

Chapter 3

CHAPTER 3

RESEARCH METHODOLOGY

This chapter deals with the nature of the research, sample design that includes the population, sample size and sampling technique, methods of data collection, time period of the study, pilot research, statistical tools and techniques used in this research. Also, the hypotheses framed based on the objectives have also been listed.

3.1 NATURE OF THE RESEARCH

The study is descriptive and analytical in nature. Dulock, H. L. (1993). Descriptive research is used to systematically and accurately describe the facts and characteristics of a given population or area of interest whereas analytical research attempts to explain complex phenomena. Analytical research is mainly concerned with testing hypotheses, specifying and interpreting relationships, by analysing the facts or information that is already available.

This study aims at describing the demographic and trading profile of the equity derivative traders in Bengaluru city and then analyse the factors that influence trading success among the traders and also the impact of these factors on successful trading.

3.2 SAMPLE DESIGN

The sample design includes the target population and the profile of the place of research, sample size and the technique used to approach the respondents.

3.2.1 Target Population

The target population of the study is the active equity derivative traders in Bengaluru city of Karnataka State in India. The city is attributed to its highly educated workforce with investible surplus making it a suitable destination for conducting the survey. There is no published information about the number of active equity derivative traders.

Kamath N. (July, 2020) The founder and CEO of Zerodha, the no. 1 Discount Broking Firm in India in a tweet guesstimated that the No. of total daily active retail investors in the F&O segment will be around 5 to 7 lakhs in India. SEBI investors survey (2015, p. 113) states that 7.2% of the investor population resides in the south zone of India.

As Bengaluru is one of the major cities in South India, it is assumed that 40% of the south Indian investors would be from Bengaluru. Therefore, the population of the study can be roughly estimated to be around 20,160 active equity F&O traders.

3.2.2 Profile of Bengaluru City

Bengaluru, the third most populated city in India is known by various names, including: The Silicon Valley of India as it is the nation's prominent IT exporter, The Garden City of India because of its abundance of greenery, The Aerospace and machine tools hub of India, the knowledge hub of India and also used to be the Pensioner's Paradise.

According to the Official Website of the International Trade Administration, Bengaluru has witnessed a drastic technology explosion since the early 1980's when the tech giant Texas Instruments discovered its potential as a high-tech city. It is now home to major IT companies like Wipro, Infosys, TCS, Cognizant, Accenture, Google, Microsoft, HP, IBM, Cisco, Intel, etc...Most of these companies have their research and development centres along with their corporate offices in the city. Progressively, the city has also seen a huge growth in biotech, healthcare and infrastructure industries. The city has about 200 biotech firms making it the major biotechnology group in India. In addition, Bengaluru boasts to have the major public sector companies, such as Hindustan Aeronautics Limited (HAL), Bharat Heavy Electricals Limited (BHEL), Bharat Electronics Limited (BEL) and HMT Limited. Bengaluru is also home to some of India's finest scientific research institutions, such as Indian Space Research Organization (ISRO), Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), National Aerospace Laboratories (NAL), Central Manufacturing Technology Institute (CMTI), National Centre for Biological Sciences (NCBS) and a couple of national defence research laboratories.

To take advantage of the huge talent pool available in the city, in the past 20 years, numerous private sector and global companies set up their base in Bengaluru. The clustering of IT, biotech, aerospace, defence and electronics-related industries in Bengaluru and the large public sector enterprises attracted people from all parts of India, making it very cosmopolitan in character and culture, very different from most other Indian cities.

The biggest discount stock broking firm Zerodha is also headquartered in Bengaluru. The city has the most highly educated workforces in the world who are tech savvy with investible surplus. This makes it an appropriate destination for conducting the research.

3.2.3 Sample size

Krejcie & Morgan (1970) the sample size should be at least 370 if the population is more than 10,000. For this study, data was collected from 420 active equity derivative traders in Bengaluru city through questionnaire using online google forms and offline hard copies. 38 responses were rejected because of insufficient information. Finally, the analysis was carried out using the data disclosed by 382 respondents.

