Chapter Three

Models, Methodology and Data

The present study adopts primarily two different types of approaches to quantitative measurement and analyses. The first is the econometric regression approach and the second is a non-econometric statistical approach. Under the first approach the OLS regression method and the binary Logit regression method are separately applied. The second is the non-econometric approach that adopts mainly statistical methods of poverty and inequality measurement.

3.1 Econometric Methods

3.1.1. Ordinary Least Squares Models

The most widespread technique to identify the contributions of various variables to poverty is regression analysis. It is commonly undertaken to identify the effects of each of the demographic, economic and social characteristics of household on per capita income (or expenditure).

The independent (right-hand side) variables may be continuous variables, such as the age of the household. Sometimes, we use categorical variables, such as type of family, gender, religion etc.. In this case we need to create a "dummy" variable. For example take gender as dummy variable. Here we assign 1 if the person is male and 0 for female. If there are, say, 10 regions in a country, each region would need to have its own dummy variable, but one of the regions needs to be left out of the regression, to serve as the point of reference. A regression estimate shows how closely each independent variable is related to the dependent variable (say, per capita consumption) holding all other influences constant.

Table 3.1.1. List of Variables for the Determinants of Poverty with Symbols and Descriptions

Variables	Symbols	Definition/Description
Dependent		
Per-capita Consumption Expenditure	PCE	
Natural log of Per-capita Consumption Expenditure	PCE (*ln*)	
Explanatory		
Type of family dummy	TOF	1=If the household is nuclear family 0=If the household is joint family
Size of household	SoH	Total number of persons in the household
Dependency ratio	DR	Ratio of persons in the "dependent" ages (generally under age 15 and over age 64) to those in "working" ages (15-64 years)
Sex ratio	SR	Females per 1000 male population. It is normalized to 100.
Age of head of household (years)	AHH	Age of the principal income earner in years
Number of earner	NoE	Number of income earner in family
Mean years of schooling (years)	MYS	Total numbers of years of schooling divided by household size
Health index (BMI)	BMI	Total body fat based on height and weight
Landholding (per household in bigha)	LAN	Total landholdings (bigha) divided by household size
Livestock (per household in Rs)	LIV	Total livestock (Rs) divided by household size
Physical assets (per household in Rs)	PHY	Total physical assets divided by household size
Distance from urban centre (km)	DUC	Distance from nearest urban centre to the residence of the household
Flood dummy	FD	1= If the household is flood and erosion affected 0=Otherwise

Model for hypothesis: One hypothesis is that the causes of poverty are beyond the control of poor households. List of dependant and independent variables are mentioned in table 3.1.2.

Table 3.1.2. List of Variables Related to Causes of Poverty with Symbols, Description and Nature

Variables	Symbol	Definition/Description	Nature
Dependant			
Per-capita Consumption Expenditure	PCE	Total monthly household consumption expenditure divided by household size	
Independent			
Age of the principal earner (years)	AH	Age of the principal earner cum head of the household (years)	IH
Size of the household	SH	Total number of members of the household	IH
Dependency ratio	DR	Ratio of persons in the "dependent" ages (below age 15 years and over age 64 years) to those in "working" ages (15-64 years) in the household.	IH
Sex ratio	SR	Females per 1000 males total population. It is normalized to 100.	IH
Mean years of schooling (years)	MYS	Total numbers of years of schooling divided by household size	IH
Health index (based on 10 indicators)	HI	Index constructed on ten (10) health related indicators.	IH
Plinth area of the residence (per capita in foot)	PAR	Level of the ground floor of a house from the land on which the building was constructed divided by household size	IH
Sanitation dummy	SD	1= If the household does have adequate sanitation (according to the MDG guidelines) 0=Otherwise	IH
Distance of school dummy	DSD	1= If distance of school from house is within 1km 0=If distance of school from house is beyond 1km	OH
Drinking water dummy	DWD	1=Presence of deep tube-well, tube-well, tap in neighbor-hood; 0=Otherwise	OH
Contribution from agriculture Dummy	PCFA	1=If contribution of agriculture to total household income is greater than 51 per cent. 0=Otherwise	OH
Distance to urban centre (km)	DUC	Distance to nearest urban centre from the residence of the household	OH
Flood dummy	FD	1=If affected by flood and erosion; 0=Otherwise	OH
Main road connectivity dummy	MRC	1= If road condition good (at least four/three wheelers can travel); 0=Otherwise	OH
Policy dummy	PD	1= If the HoH enrolled in MGNREGA; 0=Otherwise	OH

Note: IH indicates Inside the household and OH indicates Outside the household

3.1.2 Binary Logit Regression Model

To explain whether a household is poor, a logit or probit regression is used. The dependent or endogenous variable is binary, usually taking a value of 1 if the family is poor and 0 otherwise. To analyze the determinants and identify the correlates of poverty, logistic regression is to be applied to primary data. Logistic regression analysis helps to predict probability of a binary dependent variable from a set of independent variables that may be continuous, discrete, or a mix of them. Logistic regression method is a powerful technique as it is relatively free from restrictions and analyzes a mix of all types of predictors.

