

RESEARCH METHODOLOGY

Contents	3.1	<i>Introduction</i>
	3.2	<i>Research Design</i>
	3.3	<i>Data Analysis</i>

3.1 Introduction

The present study pertains to examining the influence of behavioural bias on the investment performance of equity mutual fund investors in Kerala. The first two objectives are to examine the long-run relationship between the stock market and equity mutual funds in India and to analyse the trend of the performance of equity mutual funds in India. The third objective of the study is to analyse the nature and extent of behavioural bias among equity mutual fund investors in Kerala with regard to different socio-economic variables and the fourth objective is to analyse the influence of behavioural bias among equity mutual fund investors on their investment performance. To fulfil the first and second objectives, the researcher has used secondary data obtained from the websites of AMFI, BSE and different asset management companies in India. Primary data was collected from equity mutual fund investors in Kerala for the third and fourth objectives, which was then analysed using various statistical tools. The following methodologies were used in the study:

3.2 Research Design

The study is both descriptive and analytical in nature. The study is descriptive in nature, as it describes the characteristics of the investors. The study is analytical because it developed the hypotheses and used various statistical tools to test them. The study is based on both secondary and primary data.

3.2.1 Source of Data

3.2.1.1 Secondary Data

The secondary data source for the study include journals, books, publications and websites of various mutual fund companies, websites of AMFI, NSE, BSE, SEBI, World Bank, IMF and government publications. The reports of CRISIL, RBI, MOSPI and various other wealth management organizations have also been considered.

3.2.1.2 Primary Data

Primary data for the study have been collected from individual equity mutual fund investors. The equity mutual fund investors in Kerala constitute the target population. Since the population is infinite, census survey is not possible. Hence, sample survey is used.

3.2.2 Sampling Design

Two sets of samples are required to fulfill the objectives of the study: sample equity mutual funds and sample equity mutual fund investors.

3.2.2.1 Selection of Equity Mutual Funds

As per SEBI guidelines, equity mutual funds are classified into various types based on their market capitalisation and investment strategy. In this study, equity mutual funds classified based on market capitalisation are considered.

The classification of equity mutual funds as per market capitalisation is as follows:

- i. Large-cap fund: At least 80% of its assets are invested in large-cap stocks.
- ii. Large and Mid-cap fund: At least 35% of its assets are invested in large-cap stocks and 35% in mid-cap stocks.
- iii. Mid-cap fund: At least 65% of its assets are invested in mid-cap stocks.
- iv. Small-cap fund: At least 65% of its assets are invested in small-cap stocks.
- v. Multi-cap fund: At least 65% of its assets are invested in equity and equity-related instruments.

Since the multi-cap funds invest in stocks of varying market capitalisation, only the first four categories of equity mutual funds are considered for the study.

The funds that have outperformed the benchmark for most of the years out of the past 11 years (2011-2021) have been taken as samples for the study. The following funds have been selected from each of these categories:

Large-cap fund: Canara Robeco Bluechip equity fund

Large and Mid-cap fund: Mirae Asset Emerging Bluechip fund

Mid-cap fund: UTI Mid-cap fund

Small-cap fund: Nippon India Small-cap fund

The daily net asset values of these funds and the Sensex for the period 1st January 2011 to 31st December 2021 have been considered for the study.

3.2.2.2 Selection of Equity Mutual Fund Investors

a. Population of Investors

The target population consists of individual equity mutual fund investors in Kerala. Due to the lack of an official database regarding mutual fund investors and their geographical distribution, the assistance of various banks and stock broking companies engaged in mutual fund investment, such as Geojit BNP Paribas, Motilal Oswal Financial Services, Tata Mutual Fund, Canara Bank and SBI has been sought. Furthermore, many investors are investing in direct funds using applications such as Upstox, Zerodha, Growapp and so on. Hence, it becomes difficult to identify the number of investors in Kerala. However, with the help of the banks and stock broking companies, the details regarding equity mutual fund investors were accumulated.

b. Determination of Sample size of Investors

The data relating to the number of investors and their geographical distribution is unavailable. Hence, the sample size of investors is calculated using

the following statistical equation. The highest standard deviation obtained from the pilot study was used in the equation.

$$n_0 = \left(\frac{zS}{e} \right)^2$$

n = number of sample size

z = standardized value corresponding to a confidence level (1.96 for 95% confidence level)

S = sample standard deviation (1.2903)

E = acceptable magnitude of error (assumed as 0.129)

$$n_0 = \left(\frac{1.96 * 1.2903}{0.129} \right)^2 = 19.60456^2 = 384.3387$$

Thus, the sample size of investors has been rounded off to 390.

c. Sampling Method

Multi-stage sampling method has been adopted for collecting primary data from investors in Kerala. In the first stage, the districts in Kerala have been classified into three groups based on the number of branches of mutual fund AMC's present. Mutual fund AMC's have their presence in 11 districts in Kerala, which are Thiruvananthapuram, Kollam, Alappuzha, Pathanamthitta, Kottayam, Ernakulam, Thrissur, Palakkad, Malappuram, Kozhikode and Kannur. The districts with the most number of branches of AMC's are Ernakulam, Thiruvananthapuram, Kozhikode and Thrissur. The districts with moderate number of branches are Kottayam, Kannur and Palakkad. The districts with the least number of branches are Pathanamthitta, Kollam, Malappuram and Alappuzha.

