Chapter-4

METHODOLOGY AND DATA DESCRIPTION

The methodology employed in the present work consists of construction of an index for financial inclusion, specification and estimation of a binary logistic curve and chi-square test in addition to simple techniques like standard deviation, coefficient of variation, percentages and frequency distribution tables.

Index of Financial Inclusion (IFI)

The Government of India's "Committee on Financial Inclusion on India" defines financial inclusion as the process of ensuring access to financial services and timely and adequate credit where needed by vulnerable groups such as weaker sections and low income groups at an affordable cost.[1]

For the purpose of the present work, financial inclusion is defined as a process that ensures access, availability and usage of the formal financial system for all member of an economy. Since banks are the gateway to the most basic forms of financial services, banking inclusion is treated as financial inclusion. All the three dimensions namely, accessibility, availability and usage refer to banking services for the present work. Several indicators have been used to measure the extent of financial inclusion. The most frequently used measure has been the number of accounts say per thousand adult persons. Some other indicators are number of bank branches, number of ATMs, amount of bank credit and bank deposit, geographic branch penetration, loan and deposit accounts, loan income and deposit income ratios and so on. All these measures provide important and useful indications about the outreach of the financial system of an economy when taken together. But while used individually, they provide only partial indication about financial inclusion and may, therefore, lead to misinterpretation of the extent of financial inclusion. Thus, a comprehensive measure which incorporates a number of measures representing several dimensions of financial inclusion preferably by, aggregating into a single number. Such as measure can be used to measure and compare the levels of financial inclusion across economies, across states and districts. It can

[1]Rangarajan Committee 2008

be used to evaluate the performance policy measures aimed at financial inclusion over a period of time.

We propose to use an index of financial inclusion (IFI) given by Mandira Sarma (2010). The proposed IFI takes values between 0 and 1, zero corresponds to lowest financial inclusion and 1 corresponds to complete financial inclusion.

4.1 Construction of IFI

As the inclusiveness of a financial system should be evaluated along several dimensions, we follow a multidimensional approach while constructing our index of financial inclusion (IFI). The present approach is similar to that used by UNDP for computation of some well-known development indexes such as the HDI, the HPI, the GDI and so on. As in the case of these indexes, our proposed IFI is computed by first calculating a dimension index for each dimension of financial inclusion. The dimension index for the i^{th} dimension, d_i, is computed according to **(**4.1.1**)** below. A weight w_i such that $0 \leq w_i \leq 1$ is attached to the dimension i, indicating the relative importance of the dimension i in quantifying the inclusiveness of a financial system.

$$d_i = w_i \frac{A_i - m_i}{M_i - m_i} \qquad \ldots.(4.1.1)$$

where

w_i= Weight attached to the dimension i, $0 \leq w_i \leq 1$

A_i = Actual value of dimension i

m_i = Lower limit on the value of dimension i, fixed by some pre-specified rule.

M_i = Upper limit on the value of dimension i, fixed by some pre-specified rule.

In the present work, for number of accounts per thousand, 0 and 1000 have been as minimum and maximum. In case of bank branches, 0 and highest figure have been used as minimum and maximum, while for deposit plus credit, 0 and highest figure have been used as minimum and maximum.

Formula **(**4.1.1**)** ensures that $0 \leq d_i \leq w_i$. Higher the value of di, higher the country's achievement in dimension i. If n dimensions of financial inclusion are considered, then, a country will be represented by a point $D = (d_1, d_2, \ldots. d_n)$ on the n-dimensional Cartesian space.

In the n-dimensional space, the point $O = (0, 0, 0, \ldots 0)$ represents the point indicating the worst situation while the point $W = (w_1, w_2, \ldots. w_n)$ represents the highest achievement in all dimensions. The index of financial inclusion, IFI, for a country, is then measured by the normalized inverse Euclidean distance of the point D from the ideal point $I = (w_1, w_2, \ldots. w_n)$. The exact formula is

$$IFI = 1 - \frac{\sqrt{(w_1 - d_1)^2 + (w_2 - d_2)^2 + \cdots + (w_n - d_n)^2}}{\sqrt{(w_1{}^2 + w_2{}^2 + \cdots + w_n{}^2}}. \qquad \ldots (4.1.2)$$

In formula (4.1.2), the numerator of the second component is the Euclidean distance of D from the ideal point w, normalizing it by the denominator and subtracting by 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.