In case of a binary poverty variable (i.e. being poor or non-poor), let the underling response variable y*, be defined by the regression relationship:

$$y_i^* = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \text{--------} + \beta_k X_{ki} + U_i \text{ ------------------------------(4)}$$

In matrix form, it can be written as

$$y_i^* = \Sigma X''_i \beta + U_i \text{--(5)}$$

$$X''_i = [\, 1\ X_{2i}\ \ X_{3i} \text{ -------------------} X_{ki}] \text{ and}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

Let the observable dummy variable y which represents the unobserved latten variable y*, be expressed by:

$$y_i = 1 \text{ , if } y^* > 0$$

$$y_i = 0 \text{ , otherwise --(6)}$$

From equations (5) and (6) we can derive the following expression

$$\text{Prob } (y_i=1) = \text{prob } (U_i > -\Sigma X''_i\beta) = P_i$$

$$= 1 - F(-\Sigma X''_i\beta) \text{ --(7)}$$

Where,

F= cumulative distribution of U_i

If the distribution of U_i is symmetric, we can write

$$\text{Prob } (y_i=0/\beta, X''_i) = F(-\Sigma X''_i\beta) = 1 - P_i \text{ --(8)}$$

Since, the observed values of y_i are just realization of a binomial process with probabilities given by equation (8) with X_i. Thus, the likelihood function can be given by: $[F(-)]^{1-y_i}$

$L = \prod_{y_i=0}\left[F\left(-\sum X_i'\beta\right)\right] \prod_{y_i=1}\left[1 - F(-\sum X_i'\beta)\right]$--------------------------------- (9a)

Alternatively,

L=$\prod_{y_i} P_i(1 - P_i)$-- -------- (10)

However, we can write equation (9a) as

L=$\prod_{y_i}\left[F\left(-\Sigma X_i'\beta\right)\right]^{1-y_i}\left[1 - F\left(-\Sigma X_i'\beta\right)\right]^{y_i}$---------------------------------------(9b)

The functional formed imposed on F in equation (9) depends on the assumptions made about U_i in equation (5). Maddala (1983) remarks that the cumulative normal and logistic distributions are almost similar, thus using one or the other will basically lead to the same results. Moreover, according to Amemiya (1981), it is possible to derive the would-be estimate of a probit model once we have parameters derived from the logit model. Following Alemayehu et al. (2005) a logit model will be adopted for this study assuming a logistic cumulative distribution of *i u* in F. The relevant logistic expressions are:

$1 - F\left(-\Sigma X_i'\beta\right) = \frac{e^{-\Sigma X_i'\beta}}{1-e^{\Sigma X_i'\beta}}$--- (10a)

and

$F\left(-\Sigma X_i'\beta\right) = \frac{e^{-\Sigma X_i'\beta}}{1+e^{-\Sigma X_i'\beta}} = \frac{1}{1+e^{-\Sigma X_i'\beta}}$-- (10b)

Here, X_i are the characteristics of households/individuals and β_i the coefficients of the respective variables in the logit regression. To estimate equations (9) with Maximum Likelihood (ML) techniques, equation (10a) expresses the probability of being poor $[\mathrm{Prob}(y_i = 1)]$ and (10b) the probability of being non-poor $[\mathrm{Prob}(y_i = 0$. (Maddala, 1983; Amemiya, 1985; Green, 2003, Wooldridge, 2002).

After the derivation of binary logit Model, we are to choose the explanatory variables of which some are continuous variables and some are binary variables.

Table 3.1.3. List of Variables with Symbols and Descriptions for the Logistic Estimates		
Variables	*Symbol in the Estimated Equation*	*Definition/Description*
Dependent		
Poverty	POVT	Whether a household is poor or non-poor (1=Poor; 0 =Non-Poor)
Explanatory		
Size of household	HOUS	Total number of persons in the household
Dependency ratio	DEPR	Ratio of persons in the "dependent" ages (under age 15 years and over age 64years) to those in "working" ages (15-64 years)
Sex ratio	SEXR	Females per thousand males. It is normalized into 100.
Mean years of schooling	MYS	Total number of years of schooling divided by household size
Occupation dummy	OD	1= If the occupation of the principal income earner is agriculture and allied 0 = Otherwise
Health index	HI	Score based on 10 indicators of health
Distance from urban centre	DIST	Distance from residence of the household to nearest urban location
Flood dummy	FD	1= If household is affected by flood 0=Otherwise

3.2 Non-Econometric Methods

3.2.1 Measurement of Poverty and Inequality

Measurement is necessary but not sufficient. It is justified in the sense that, as Ravallion (1998) argues, "[A] credible measure of poverty can be a powerful instrument for focusing the attention of policy makers on the living conditions of the poor."

Indices of Absolute Poverty

Absolute poverty measures are of four types. The first type measures the number of the poor, that is, the headcount. A second type of poverty measurement is with the amount of income required to move the poor out of poverty, that is, the poverty gap or the income shortfall. The third class of poverty measures is concerned with the distribution of income among the poor, that is, squared poverty gap. Finally, there are composite poverty measures such as the Sen Index (Lackwood & Lynch, 1994:569).