In the second stage, one district has been selected from each of these groups as sample districts using simple random sampling method. Kozhikode was selected from the group of districts with the highest number of branches, Kollam

from the group with the least number of branches and Kannur from the group with moderate number of branches.

In the third stage, 130 investors were selected from each of these districts. For identifying investors in these regions, the assistance of various banks and stock broking firms engaged in mutual fund investment such as Geojit BNP Paribas, Motilal Oswal Financial Services, Tata mutual fund, Canara Bank and SBI has been sought.

3.2.3 Research Instrument

The researcher has used structured questionnaire as the instrument for collecting the primary data from the sample investors. Initially, a pilot study was conducted among 50 equity mutual fund investors in the Ernakulam district. Several experts in the fields of finance, behavioural finance and research have been consulted and their suggestions were incorporated while preparing the questionnaire to ensure the content validity of the instrument. Based on the pilot study, some of the questions were refined and the questionnaire was finalized. The researcher personally met some of the investors and gave them the questionnaires and mailed them to the rest of them to get the questionnaires filled. The period of the actual survey was from August 2020 to December 2021.

The questionnaire consisted of three parts which are as follows:

Part 1: Questions related to the socio-economic profile of the respondents

Part 2: Questions for collecting the responses related to behavioural bias

Part 3: Questions for collecting the responses related to investment performance

3.2.4 Variables used for the study

The present study examines the nature and extent of behavioural bias with regard to different socio-economic factors and the influence of behavioural bias on the investment performance of equity mutual fund investors in Kerala. For this purpose, the following variables are used.

Table 3.1
Variables Used for the Study

Sl. No.	Nature of the Variable	Name of the Variables
1	Socio-economic Variables	1. Gender 2. Age 3. Education level 4. Occupation 5. Marital status 6. Annual Income 7. Investment Experience
2.	Behavioural bias	A. Belief Perseverance Bias 1. Representativeness Bias 2. Confirmation Bias 3. Cognitive Dissonance bias 4. Illusion of Control Bias B. Information Processing Bias 5. Anchoring bias 6. Availability bias 7. Self Attribution bias 8. Mental Accounting bias C. Emotional Bias 9. Overconfidence bias 10. Loss Aversion bias 11. Regret Aversion bias 12. Herding bias
3.	Investment Performance	Investment Performance

3.2.5 Reliability Analysis

In the present study, reliability of the measurement scales is tested by using Cronbach's Alpha Reliability Co-efficient. The results of the reliability analysis are presented in table 3.2.

Table 3.2
Reliability Analysis

Sl. No.	Variables	Number of Items	Alpha Value
Behavioural Bias			
1	Representativeness Bias	4	0.807
2	Confirmation Bias	4	0.843
3	Cognitive Dissonance Bias	2	0.810
4	Illusion of Control Bias	3	0.754
5	Anchoring Bias	5	0.844
6	Availability Bias	5	0.786
7	Self Attribution Bias	3	0.787
8	Mental Accounting Bias	2	0.756
9	Overconfidence Bias	6	0.892
10	Loss Aversion Bias	4	0.777
11	Regret Aversion Bias	3	0.778
12	Herding Bias	5	0.851
Investment Performance		3	0.863

Source: Survey Data

Since all the values of Cronbach's alpha are above 0.7, it can be inferred that the scale is reliable in terms of internal consistency (Nunnally, 1967).

3.2.6 Normality Analysis

Normality test determines whether the data set is well-modeled by a normal distribution or not. One sample K-S test is commonly used to examine the normality of data.

Table 3.3
Kolmogorov-Smirnov Test of Normality

Sl. No.	Variables	N	Mean	Std. Deviation	p-value
Behavioural Bias					
1	Representativeness Bias	390	14.37	3.31	.000
2	Confirmation Bias	390	14.56	3.29	.000
3	Cognitive Dissonance Bias	390	7.04	1.80	.000
4	Illusion of Control Bias	390	10.27	2.41	.000
5	Anchoring Bias	390	16.00	4.06	.005
6	Availability Bias	390	17.67	3.67	.000
7	Self Attribution Bias	390	9.92	2.01	.000
8	Mental Accounting Bias	390	7.55	1.55	.000
9	Overconfidence Bias	390	21.21	4.67	.000
10	Loss Aversion Bias	390	13.75	2.86	.008
11	Regret Aversion Bias	390	9.84	2.40	.000
12	Herding Bias	390	16.07	4.12	.001
Investment Performance		390	11.68	2.43	.000

Source: Survey Data

The results in table 3.3 revealed that none of the data set is normally distributed. Hence, the researcher has employed Skewness and Kurtosis tests for checking the normality of the data. Skewness relates to the symmetry or asymmetry of a distribution while kurtosis relates to the peakedness of the distribution. Chou & Bentler (1995) suggested that the data will be normal when the values of skewness fall between -3 and +3 and the values of kurtosis fall between -10 and +10. The studies by Black, Hiar, Babin and Anderson (2006) suggested that the data become normal when the values of skewness and kurtosis are in the range of ± 2.58 and ± 1.98 . The results of skewness and kurtosis tests are presented in table 3.4.