For simplification, if we consider all dimensions to be 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

$$IFI = 1 - \frac{\sqrt{(1 - d_1)^2 + (1 - d_2)^2 + \cdots + (1 - d_n)^2}}{\sqrt{n}} \qquad \ldots (4.1.3)$$

The IFI so defined, can be used to measure financial inclusion at different time points and at different levels of economic aggregation (village, province, state, nation and so on).

4.2 Binary Logistic Curve

For assessing the influence of socioeconomic factors on the decision of a household about opening or not opening an account in a bank, a binary logistic curve is most suitable. Since the dependent variable is binary taking two values, the household action can be described by a logistic curve,

$$P = \frac{1}{1 + e^{-\beta x}} \qquad \ldots.(4.2.1)$$

β is a vector of parameters and x is a vector of socio-economic factors, P is the probability of opening account.

It is easy to verify that as βx ranges from -∞ and +∞, P_i ranges from 0 and 1 and that P is related non-linearly with x and β. But since P is nonlinear in both β and x, this means that we can not estimate it 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

$$1\text{-}P = \frac{1}{1+e^{\beta x}} \qquad \ldots(4.2.2)$$

Thus we can write

$$\frac{P}{1-P} = \frac{1+e^{\beta x}}{1+e^{-\beta x}} \qquad \ldots (4.2.3)$$

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.

$$\ln\left(\frac{P}{1-P}\right) = \beta x \qquad \ldots.(4.2.4)$$

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 (4.2.4), the term $\ln\left(\frac{P}{1-P}\right)$ 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 using SPSS package. In the present work we have considered to examine the effects of socio-economic factors such as income of the households (Y), land ownership (L), education level

of the household (E) and age of the head of the household (A) on the decision of opening an account and have specified the logistic function as:

$$P = \frac{1}{1+e^{-(\beta 0+\beta 1Y+\beta 2L+\beta 3E+\beta 4A)}} \quad \ldots (4.2.5)$$

which is non-linear in parameters and variables. The results based on ML estimates and their interpretations are given in the next chapter.

4.3Sources and Description of Data

For the purpose of fulfillment of the objectives of the study, both primary and secondary sources of data were used. For the assessment of the extent of financial inclusion at the district level, the relevant data on the number of bank branches, number of saving bank accounts, and amount of loans and credits were obtained from branch banking statistics (RBI) which furnish district wise information for the year 2008-2009. Since the size of districts is not same, it would be erroneous to use the absolute figures for inter-district comparison. To avoid this inconsistency the data were adjusted for the size by expressing the figures in terms per thousand of population. The district level figures of population pertaining to 2008-2009 are not available readily. Though the district wise population figures for the year 2011 are available from census data. The district wise population figures for the year 2008-2009 were estimated with the help of population growth rates available from 2011. For the purpose of village level analysis, primary data was employed which was collected by the researcher himself from the Atail village in Rohtak district. The data was collected through structured schedule containing closed as well as open ended questions. Before conducting the actual survey, the researcher acquired the population frame- containing the list of households residing in the village. On the basis of this list containing 1050 households in all, a sample of 100 households was selected using a stratified random sampling technique. Presuming that the caste may be a factor in the determination of financial inclusiveness, the population was first divided into three strata comprising General households (650), Backward households (125), and S.C households (275).Then the sample households were drawn randomly from the three groups in proportion to their respective shares (62 from General, 12 from Backward and 26 from S.C households) in the total households in the village. To make the selection random, the method of lottery was adopted. A table depicting the population and sample distribution of households is shown below.

Table: 4.1

Distribution of Population and Sample Households.

Strata	Population	Sample
General	650	62
Backward Class	125	12
Dalit Class	275	26
Total	1050	100

Source: Primary Survey

For the purpose of collecting information on financial inclusion, household head was taken as a unit of observation. Even if there are households having more than one account, it was counted as one. Though it would have been better to consider an individual as a unit of observation, but it was not done due to two considerations, first, it was not possible for the investigator (in this case the researcher himself) to contact each and every individual respondent, second and more important reason is that in villages all the fixed assets and income etc., are shared by all the members of the household equally and jointly. This would have made it difficult to study the influence of socio-economic factors such as literacy, age, land holding and income on the financial inclusion. A copy of the unfilled schedule is attached as an Appendix 4.1.