Chapter 3 Research Methodology

Research methodology provides logical and systematic sequence of various research steps and methods which are based on scientifically proved procedures in order to achieve the objectives of research. Methodology ensures objectivity and consistency in procedures thus provides more reliable results. Many research works have been published on different aspects of financial inclusion. In present research, combination of primary survey, analysis of secondary data and development of an index have been employed to study various dimensions of financial inclusion among rural households in Haryana. In the primary survey, 525 households in 21 villages spread over 4 districts were studied through structured interview schedule. Secondary data have been collected from various sources such as Reserve Bank of India (RBI) annual reports and publications, NABARD annual reports, Financial Access Survey 2014, government reports and other committees' reports on financial inclusion etc. regarding the status of financial inclusion in developing countries, developed countries and state-wise in India.

3.1 Research Design of the Study

The nature of present study is descriptive as well as empirical research, whereas researcher attempts to find out the current status of financial inclusion among rural households in Haryana. The field survey is conducted in to order to know the level of financial literacy and reasons of financial exclusion of rural households in Haryana. The present research work is based on primary as well as secondary data. A structured plan has been designed for sample area, sample size, methods of data collection, methods of data interpretation and techniques of empirical analysis. Sample survey is conducted in

order to fulfill the objectives of this research. The 'structure interview schedule' statistical technique of data collection has been used to conduct a field survey in rural areas of Haryana. Primary data have been collected from rural households of selected four districts of Haryana namely Mahendergarh, Fatehabad, Karnal and Gurgaon. Appropriate statistical and econometric tools have been employed to measure financial inclusion and to analyze the impact of socio-economic variables on financial inclusion.

3.1.1 Data Description

In order to achieve the research objectives of the study, the primary and secondary data have been used for empirical analysis. The primary data have been collected by filling structured interview schedules containing the closed ended as well as open ended questions. Open ended questions were asked to develop understanding about various issues and such findings were again discussed with experts to draw meaningful inferences. Secondary data have been obtained from the primary sources and secondary sources such as Handbook of Statistics on Indian Economy, Basic Statistical Returns of Schedule Commercial Banks 2013, Reserve Bank of India (RBI) bulletins, RBI annual reports, NABARD annual reports, Insurance Regulatory and Development Authority (IRDA) annual reports, Report of the Committee on Financial Inclusion 2008, Report of the Committee on Financial Sector Reform 2009, Report of the Technical Committee on Mobile Banking 2014, UNDP reports and Research work, Financial Access Survey (IMF) 2014, Global Findex Database, Financial Inclusion 2020 A Global Forum, Report of Committee on Comprehensive Financial Services for Small Business and Low Income Households 2013, Haryana Statistical Abstracts 2013-14, Economic Survey of India 2013-14, Ministry of Commerce and Industry, Ministry of Statistics and Programme

Implementation (MOSPI) publications, 11th Five Year Plan (2007-12), and 12th Five Year Plan (2012-17) etc. Secondary data have been obtained to present global, national and regional scenario with respect to financial inclusion. The composite Index of financial inclusion (IFI) has been constructed on the basis secondary data on number of bank account, number of bank branches and deposit/credit at a point of time (March, 2013).

An 'unstructured interview' method has also been used to obtain data from Gram Panchs/Panchayats and employees of financial institutions - scheduled commercial banks (SCBs), Primary Agriculture Cooperative Society (PACS), Microfinance Institutions (MFIs), Post Office Saving Banks (POSBs) etc. The employees of formal financial system have been interviewed on some issues regarding financial inclusion such as financial literacy, usage of technology in banking sector, reasons for not availing credit from formal financial system, performance of formal financial system etc. By conducting interview, the researcher has come to know the nuances of financial inclusion among rural households in Haryana.