3.2.4 Sampling technique

Exponential non-discriminative Snowball sampling method is used in the study. Snowball sampling or chain-referral sampling is a non-probability sampling technique. It's called so because (in theory) once the ball gets rolling, it picks up further "snow" on its way and turns out to be bigger and more. For this research study, the initial respondents were requested to provide multiple referrals to identify potential subjects required and each referral gave some more and the process continued until necessary sample size was reached. As the majority of the traders operate online and hardly any turn-up to the broking firms, it was difficult to locate the subjects and therefore the snowball sampling technique came handy.

3.3 DATA COLLECTION

Primary data was collected using a self – administered questionnaire. The questionnaire was prepared based on the extensive literature review and opinion from experts in the field of equity derivatives trading. The profiles of the experts and the copy of the questionnaire are enclosed in the annexures.

The questionnaire consists of 57 questions, out of which the first 14 questions are categorical regarding the demographic details of the respondents and certain basic questions to understand the profile of the trader like their no. of years of trading experience, whether they are full time or swing trader, etc. The rest are ordinal 5-point Likert scale questions covering all the independent factors such as personal habits and

lifestyle, trading behaviour, knowledge, risk management skills, emotional factor, technology and social factor. The last part includes questions related to variables measuring the dependent factor, i.e.: trading success. The study also uses Secondary data gathered from the classic books related to the field of research such as market wizards, Reminiscences of a Stock Operator, etc. Newspapers such as Economic times, Hindu Business Line, etc. Research journals such as Journal of finance, International journal of financial and quantitative analysis, etc. websites such as <https://www.sebi.gov.in/>, <https://www.nseindia.com/>, etc.

3.4 PERIOD OF THE STUDY

The primary data used in the study was collected between August 2019 and February 2020.

3.5 PILOT STUDY

Pilot study was conducted among 50 respondents using self- administered questionnaire to test the validity and reliability of the data collection instrument (Questionnaire). Based on the results of the study and opinion from experts, few changes were made in the questionnaire to suit the present study.

3.6 TECHNIQUES AND TOOLS USED FOR DATA ANALYSIS

Objective 1 – Percentage analysis is used to understand the demographic and trading profile of equity derivative traders in Bengaluru city.

A percentage frequency distribution is the presentation of data that lays down the percentage of observations that occur for each data point or grouping of data points. Lavrakas, P. J. (2008) It is a predominantly convenient method of conveying the relative frequency of survey responses and further data. Several times, percentage frequency distributions are presented as tables or bar graphs or pie charts. The method of calculating the percentage frequency distribution includes first recognizing the entire number of observations to be represented; then counting the total number of observations within each data point or grouping of data points; and then dividing the number of observations within each data point or grouping of data points by the total number of observations and multiplying by 100.

Objective 2 - Cross Tabulation, Chi-square analysis and Cramer's V is used to explore the relationship between demographic and trading profile of the traders and their trading success.

(Guide to cross-tabulation analysis, qualtrics.com) Cross-tabulation is one of the most valuable analytical tools and a backbone of the market research business. Qualtrics presumes that cross-tabulation and single variable frequency analysis accounts for more than 90% of all research analyses. Cross-tabulation analysis, also known as contingency table analysis, is quite normally used to examine categorical (nominal) data. Cross-tabulation is a two- (or more) dimensional table that records the number (frequency) of respondents that have the particular characteristics described in the cells of the table. A Cross-tabulation table is used to study relationships within data that may not be readily obvious. It is usually used in statistical analysis to find patterns, trends, and probabilities within raw data.

The Pearson chi-square test essentially tells us whether the results of a cross-tabulation are statistically significant. That is, are the two categorical variables independent (unrelated) of one another. Cross tabulation presents the distributions of two categorical variables simultaneously, with the intersections of the categories of the variables appearing in the cells of the table. The Test of Independence assesses whether an association exists between the two variables by comparing the observed pattern of responses in the cells to the pattern that would be expected if the variables were truly independent of each other. Calculating the Chi-Square statistic and comparing it against a critical value from the Chi-Square distribution allows the researcher to assess whether the observed cell counts are significantly different from the expected cell counts.