Headcount Measure or Headcount Index (Incidence of Poverty) (HI or P_0): It measures the proportion of people below the poverty line. It is the most widely used and oldest measure of poverty. This measure literally counts heads, allowing policymakers and researchers to track the most immediate dimension of the human scale of poverty. It is popularly known as Headcount Ratio (P_0) and symbolically expressed as

$$P_0 = N_p / \mathrm{N}$$

Where,

N_p=The number of poor (i.e. persons/households below the poverty line)

N=Total population (or sample)

It can be rewrite as follows

$$\mathrm{H} = \frac{1}{N} \sum_{i=1}^{N} I(\mathrm{y_i} < \mathrm{z}).$$

Here, if expenditure ($\mathrm{y_i}$) is less than the poverty line (z), then I(·) equals 1 and the household would be counted as poor.

Poverty Gap Measure or Poverty Gap Index (Depth of Poverty) (PGI or P_1): The poverty gap index (P_1) measures the extent to which individuals fall below the poverty line (the poverty gaps) as a proportion of the poverty line (Haughton and Khander, 2010:67). This measure captures the mean aggregate income or consumption shortfall relative to the poverty line across the whole population. It is obtained by adding up all the shortfalls of the poor (assuming that the non-poor have a shortfall of zero) and dividing the total by the population. In other words, it estimates the total resources needed to bring all the poor to the level of the poverty line (divided by the number of individuals in the population) (Klugman, 2002:35). It is expressed as percentage of the poverty line. More specifically, define the poverty gap (G_i) as the poverty line (z) less actual income (y_i) for poor individuals; the gap is considered to be zero for everyone else. Using the index function, we have

G_i = (z – y_i)) × I(y_i) < z)

Then the poverty gap index (P_1) may be written as

$P_1 = \frac{1}{N} \sum_{i=1}^{N} \frac{G_i}{z}$.

This measure is the mean proportionate poverty gap in the population (where the non-poor have zero poverty gap (Haughton & Khander, 2010:67-70).

Squared Poverty Gap Measure or Index (Poverty Severity) (SPGI or P_2)**:** This takes into account not only the distance separating the poor from the poverty line (the poverty gap), but also the inequality among the poor. That is, a higher weight is placed on those households further away from the poverty line (Klugman, 2002:35).Hence,by squaring the poverty gap index, the measure (P_2) implicitly puts more weight on observations that fall well below the poverty line. Formally,

$P_2 = \frac{1}{N} \sum_{i=1}^{N} (\frac{G_i}{z})^2$

(Haughton & Khander, 2010:71-72).

FGT Poverty Measures: The Foster, Greer and Thorbecke (FGT) class of poverty has become the most popular class of poverty indices in the last two decades used in both theoretical and empirical studies of poverty.The headcount index, the Poverty Gap (PG) and Poverty Gap Index (PGI) and the Squared Poverty Gap Index (SPGI) all belong to the Foster-Greer-Thorbecke class of measures.

Using similar notation:

$P_\alpha = \frac{1}{N} \sum_{i=1}^{N} (\frac{G_i}{z})^\alpha$, (α ≥ 0)

where α is a measure of the sensitivity of the index to poverty and the poverty line is z, the value of expenditure per capita for the *i*th person's household is y_i, and the poverty gap for individual i is G_i = z – y_i (with G_i = 0 when y_i> z). The measures are defined for α ≥ 0

If we use α=0, we have the headcount index

If we use α =1, we have the poverty gap index

If we use α =2, we have the squared poverty gap index (Foster, Greer, & Thorbecke, 1984: 763).

The Sen Index

The Sen index is one of the generalised poverty gap a measure, which is built on the poverty gap. The Sen Index that integrated two simple poverty indices-the headcount ratio and poverty gap can be expressed as follows:

$$S = HC\left[PG + (1 - PG)G_p\right] = HC\left[1 - \frac{\bar{y}_p}{z} + \left(1 - 1 + \frac{\bar{y}_p}{z}\right)G_p\right]$$

$$= HC\left[1 - \frac{\bar{y}_p}{z}\left(1 - G_p\right)\right]$$

Where,

HP $=\frac{P}{N}$; [Here, P= Total No. of poor; N= Total population]

$\bar{y}_p$=Mean Income of poor people

PG=Poverty Gap = 1-$\frac{y_p}{z}$

G_p= Gini Co-efficient

Thus, the Sen index is the combination of three characteristics

a) The Head-count ratio (HC)
b) The Poverty Gap (PG)
c) The Gini coefficient (G_p)

Because of these characteristics, the Sen Index is said to include the three 'I's of poverty: Incidence, Intensity and Inequality.

The Kakwani Index (KA)

The Kakwani index (KA) that is a generalization of both the FGT and the Sen Index. It is of the following form:

$$KA = \frac{P}{Nz\left(\sum_{i=1}^{p} i^{\propto}\right)} \sum_{i=1}^{p} (z - y_i)(P + 1 - i)^{\propto}$$

The α power is used to give relatively more weight to the poorer people. The value of (P+1+i) is larger in case of extremely poor people. The KA collapses to the

Sen index with α=1 and the KA collapses to the FGT with α=0. For empirical applications, the use of KA index is meaningful if α>1.