3.1.2 Sampling Design

The state of Haryana has wide physical area having 21 districts with large number of population of 2.54 crore (Census, 2011). Therefore, it is not possible for researcher to investigate the census population. Thus, planning of sampling design is very important for the field survey. The sample area is selected on the basis of multi-stage sampling. Four districts have been selected out of 21 districts on the basis of purposive sampling (one from North, one from South, one from East and one from West). Every district has around five to six blocks, whereas researcher has selected one village from each block of

the district. Therefore, 21 villages have been selected for the primary data collection though structured interview schedule method. The sample size of 25 respondent households has been taken from each village, whereas one village has been chosen from each block of districts. Therefore, total size of sample is 525 respondent households from rural areas of Haryana. The sample of rural households is segmented in three strata comprising General caste households, backward caste households and schedule caste households. On the basis of strata sampling, 9, 8 and 8 samples have been taken from general caste (Gen.) households strata, schedule caste (SC) households strata and other backward caste (OBC) households strata.

Table 3.1 District wise Sample Size

Segment	Mahendergarh	Fatehabad	Karnal	Gurgaon
General	45	54	54	36
OBC	40	48	48	32
SC	40	48	48	32
TOTAL	125	150	150	100

Source: Field Survey

3.1.3 Study Area

The primary survey based study is designed to conduct a research on rural households of Haryana, but rural area has too wide range. Haryana state has more than 6000 villages. Therefore, sample area has been designed to collect the primary data from rural households in Haryana. Four districts have been selected to represent the state of Haryana (one from north, one from south, one from east, and one from west Haryana). One village

has been selected from each block of the districts; therefore, 21 villages have been selected for conducting field survey. The district wise detail of blocks and village has been depicted in table 3.2.

Table 3.2 Detail of Study Area for Sample Survey

Name of District	Name of Block	Name of Village
Mahendergarh	Narnaul	Meharampur
	Mahendergarh	Majra Kalan
	Kanina	Pota
	Nangal Sirohi	Meghot Halla
	Ateli	Bhori (Gujarwas)
Fatehabad	Fatehabad	Buthan Kalan
	Bhuna	Dhani Dulat
	Jakhal	Baijjalpur
	Tohana	Dhani Bhojraj
	Ratia	Hasinga
	Bhattu Kalan	Dhingsara
Karnal	Karnal	Kachhwa
	Indri	Uchana
	Assandh	Salwan
	Gharaunda	Shekhpura
	Nilokheri	Sagga
	Nissing	Chirao
Gurgaon	Farukhnagar	Joniawas
	Gurgaon	Dhankot
	Pataudi	Ransika
	Sohna	Badshahpur Tethar

Source: Designed by Researcher for conducting field survey

3.1.4 Definitions of Rural Area and Rural Household

The 'Rural Area' and 'Rural Household' are key concepts in this study. Thus, precisely defining these terms is very important. In this research, these definitions have been taken from Census 2011.

3.1.4.1 Defining the Rural Area as per Census 2011

Rural area may be defined as an area which is having at least one feature out of below explained three features:

1. Area should not city and municipality.

2. Area should not have population more than 5000.

3. More than 75 percent of population should be engaged in agriculture and allied agriculture activity.

3.1.4.2 Definition of Rural Household as per Census 2011

"A household is usually a group of persons who normally live together and take their meal from a common kitchen". Rural household is a group of persons who are living in rural area, whether few members of household may go daily to nearby city for job or earning purpose. Person in household may be related or unrelated or mix of both. However, if a group of unrelated persons live in a census house but do not take their meals from the common kitchen, then they are not constituent of a common household. Each such person should be treated as a separate household.

3.2 Research Methods

According to the nature of data, an appropriate research methodology is employed in order to achieve objectives of the study. Both primary and secondary data have been collected for the empirical analysis of the study. The index of financial inclusion (IFI) has been constructed to measure the district wise financial inclusion in rural areas of Haryana.

3.2.1 Index of Financial Inclusion (IFI)

The extent of financial inclusion can be measured by using several indicators such number of bank accounts (per 1000 population), number of bank branches (per 100,000 population), amount of bank credit and deposit, geographical ATM penetration, demographic ATM penetration, loan income ratio, deposit income ratio, cash deposit ratio, credit and deposit accounts, financial market size and depth, etc. (Sarma, 2008), (Chattopadhyay, 2011), (Sarma, 2012), (CRICIL Inclusix, 2013), (Global Findex, 2012).

