1, 2, \ldots, 11
\]

**iv. Eigen value:**

It is the sum of squares of the factor loadings of all variables on a factor. Eigen Value of the factor \( j = \sum_{i=1}^{n} L_{ij}^2 \)

If the Eigen value of a factor is more than or equal to one, then that factor is to be retained, otherwise that factor has to be dropped. If no factor is dropped, then the sum of the Eigen values of all factors is equal to the sum of communalities of all variables.

**v. Rotation:**

After obtaining factor loadings, it is to be examined whether the factor loading matrix possesses simple structure. If the factor loading matrix has a simple structure then, it is easy to make interpretations about the factors. A simple structure means that each factor consist of very high factor loading, as high as 1 on one of the factors and very low factor loadings, as low as 0 on other factors. If there is no simple structure, then the n-dimensional space of the factors has to be rotated such that factor loadings are revised to have a simple structure. This rotation simplifies the process of interpretation of the factors and is called the *rotation of factors*. The communalities of each variable before and after the factor rotation will be the same.

In the present study Varimax method of rotation of factors has been used. It is a popular method of rotation of factors and employs orthogonality between different pairs of factors axes. This means that even after the rotation, angles between different pairs of factors
axes are 90°. Also, in the present study, Principal Component Methods of the factor analysis has been used that maximizes the sum of squares of loadings of each identified factor. This is a popular technique that determines loadings of variables on different factors by using the standard normal values of the observations of the original / input variables.

Apart from the Analysis of Variance (ANOVA) and Factor Analysis, the Mean and Ranking methods have also been used in studying the perception of individual investors for mutual funds and other investment options.

Thus it can be concluded that ten attributes viz., LOAD, EXPENSE, MINII, RISK, SHARPE, ALPHA, LAGE, ASSETS, ASSETR and LSHARPE have been used to analyse the performance of mutual funds. For fulfilling this objective, six hypotheses have been formed out of which, one is for analysing the performance of mutual fund schemes in terms of their efficiency and the rest five have been formulated for exploring the relationship between mutual funds’ attributes and their performance. Performance of mutual fund schemes in terms of their efficiency has been analysed through DEA technique and the relationship between their attributes and efficiency score has been explored through logistic regression model. The study has been done over a period of six years from April, 2006 to March, 2012 on a sample of 119 mutual fund schemes. Investors’ behaviour has been analysed through two questionnaires, one for MFI and the other for NMFI. For analysing primary data ANOVA, Factor Analysis, mean and ranking methods have been used. Interpretation of the results of secondary as well as primary data have been provided in next chapter i.e., chapter 5.