McHugh, M. L. (2013) The Chi-square is a significance statistic, and should be followed with a strength statistic. Cramer's V is the most common strength test used to test the data when a significant Chi-square result has been obtained. Advantages of the Chi-square include its robustness with respect to distribution of the data, its ease of computation, the detailed information that can be derived from the test. One of the major limitations is the tendency of the Cramer's V to produce relatively low correlation measures, even for highly significant results.

The calculation of the Chi-Square statistic is quite straight-forward and intuitive:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Where f_o = the observed frequency (the observed counts in the cells)

f_e = the expected frequency if NO relationship existed between the variables

As depicted in the formula, the Chi-Square statistic is based on the difference between what is actually observed in the data and what would be expected if there was truly no relationship between the variables.

Chi-square says that there is a significant relationship between variables, but it does not say just how significant and important this is.

Cramer's V is a way of calculating correlation in tables which have more than 2x2 rows and columns. It is used as post-test to determine strengths of association after chi-square has determined significance.

V is calculated by first calculating chi-square, then using the following calculation:

$$V = \text{SQRT} (\chi^2 / (n (k - 1)))$$

Where χ^2 is chi-square and k are the number of rows or columns in the table. Cramer's V varies between 0 and 1. Close to 0 it shows little association between variables. Close to 1, it indicates a strong association.

Objective 3 - Principal component analysis (PCA) is used to study the various factors that influence trading success among equity derivative traders.

Principal component analysis is a technique that uses sophisticated mathematical models to reduce a large set of possibly correlated variables into few principal components. Shlens, J. (2014) "PCA is the mainstay of modern data analysis. The goal of PCA is to identify the most meaningful basis to re-express a data set. The hope is that this new basis will filter out the noise and reveal hidden structures".

Abdi, H., & Williams, L. J. (2010) PCA examines a data table representing observations described by a number of dependent variables, which are, in general, inter-correlated. Its aim is to extract the vital information from the data table and express it as a set of new orthogonal variables called principal components. PCA also represents the pattern of similarity of the observations and the variables by displaying them as points in maps. It is a technique for bringing down the dimensionality of datasets, improving interpretability but at the same time reducing information loss. It does so by producing new uncorrelated variables that successively maximize variance.

Dimensionality: It is the number of random variables in a dataset or simply the number of features, or rather more simply, the number of columns present in the dataset.

Correlation: It shows how strongly two variables are related to each other. The value ranges between -1 and +1. Positive value indicates that when one variable increases, the other increases as well, while negative value indicates the other decreases on increasing the former.

Objective 4 – PLS SEM (Partial Least Square Structural Equation Modelling) is used to test theoretical model from a prediction perspective

Hair, J. F., et. al. (2011) “SEM has become a quasi-standard in management research when it comes to analysing the cause-effect relations between latent constructs.” SEM is the methodology for demonstrating, estimating and evaluating a network of complex cause-effect relationship models with latent variables. Hair et al. (2019) PLS SEM is used when the analysis is about testing a theoretical framework from a prediction standpoint.

The PLS SEM consists of two sub-models: the measurement model and structural model. The measurement model represents the relationships between the latent variables and the observed data. The structural model represents the relationships amongst the latent variables.

An iterative algorithm solves the structural equation model by estimating the latent variables by using the measurement and structural model in alternating steps, hence the procedure's name, partial. The measurement model estimates the latent variables as a

weighted sum of its manifest variables. The structural model estimates the latent variables by means of simple or multiple linear regressions between the latent variables estimated by the measurement model. This algorithm repeats itself until convergence is achieved.

3.7 HYPOTHESIS TO BE TESTED

- The demographic and trading profiles of the equity derivative traders have no association to their trading success.
- There is no significant influence by factors such as personal habits, trading behaviour, knowledge, emotional intelligence, risk management skills, technology and social factors on the trading success of the participants in the market.

The above research hypotheses are converted into statistical hypotheses and are discussed in detail in analysis and interpretation chapter.