Table 3.3 Indicators to Measure Financial Inclusion

Sr. No.	Indicators	Measurement
1	Bank Accounts	Number of Bank Accounts per 1000 population
2	Deposit Penetration	As percentage of GDP
3	Credit Penetration	As percentage of GDP
4	Demographic Branch Penetration	Number of Bank Branches per 100,000 population
5	Geographic Branch Penetration	Number of Bank Branches per 1000 km^2
6	Demographic ATM Penetration	Number of ATM per 100,000 population

7	Geographic ATM Penetration	Number of ATM per 1000 km^2
8	Demographic Loan Penetration	Number of Loans per 100000 population
9	Loan Income Ratio	Average size of loan to GDP per capita
10	Deposit Income Ratio	Average size of deposit to GDP per capita
11	Cash Deposit Ratio	Cash in circulation to total bank deposits

Source: (Chattopadhyay, 2011), (Sarma, 2008), (CRICIL Inclusix, 2013).

All above mentioned indicators provide important and useful information on inclusiveness of the financial system of an economy. If single indicator is used to measure the financial inclusion, it will present only partial information on the inclusiveness of the financial system. The use of individual indicator may lead to misinterpretation of the extent of financial inclusion. Therefore, more than one indicator should be used to measure the inclusiveness of financial system. By reviewing the literature, it is found that most important indicators to measure the extent of financial inclusion are banking penetration, availability of bank branches and usage in term of deposit and credit.

Thus, a comprehensive measure incorporates several dimensions of financial inclusion by aggregating into a single number. Comprehensive measure can be used to measure and compare the levels of financial inclusion across economies, across states and districts, which can be used to evaluate the performance of policy measures aimed at financial inclusion over a period of time.

In the present index, three indicators of financial inclusion have been used to measure the extent of financial inclusion in rural area of Haryana. These indicators are Banking Penetration, Availability of Bank Branches, and Usage of deposit and credit. These indicators have been described as follow:

I. **Banking Penetration**: Banking penetration means number of bank account in respect to number of population. The size of the banked people is a measure of the banking penetration of the inclusive system. It means if every person has a bank account, then the value of this measure would be one. However, the number of bank account per 1000 population is used to measure the index of banking penetration in the present index of financial inclusion.

II. **Availability of Banking Services:** The second important indicator of financial inclusion is availability of banking services. The availability of banking services can be measured by the number of bank branches (per 1,00,000 population) or numbers of ATMs per 1,00,000 population. The number of banking personnel can also be used as another indicator of the availability of banking services. In the present index, the number of bank branches per 1,00,000 population is used as an indicator of availability of banking services to measure the extent of district wise financial inclusion in rural areas of Haryana.

III. **Usage:** third important indicator to measure the extent of financial inclusion in rural areas Haryana is usage of banking services in term of deposit and credit. In constructing the usage dimension index, two basic important banking services i.e. credit and deposit in bank accounts have been taken for this study. The amount of sum of rural deposit and rural credit (per 100000 population) has been used as an indicator of usage of banking system in present index.

The observed maximum and minimum figures in particular dimensions have been used as maximum and minimum values in order to compute dimensional index (Kodan, 2013).

Based on eminent research studies, level of financial inclusion has been studied based on following criteria (Sarma, 2008):

Table 3.4 Level of Financial Inclusion

Index Score	Level of Financial Inclusion
Above 0.5	High
Between 0.3 to 0.5	Medium
Below 0.3	Low

This index of financial inclusion takes values between 0 to 1. Zero indicates lowest financial inclusion means completely financial exclusion and one indicates highest financial inclusion that shows complete financial inclusion.

3.2.1.1 Methodology of Index

In this research work, multidimensional approach is adopted for constructing the composite index of financial inclusion. The methodology of computation of index of financial inclusion is derived from some eminent research studies such as (Sarma, 2008), (CRISIL Inclusix, 2013), (UNDP HDI methodology, 1994), Global Findex (2012), (Kunt, & Klapper, 2013) etc. Three dimensions have been used to compute the index of financial inclusion (IFI). In order to compute composite index, firstly a dimension index is calculated for each dimension to measure the financial inclusion and later on all

dimension indices are clubbed together for constructing composite index on based of below said methodology.