The Thon Index (TH)

The Thon index (TH) is derived from the Sen Index. The main difference between the two is that the Thon index assumes (N+1-i) instead of (P+1-i) of the Sen Index. That is, the Thon index considers the total number of individuals (N) rather than the number of poor individuals (P). The index can be normalized as

$$TH = \frac{2}{(N+1)Nz}\sum_i (z - y_i)\,(N + 1 - i)$$

Relative Poverty (Inequality) Measures

Cowell defined "Inequality measure" as "a scalar numerical representation of the interpersonal differences in income within a given population." (Cowell, 2009:7). There are wide range of relative poverty (inequality) measures, such as, coefficient of variation, standard deviation of logarithms, Gini coefficient, Theil's entropy measure, income shares by percentiles, deciles and quintiles, Dalton's measure and Atkinson's measure. Some of the measures are mentioned as follows:

The Theil Index: The Entropy Class of Inequality Indices

The entropy class of inequality index is based on the concept of entropy that is a measure of disorder in theory of thermo-dynamics. When it is applied to income distributions, the entropy (disorder) means deviations from perfect equality.

The generalized inequality index is as follows:

$$E(\propto) = \frac{1}{n(\propto^2 - \propto)}\sum_i \left[\left(\frac{y_i}{\bar{y}}\right)^{\propto} - 1\right]$$

The equation expresses a class as the index E($\propto$) gives different forms depending on the value assigned to $\propto$. $\propto$ is a parameter ranging from minus infinity to infinity.

With $\propto = 0$, the expression becomes

$$E(0) = -\frac{1}{n}\sum_i \text{In}\left(\frac{y_i}{\bar{y}}\right)$$

With $\propto = 1$, the expression becomes

$E(1) = -\frac{1}{n}\sum_i \left(\frac{y_i}{\bar{y}}\right) \text{In} \left(\frac{y_i}{\bar{y}}\right)$

E(0) index is called the mean logarithmic deviation. E(1) is called the Theil Index. For the purpose of an operational approach, a class of Relative Entropy inequality indexes (RE) is defined. For all other numbers, i.e. α and α≠0, it is worth defining relative indexes:

$$RE(1) = \frac{E(1)}{\max\ E(1)} = \frac{\frac{1}{n}\sum_i \left(\frac{y_i}{\bar{y}}\right) \text{In} \left(\frac{y_i}{\bar{y}}\right)}{\text{In}\, n}$$

Here, RE(1) can be called the Relative Theil Index

The Atkinson Index

It is one of the most popular welfare-based measures of inequality.

The Atkinson Index may be expressed in following form:

$$A(\varepsilon) = 1 - \frac{y_{EDE}{}^{*}\sqrt{2}}{\bar{y}^{*}\sqrt{2}} = 1 - \frac{y_{EDE}}{\bar{y}}$$

Here, EDE = Equally Distributed Equivalent

For the EDE, we can get the following expression

$$y_{EDE} = \left[\frac{1}{n}\sum_i y_i{}^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$$

Atkinson Index gives indication of how much income are disposed to give up in order to have equal incomes. Given any distribution of income, EDE can be calculated easily for different levels of inequality aversion. Different levels of inequality aversion ε provide different values of y_{EDE} (Atkinson, 1970& 1983; Bellù & Liberati, 2006).

Gini Coefficient

The Gini coefficient, developed by Gini (1912), is the most common measure of distribution of income, expenditure, wealth and other attribute within a given population. The value of Gini coefficient ranges from 0 to 1. The value 0 represents a situation of perfect equality such that income (say) is identical across all households, whereas value 1 represents a situation of extreme inequality such that all income is concentrated in a single household. Between 0 and 1, the higher values of Gini coefficient are associated with higher level of inequality. It is strictly linked to the representation of income inequality through the Lorenz curve (proposed by Lorenz, 1905) that indicates which proportion of total income is in the hands of a given percentage of population. In particular, Gini coefficient is the ratio of the area between the Lorenz curve and the equi-distribution line (henceforth, the concentration area) to the area of maximum concentration area. In other words, the Gini coefficient is defined as the proportion of the total area under the diagonal that is between the diagonal (equality line) and the Lorenze curve (Bhagwati & Panagariya, 2014:258).

Formally, let x_i be a point on the x-axis, and y_i a point on the y-axis. Then,

Gini=$1-\sum_{i=1}^{N}(x_i\text{-}x_{i-1})(y_i+y_{i-1})$.

When there are N equal intervals on the x-axis, the equation simplifies to

Gini=$1-\frac{1}{N}(y_i+y_{i-1})$.(Haughton &Khander,2010:104).

Main weakness the Gini coefficient is that it cannot differentiate different kinds of inequalities. Lorenz curves representing different patterns of income distribution may intersect resulting very similar Gini coefficient values (Atkinson 1975; Cowell 1995; Bhagwati & Panagariya, 2014:261).

Multidimensional Poverty Index (MPI)

The MPI is an index of acute multidimensional poverty. The MPI reveals a different pattern of poverty than income poverty, as it illuminates a different set of deprivations. The MPI has three dimensions: health, education, and standard of living. These are measured using ten indicators. Poor households are identified and an aggregate measure is constructed using the methodology proposed by Alkire and Foster (2007, 2009).

Each dimension is equally weighted at one-third. Each indicator within a dimension is also equally weighted. Thus the health and education indicators are weighted at one-sixth each, and standard of living at one-eighteenth.