Table 3.4
Skewness and Kurtosis Results

Sl. No.	Variables	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Behavioural Bias					
1	Representativeness Bias	-.448	.124	-.476	.247
2	Confirmation Bias	.071	.124	-.733	.247
3	Cognitive Dissonance Bias	.155	.124	-.010	.247
4	Illusion of Control Bias	-.003	.124	-.180	.247
5	Anchoring Bias	-.422	.124	.125	.247
6	Availability Bias	.151	.124	.822	.247
7	Self Attribution Bias	-.133	.124	-.949	.247
8	Mental Accounting Bias	.147	.124	-.408	.247
9	Overconfidence Bias	.246	.124	-.451	.247
10	Loss Aversion Bias	-.280	.124	-.059	.247
11	Regret Aversion Bias	-.111	.124	-.473	.247
12	Herding Bias	-.492	.124	.315	.247
Investment Performance		-.661	.124	.171	.247

Source: Survey Data

The results of table 3.4 implied that the skewness and kurtosis values of all the variables fall within the acceptable ranges. Hence, normality can be assumed and parametric tests can be used for analysis in the study.

3.3 Data Analysis

3.3.1 Secondary Data Analysis

1. Unit Root Tests

In statistics, unit root tests are used to check the stationarity of the time series data. In the present study, the Augmented Dickey Fuller testis employed to examine whether the data are non-stationary and possess a unit root. Data are said to be stationary when their statistical properties such as mean, variance, etc. are constant over time.

2. Vector Auto Regression (VAR)

VAR is a widely accepted method adopted to determine the optimum lag length of each variable. Various criteria which are used for the selection of optimum lag length are the Likelihood Ratio, Akaike Information Criteria, Final Prediction Error, Hannan-Quinn Information Criteria and Schwarz Information Criteria.

3. Co-Integration tests

Co-integration is a statistical method used to find a possible correlation between time series processes in the long run. Johansen Co-integration test is used to determine the number of co-integrating equations among the variables.

4. Vector Error Correction Models (VECM)

VECM is a co-integrated VAR model. It is used to examine the short-run dynamics and long-run equilibrium of the model.

5. Causality tests

If there is co-integration among the variables, the exogeneity tests are applied based on VECM. Causality refers to the ability of one variable to predict and influence the value of another variable. It reveals which variable is endogenous and which one is exogenous. Engle & Granger (1987) found that a causal relationship exists among variables in one direction if they are co-integrated. In this study, the Granger causality test is applied to examine the causality between the variables.

6. Variance Decomposition Analysis

Variance decomposition is used to assess the proportion of the movement of dependent variables due to their own shock and due to the shock of other independent variables. In this study, Cholesky decomposition is used to obtain variance decomposition.

7. Impulse Response Function

The impulse response function is used to determine the direction, magnitude and duration of the variables in the system which is affected by an external variable's shock. Cholesky decomposition is used to obtain the impulse response of linkages.

8. Auto Regressive Integrated Moving Average (ARIMA)

ARIMA is a statistical analysis model in which time series data have been used to understand the data and to predict the future trends. ARIMA forecasts future values based on past values assuming that the future will resemble the past.

3.3.2 Primary Data Analysis

1. Mean

Mean is a measure which represents the entire data by a single value. It is the average value of the given set of data.

2. Standard Deviation

Standard deviation measures the deviation of values from the mean value. It is the degree of dispersion relative to its mean.

3. Independent Sample t Test

The Independent Sample t-test is used for comparing the means of two independent groups to examine whether significant difference exists between them. Levene's test is used to assess the equality of variances of the group of variables. It tests the null hypothesis that the variance of the group is homogeneous. If the p-value of Levene's test is less than 0.05, the variance is said to be heterogeneous (Garson, 2012). In such cases, the second set of analysis which is 'equal variance not assumed' has to be considered.

4. One-way Analysis of Variance/ Welch F

The One-way ANOVA is used to examine whether significant difference exists among the means of three or more independent groups. The assumption of homogeneity of variance is tested using Levene's test. The null hypothesis is that the variance of the group is homogeneous. If the p-value of Levene's test is less than 0.05, the variance is said to be heterogeneous (Garson, 2012). In such cases, the values of the Welch F test are considered instead of ANOVA.

5. Tukey HSD/Tamhane's T2 Post Hoc Test for Multiple Comparisons

When there exists a significant difference among the independent groups using ANOVA, post hoc tests are employed to examine the exact difference between the groups. Post hoc tests such as the Tukey HSD test and Tamhane's T2 test are widely used. The Tukey HSD test is used when equal variances are assumed and Tamhane's T2 is used when equal variances are not assumed.

6. Multiple Regression Analysis

Multiple regression is a statistical test employed to analyse the relationship between a single dependent variable and many independent variables (Hair, Black, & Anderson, 2015).