The dimension index for i^{th} dimension, D_1 is computed by using following formula:

$$d_i = W_i \frac{A_i - m_i}{M_i - m_i} \qquad \ldots\ldots\ldots \qquad 3.1$$

Where

W_i = weight given to i^{th} dimension, $0 \quad W_i \quad 1$

A_i = Actual value of ith dimension

m_i = minimum value of ith dimension

M_i = maximum value of ith dimension

Formula (1) ensures that the value of index D_1 lies between 0 to 1 that means $0 \quad d_1 \quad 1$. Higher the value of d_i shows the higher achievement in dimension i and vice –versa.

The researcher used the criteria given by Sarma (2008) to attach weight with dimensions. In the n number of dimension space, the point 0 = (0, 0, 0, …..0) represent the point indicating the worst situation while point W = (w_1, w_2, w_3, ….. w_n) represent the highest achievement in all dimensions. The present index of financial inclusion is measure by the normalized inverse

Euclidean distance of the point D from the ideal point $1 = (w_1, w_2, w_3, \ldots w_n)$. Euclidean distance is the ordinary distance between two points (values of weight and dimension index) in Euclidean space.

The formula is

$$\text{IFI} = 1- \frac{\sqrt{(w_1-d_1)^2+(w_2-d_2)^2+\cdots+(w_n-d_n)^2}}{\sqrt{(w_1^2+w_2^2+\cdots+w_n^2)}} \qquad \ldots\ldots\ldots \quad 3.2$$

In formula equation 3.2, the numerator of the second component is the Euclidean distance of D from the ideal point w, normalizing it by the denominator and subtracting from 1 gives the inverse normalized distance. The normalization is done in order to make the value lie between 0 and 1 and the inverse distance is considered so that higher value of the IFI corresponds to higher financial inclusion.

Formula for the present composite index of financial inclusion after attaching weights to selected dimensions:

$$\text{IFI} = 1- \frac{\sqrt{(1-P)^2+(.5-A)^2+(.5-U)^2}}{\sqrt{1.5}} \qquad \ldots\ldots\ldots\ldots \quad 3.3$$

Weights have been attached to the all the variables (1 to Banking Penetration, .5 to Availability and .5 to Usage).

For simplification, all dimensions can be considered as equally important in measuring the inclusiveness of a financial system, then $w_i = 1$ for all i. In this case, the ideal situation will be represented by the point $I = (1, 1, 1, \ldots, 1)$ in the n-dimensional space and the formula for IFI will be

$$IFI = 1 - \frac{\sqrt{(1-d_1)^2 + (1-d_2)^2 + \cdots + (1-d_n)^2}}{\sqrt{n}} \qquad \ldots\ldots\ldots\ldots \quad 3.4$$

The IFI can be used to measure financial inclusion at different time points and for different geographic area (district, state, nation and so on).

3.2.2 Binary Logistic Regression

A binary logistic regression is most suitable econometric model in order to analyze the impact of socio-economic variables on the decision of a household about opening or not opening an account in a bank. The dependent variable is binary taking two values (0, 1), the decision of household about opening of bank account can be described by a logistic regression (Gujarati, 2013).

$$P = \frac{1}{1 + e^{-(\beta\lambda)}} \qquad \ldots\ldots\ldots\ldots \quad 3.5$$

is a vector of parameters and X is a vector of socio-economic variables, P is the probability of opening account.

For ease of exposition, this equation (3.5) can be written as:

$$P = \frac{1}{1 + e^{-Z}} = \frac{e^Z}{1 + e^Z} \qquad \ldots\ldots\ldots \quad 3.6$$

Where $Z = \beta X$

It is easy to verify that as Z ranges from - and + , P ranges from 0 and 1 and that P is related non-linearly with Z. this means that it cannot be estimated with the help of

familiar liner OLS technique. But this problem is more apparent than real as the functional form can be linearized as follows;

If P is the probability of opening an account, than the probability of not opening the account is given by

$$1\text{-}P \quad = \quad \frac{1}{1+e^Z} \quad \dots\dots\dots\dots \quad 3.7$$

Thus we can write

$$\frac{P}{1-P} = \frac{1+e^Z}{1+e^{-Z}} = e^Z \quad \dots\dots\dots\dots \quad 3.8$$

Now $P/_{1-P}$ which is the ratio of probability of opening an account to the probability of not opening an account, is simply the odds ratio in favour of opening an account - the ratio of the probability that a household will open an account to the probability that it will not open an account.