A person is identified as multi-dimensionally poor (or “MPI poor”) if he/she is deprived in at least one third of the weighted indicators. In other words, the cutoff for poverty (k) is 33.33 per cent. If a person is deprived in 20-33.3 per cent of the weighted indicators he/she is considered as “vulnerable to poverty”, and if he/she is deprived in 50 per cent or more (i.e. k=50per cent), he/she is identified as being in “severe poverty”.

Deprivation cutoffs: The MPI first identifies who is deprived in each of the 10 indicators. Here, the household is considered as the unit of analysis. In case of standard of living indicators, a person is deprived if *their household* is deprived in that particular indicator. On the other hand, for health and education indicators, a person’s deprivations depend on the achievements of *other household members*.

The MPI value is the product of two measures- the multidimensional headcount ratio and average intensity (or breadth) of poverty.

The headcount ratio (H) or the incidence of poverty is the proportion or percentage of the population who are multi-dimensionally poor:

$$H = \frac{q}{n}$$

Where, q is the number of people who are multi-dimensionally poor and n is the total population.

The intensity of poverty (A) reflects the average share or proportion of indicators in which poor people are deprived. For poor households only, the deprivation scores are summed and divided by the total number of poor persons:

$$A = \frac{\sum_i^q c_i}{q}$$

Where, c is the deprivation score of the i-th poor individual. The deprivation score c of a poor household as been expressed as the sum of deprivations in each dimension j (j = 1, 2, ……n)

The MPI is calculated by multiplying the incidence of poverty by the average intensity of poverty across the poor (MPI = H x A); as a result, it reflects both the share of people in poverty and the degree to which they are deprived.

The contribution of dimension j to multidimensional poverty can be expressed as follows:

$Contribution_j = \frac{\sum_1^q c_j}{n}$ / MPI (UNDP, 2014).

Table 3.2.1. Dimensions, Indicators, Deprivation Criteria and Weights of the MPI

Dimension	Indicator	Deprived if…	Related to	Weight	
Health	Child Mortality	If any child has died in the family	MDG 4	16.67per cent	each indicator weighted equally at 1/6
	Nutrition	If any adult or child in the family is malnourished*	MDG 1	16.67per cent	
Education	Years of schooling	If no household member has completed 5 years of schooling	MDG 2	16.67per cent	each indicator weighted equally at 1/6
	Children enrolled	If any school-aged child is out of school in years 1 to 8	MDG 2	16.67per cent	
Standard of Living	Electricity	If household does not have electricity	-	5.56per cent	each of the six indicators weighted equally at 1/18
	Cooking Fuel	If the household cooks on wood, dung or charcoal	MDG 7	5.56per cent	
	Floor	If the household's floor is dirt, sand of dung	-	5.56per cent	
	Sanitation	If the household does not have adequate sanitation (according to the MDGguidelines) or it is shared **	MDG 7	5.56per cent	
	Water	If the household does not have clean drinking water (according to MDGguidelines) or it is more than a 30 minute walk away***	MDG 7	5.56per cent	
	Assets	If the household does not own more than one of: radio, television, telephone, bicycle, motorbike, or refrigerator; and does not own a car or truck.	MDG 7	5.56per cent	

Notes: MDG 1 is Eradicate Extreme Poverty and Hunger; MDG 2 is Achieve Universal Primary Education; MDG 4 is Reduce Child Mortality; MDG 7 is Ensure Environmental Sustainability.
*Adults are considered malnourished if their BMI is below 18.5 m/kg^2. Children are considered malnourished if their z-score of weight-for-age is below minus two standard deviations from the median of the reference population.
**A household is considered to have access to improved sanitation if it has some type of flush toilet or latrine, or ventilated improved pit or composting toilet, provided that they are not shared.
***A household has access to clean drinking water if the water source is any of the following types: piped water, public tap, borehole or pump, protected well, protected spring or rainwater, and it is within a distance of 30 minutes walk (roundtrip).

Sources: Alkire and Santos , 2010:7; Alkire, Conconi and Seth, 2014:16

3.2.2 Measurement of Human Development

Inequality-adjusted Household-level Human Development Index (IHHDI)

One of the most serious weaknesses of the human development index (HDI) is that it considers only average achievements and does not take into account the distribution of human development within a country or by population subgroups. All previous studies attempt to capture inequality in the HDI with no existence of HDI at the household level. Here, we calculate the HDI at the household level.

The Household-level Human Development Index (HHDI) measures the average achievements in three basic dimensions of human development: a long and healthy life, knowledge and a decent standard of living. In other words, the HHDI is the geometric mean of the normalized three dimension indices.

Table 3.2.2. Inequality-adjusted Household-level Human Development Index

Dimensions	*A long and healthy life*	*Knowledge*	*A decent standard of living*
Indicators	Multi-dimensional health/Health score	Mean years of schooling	Per-capita Consumption Expenditure
Dimension Index	Health Index	Education Index	Consumption Index
Inequality Adjusted Index	Inequality-adjusted Health index	Inequality-adjusted education index	Inequality-adjusted consumption index
	Inequality-adjusted Household level HD Index (IHHDI)		

Source: Based on UNDP, 2010:215

The sub-indices of each dimension indices are calculated as follows:

Dimension Index= $\frac{Actualvalue\ -Minimumvalue}{Maximumvalue\ \ -Minimumvalue}$

Health Index[34] is defined as the composite measure of ten indicators, such as, vaccination schedule followed, place of child birth, BMI of children, children dying

[34]Life expectancy at age x (e_x) is the average number of additional years to be lived by a member of the cohort who survives to age x. Estimation of Life expectancy at birth (e_0) is very difficult in India due to incomplete registration of death. The Sample Registration System (SRS) of India provides data on e_0 only for the 16 major states based on sex and residence. But, the districts level data are not

before age 5 years in the last 5 years, the age of mother during the birth of a child, access to improved sanitation, access to improved drinking water, age at death of adult person, house with acceptable roof, floor and boundary, and visit to doctor /health centre. These ten indicators are supposed to reflect all aspects of human health.

Education Index[35] is computed based on mean years of schooling that is the ratio of total number of years of schooling to the size **of** household.

Income (or Consumption) Index is calculated using the Monthly Per-capita Consumption Expenditure of the household. It is also adjusted for inflation and inequality. Because, per capita income is not considered an accurate measure of the decent standard of living in context of India and the income figures are not available for social groups. The Gini coefficient (G) of inequality of MPCE is calculated and then actual, maximum and minimum MPCE values are multiplied by the corresponding equality coefficient (1-G). Thus,

$$\text{Income (or Consumption) Index} = \frac{MPCE_{actual}\ (1-G) - MPCE_{min}\ (1-G)}{MPCE_{max}\ (1-G) - MPCE_{min}\ (1-G)}$$

(IHDR, 2011:249)

There are three steps in calculating IHHDI as follows (UNDP, 2010:218-219):

Step 1.*Measuring inequality in the dimensions of the Household-level Human Development Index (HHDI)*

The IHHDI draws on the Atkinson (1970) family of inequality measures and sets the aversion parameter ε equal to 1. In this case the inequality measure is $A = 1 - g/\mu$, where g is the geometric mean and μ is the arithmetic mean of the distribution.

available (Sharma & Choudhury, 2014:180-181). However, the national and state level life expectancy values may not be uniformly applicable at the sub-state level, and there may be unsuspected life expectancy differentials across various districts of a state (Swanson & Stockwell, 1986).As life expectancy rate at birth and expected years of schooling cannot be calculated at sample survey within a community of a district; hence household level Health Index (HI) is newly formulated based on following ten (10) indicators.

[35] The household level Education Index (EI) has been calculated based on Mean years of schooling. Here, expected years of schooling are not taken into consideration as it cannot be calculated within a community of a district.

Table 3.2.3. Indicators of Health Index

Serial No.	*Indicators*	*Description and Assignment of value*
1	Whether vaccination schedule or immunization of children (below 5 years) followed completely	1= If followed without any fail 0=Otherwise
2	Place of child birth	1=If birth of child took place in hospital or medical practitioner 0=If at home
3	Presence of at least one children below normal BMI*	1= If no children below normal BMI 0=Otherwise
4	At least one children dying before age 5 years in the last 5 years	1=If no child dying before age 5 years in the last 5 years 0=Otherwise
5	In case of child below 10 years present at home, the age of mother during the birth of this child	1= If the age of mother is above 20 years at the birth of child 0= Otherwise
6	Access to improved sanitation (according to the MDG guidelines)**	1= Access to improved sanitation 0= Otherwise
7	Access to improved drinking water (WHO prescribed)***	1= Access to improved drinking water 0= Otherwise
8	Age at death of adult person who last died in the household****	1=If died at the age above 63.3 years 0= Otherwise
9	House with acceptable roof, floor and boundary	1= If acceptable 0= Otherwise
10	Visit to doctor /health centre/hospital, whenever required	1= If visited 0= Otherwise

Notes: *Adults are considered malnourished/ undernourished if their BMI is below 18.5.

**A household is considered to have access to improved sanitation if it has some type of flush toilet or latrine, or ventilated improved pit or composting toilet, provided that they are not shared.

***A household has access to clean drinking water if the water source is any of the following types: piped water, public tap, borehole or pump, protected well, protected spring or rainwater, and it is within a distance of 30 minutes walk (roundtrip).

**** Life expectancy rate at birth (2009-13) in India and Assam is 67.5 and 63.3 years respectively (Sample Registration System and SRS Statistical Report, Office of the Registrar General of India, Ministry of Home Affairs; Economic Survey 2014-15:A129.)

This can be written as:

A_x=1- $\frac{\sqrt[n]{X_1 ----- X_n}}{\bar{X}}$

Where, {X_1 ..., X_n} denotes the underlying distribution in the dimensions of interest. *Ax* is obtained for each variable (health score, mean years of schooling and per capita consumption expenditure).