Now, if we take natural logarithm, we obtain an interesting linear form which can be estimated by linear techniques of estimation.

$$L = \ln\left(\frac{P}{1-P}\right) = Z = \beta X \quad \dots\dots\dots \quad 3.9$$

L is log of the odds ratio and it is not only linear of X, but also linear in the parameters. L is called the **logit**. But in the present work the variable P is a dichotomous or binary variable defined as taking value 1 if the household open account and zero if household does not open an account. If we substitute $P = 0$ and 1 in the left hand expression of

equation (3.9), the term $\ln(\frac{P}{1-P})$ is not defined and is meaningless. Therefore, the linearized version of the Lorenz curve can not be estimated by linear techniques like OLS. In such a situation, maximum likelihood (M. L) method should be applied. Thus, it is the maximum likelihood method that has been employed to estimate the binary logit function by using SPSS package.

In the present work, four socio-economic variables have been considered in order to examine the impact of socio-economic variables on the decision of opening an account such as household income (Y), education level of the household (E), Occupation of Family head (O) and land ownership (L).

The household income is denoted by four categories for applying the logit model. One is denoted for below Rs. 5,000/- per month family income. Two is denoted for household income between Rs. 5,000/- to Rs. 10,000/- per month. Three is denoted for household income between Rs. 10,000/- to Rs. 20,000/- per month and four is denoted for two is denoted for above Rs. 20,000/- per month household income.

The education qualification of household head is denoted by four categories. Zero is denoted for nil education qualification. One is denoted for below metric qualification and two is denoted for education qualification up to graduation education qualification. Third is denoted for above education qualification.

Occupation of household head is also denoted in five categories to analyze the impact of occupation on decision of bank account opening. One is denoted for farmer, two is denoted for laborer, three is denoted for self employed, four is denoted for government employees and five is denoted for private employees.

94

In order to analyze the impact of land holding on financial inclusion, land holding pattern of households is denoted in four categories. Zero is denoted for nil agriculture land, one is denoted for up to 2 acres, two is denoted for between 2 to 5 acres and three is denoted for above 5 acres agriculture land of household.

Logistic regression is estimated in order to analyze the impact of socio-economic variables on decision of bank account opening.

$$P = \frac{1}{1+e^{-Z}} \qquad\qquad \text{……..} \quad 3.10$$

$$Z = \beta0 + \beta1Y + \beta2E + \beta3O + \beta4L$$

Whereas,

$\beta0, \beta1, \beta2, \beta3$ and $\beta4$ are vectors of parameters.

Y = household income

E = education qualification of household head

O = occupation of household head

L = land holding by household

Therefore, this equation can be written:

$$P = \frac{1}{1+e^{-(\beta0+\beta1Y+\beta2E+\beta3O+\beta4L)}} \qquad \text{…………..} \quad 3.11$$

OLS procedure cannot be used to estimate the parameters. In such a situation, maximum likelihood (M. L) method should be applied. Thus, the maximum likelihood method has been employed to estimate the binary logit function by using SPSS package

$$L = \ln\left(\frac{P}{1-P}\right) = (\beta 0 + \beta 1 Y + \beta 2 E + \beta 3 O + \beta 4 L) \quad \ldots\ldots\ldots\ldots 3.12$$

This equation has been elaborated as earlier in this chapter. The results based on ML estimates and their interpretations are given in the fifth chapter.

3.2.2.1 Hypotheses Testing

1. H_1 = Household Income has positive impact on decision of opening bank account.

2. H_2 = Education of household head has positive impact on decision of opening bank account.

3. H_3 = Occupation of household head has positive impact on decision of opening bank account.

4. H_4 = Land holding of household has positive impact on decision of opening bank account.

The results have been discussed in fifth chapter.