Step 2.*Adjusting the dimension indices for inequality*

The mean achievement in an HHDI dimension, $\bar{X}$, is adjusted for inequality as follows:

$$\bar{X}^* = \bar{X}^* \text{ (1-}A_x\text{)} = \sqrt[n]{X_1 ---- X_n}$$

Thus, the geometric mean of the distribution ($\bar{X}^*$) reduces meanaccording to the inequality in distribution. Thus, inequality-adjusted dimension indices$\left(I_{I_x}\right)$ are obtained from the human development dimension indices (*Ix)* as follows:

$$I_{I_x} = (1 - A_x)I_x$$

Where,

I_{I_x}= Household-level Inequality-adjusted dimension indices

I_x= Household-level Human Development dimension indices

Ax= Atkinson measure of each dimension

Step 3.*Combining the dimension indices to calculate the Inequality-adjusted Household-level Human Development Index (IHHDI)*

The IHHDI is the geometric mean of the three dimension indices adjusted for inequality. First, the IHHDI that includes the unlogged income index (IHHDI*) is calculated:

$$\text{IHHDI*} = \sqrt[3]{I_{Health} I_{Education}\ I^*_{Consumption}} =$$

$$\sqrt[3]{(1 - A_{Health})I_{Health}(1 - A_{Education})I_{Education}\left(1 - A_{Consumption}\right)I^*_{Consumption}}$$

The HHDI based on unlogged income index (*HHDI**) is then calculated:

$$\text{HHDI*} = \sqrt[3]{I_{Health} I_{Education}\ I^*_{Consumption}}$$

The percentage loss to the *HHDI** due to inequalities in each dimension is calculated as:

$$\text{Loss}=1-\frac{IHHDI^*}{HHDI^*}=1-\sqrt[3]{(1-A_{Health})(1-A_{Education})(1-A_{Consumption})}$$

Assuming that the percentage loss due to inequality in income/consumption distribution is the same for both average income/consumption and its logarithm, the IHHDI is then calculated as:

$$IHHDI=\left(\frac{IHHDI^*}{HHDI^*}\right)$$

$$\text{HHDI}=\sqrt[3]{(1-A_{Health})(1-A_{Education})(1-A_{Consumption})}HHDI$$

The co-efficient of human inequality is defined as "an un-weighted average of inequalities in health, education and income (or consumption)" as follows (UNDP, 2014):

$$\text{Co-efficient of human inequality}=\frac{A_{Health}+A_{Education}+A_{Consumption}}{3}$$

3.3 Data: Nature and Sources

For the proposed study, both the primary and secondary data shall be used. The primary data were purposefully collected by the researcher in a set of well-designed questionnaire from the sample respondents. The survey was conducted during January–December 2014. Before using the household questionnaire for actual survey, it was tested by conducting a pilot survey among the thirty households of Mishing tribe in Dhemaji and Sivasagar district. It was carried out to ensure that the respondents understand the questions properly and confirm the appropriateness of questions. The questionnaire was revised keeping in mind the errors detected in pilot survey. In addition to the household survey, data were collected from the official records of Autonomous Councils, the offices of the health, education and other government departments. A direct participatory and observation technique was also being employed where necessary.

The secondary data were collected from the different sources, such as, National Sample Survey (NSS), Central Statistical Organisation (CSO), Planning

Commission, UNDP, World Bank, Different Ministries of India, Department of Statistics and Economics, Gazetteer, Official Records, Block Level Stations, Statistical Abstracts, Journals, Books and the leading libraries and socio-economic research centres of the North-eastern states, and institutions. In addition to these, a few internet data and literatures were also used. These shall be used for the research work after thorough examination of their accuracy.

We selected three blocks from each district and two villages from each blocks. Thus, total twelve sample Mishing villages are selected from two sample districts. The blocks and villages have been selected purposively, but households are selected randomly. While selecting the villages, we lay emphasis on variability or diversity of data, such as, population size of the village, access to electricity and safe drinking water, distance to town, transport and communication, pattern of occupation, pattern of income and consumption, literacy rate, school enrolment etc. These factors are considered to influence poverty, inequality and human development among Mishing tribe. From each district six villages are selected. Out of which two villages are regularly flood and erosion affected, two villages are less flood affected and remaining two are mixed-populated. After having selected the sample villages and the number of sample households from each village, we drew random samples of Mishing households from selected village. In the first step, we enumerated the Mishing households from each selected village with their names. In second step, the names of the Mishing households were arranged in an ascending lexicographic order assigning a serial number 1 to $N_{i.}$ Here, N_i is the total number of Mishing households in the *i*th village (i=1, 2, 3….12). Then, uniformly distributed random numbers lying between 1 and N_i were generated for each village. Finaly 25 per cent of Mishing households were randomly chosen from each selected village giving a total sample size of 373 households for the present study.

In all Assam, total numbers of Mishing households and population are 1,17,825 and 680,424 respectively as per 2011 census. Among all districts of Assam, the highest numbers of Mishing people (32.34 per cent) live in Dhemaji district. Recently, the total Mishing populations living in Dhemaji and Sivasagar districts are 111,732 and 27,834 respectively (as per census 2011). Total 97.27 per cent of Mishing people of Dhemaji district live in rural area and that figure in Sivasagar district is 97.89 per cent. For the present study, the village-level Mishing household list was prepared based on ASHA[36] register.

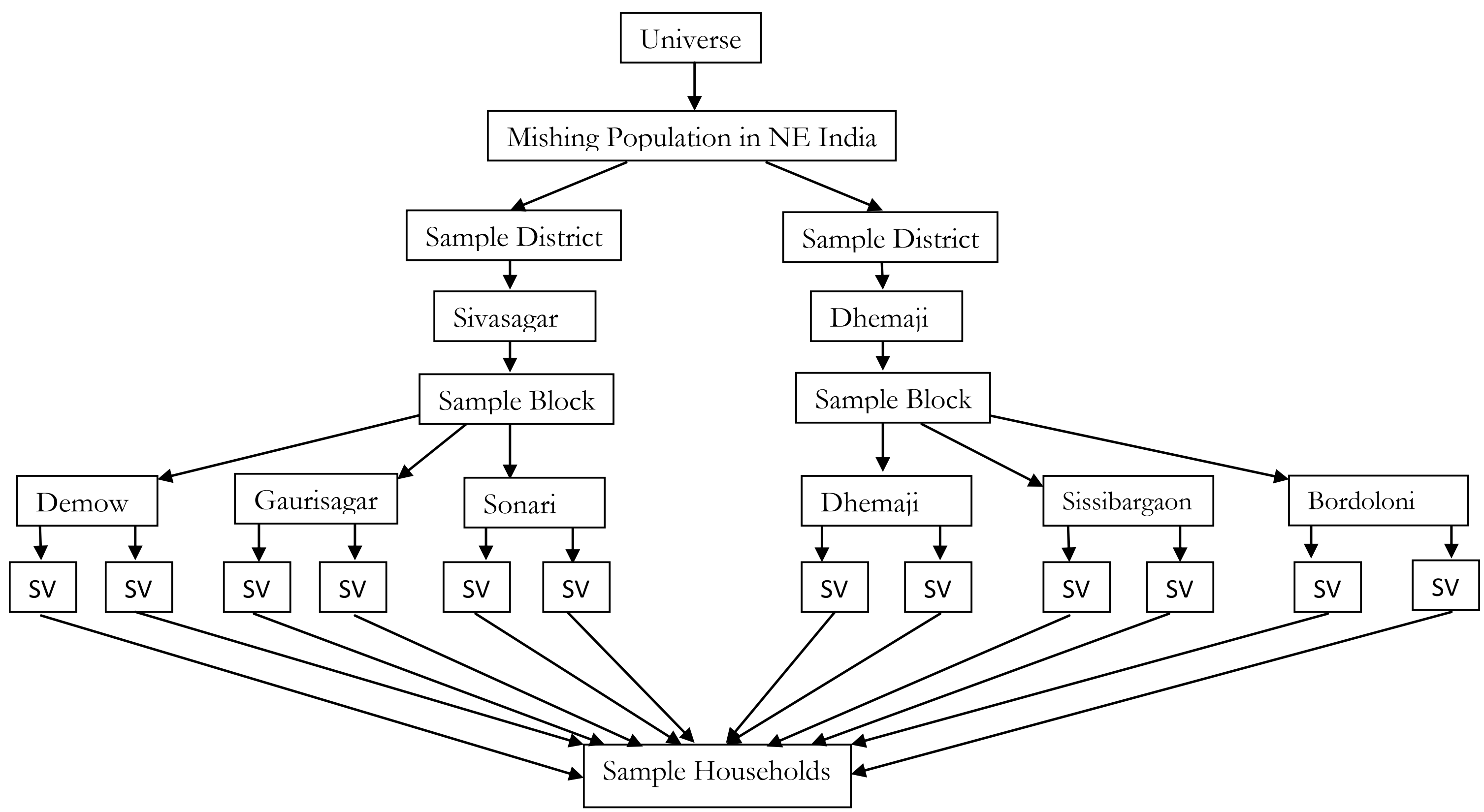
Table 3.3.1. District and Block-wise Sample Mishing Villages, Total and Sample Mishing Households

District	*Block*	*Village*	*Total Number of Mishing Households*	*25 per cent of the Total Sample Households*	*Total Sample Households*
Dhemaji	Dhemaji	Kekuri Mishing	211	53	181
		Bar Mathauri	107	27	
	Sissibargaon	Bormuria Bokajan	103	26	
		Muktiar	143	36	
	Bordoloni	No.1 Mainapara	101	25	
		Ratua Pathar	56	14	
Sivasagar	Demow	Dimowmukh	133	33	192
		Dolopa	175	44	
	Gaurisagar	Thekeratal	103	26	
		No.1 Alimur	92	23	
	Sonari	No. 2 Balikhuti	85	21	
		Ramnagar	180	45	
Two Districts	Six Blocks	Twelve Villages			373

Source: Field Study

36 Accredited Social Health Activists (ASHAs) are the community-level health workers under the National Rural Health Mission (NRHM) of the Ministry of Health and Family Welfare (MoHFW) of India. ASHAs are literate woman resident of a village aged 25 to 45 years. They are selected incorporating different community groups, self-help groups, Anganwadi Centre, Block Nodal officer, District Nodal officer, Village Health Committee and the Gram Sabha. The ASHAs, who have to undergo series of training, mobilize the community and assist them in availing health services of different government schemes, counsel women on birth preparedness, and provide good health practices. They generate awareness on health and its determinants, such as nutrition, sanitation, living and working conditions, & family welfare services. They always maintain a register where they keep all the up-to-date demographic records, such as, male, female, death, birth, age etc. of the members of a household.

Figure 2. Sampling Design



Note: SV means Sample Villallage