Chapter II
Methodology

2.1. Sources of Data

The present study used primary as well as secondary data to fulfil its stated objectives. The study first takes an overview of out-migration as source of livelihood in the state of Bihar and then it moves to an area where out-migration is concentrated to understand the latent force in migration process in diversifying the rural and agricultural economy. To understand the scenario of out-migration from Bihar the migration table presented in D – Series of 2001 census has been used, which counts migrants on the basis of Place of Birth and Place of Last Residence. Apart from measuring the volume and stream of migration from census, causes and socio-economic determinants of migration have been understood from NSS 64th round data. In Schedule 10.2 of NSS 64th round (July 2007 – June 2008) information was collected regarding the employment-unemployment characteristics and migration particulars. Information of out-migration was collected from left behind households. Therefore, it opens up the scope to understand out-migration in the context of area of origin. The other objectives of the study have been fulfilled through data collected from primary survey.
2.2. Primary Survey

2.2.1. Selection of study area

Barauli block in Gopalganj district is selected as the area of study on the basis of two criteria. First is higher out-migration from the area. Proportion of out-migrant is third highest in Gopalganj, and other is sex ratio (male/female) which is also very low in the study block (947 male per 1000 female according to census 2001), which is a clear indication of higher male out-migration from the block. This area shows a typical characteristic of dependency on agriculture, among the main male work force, 47 percent are cultivators and 30 percent are agriculture laborer.

2.2.2. Sampling Design

There are 71 villages in this block out of which 10 villages will be selected according to Probability Proportion to Size (PPS) method. Household is the unit of sample. Circular systematic random sampling method is adopted for sampling.

2.2.3. Sample Size

Formula:

\[ n = \frac{t^2 \times p(1-p)}{m^2} \]
Description:

\[ n = \text{required sample size} \]

\[ t = \text{confidence level at 95\% (standard value of 1.96)} \]

\[ p = \text{estimated prevalence expressed as decimal, which is considered 0.3 as per the evidences from NSSO data.} \]

\[ m = \text{margin of error at 5\% (standard value of 0.05)} \]

The sample size by applying this formula is 323. Taking account of non-response total 350 households will be selected. Considering the design effect 450 samples were fixed for the study.

2.2.4. Experiences and observations from field

Villages were well connected with road transport with block head quarter, which was a visible sign of development of the area. Due to good connectivity, it was possible to stay in block and commute to villages. Many small market villages have been evolving in the area, where many facilities and amenities such as internet, zerox, fast-food stall and hotels are available. Such villages were taking away the image of backwardness and marginalization of the area. The abundance of Chaomin stalls in these small market town and villages was symbolizing that this area is not secluded from ongoing modernizations and changes in preferences. Almost all the affluent households have been affording a bike for commuting to block and district headquarters. It shows that they are realizing the need of greater connectivity. Although the proportion of Muslims was sizeable in the area, they were usually not living in separate settlements. It was found that both communities were participating in the festivals of
each other. In some of the villages Muslims were able to attain better socio-economic status compared to Hindus due to higher participation in international migration. But, their Hindu neighbours were also found interested to go. Because of the emergence of print media at local level and several social security schemes of government, politics at local level were found vibrant.

During the survey, people were not found sceptical towards outsiders and at many instances they have shown welcoming gesture. Poor households showed lesser tendency of hiding their socio-economic status and income, while at several instances richer households also proudly declared their wealth. Some kind of informal conversation has been very helpful to know their real socio-economic situation.

2.3. *Concepts and definitions*

**Household or Family:** Households and families are basic units of analysis in demography. "Family" refers only to those kin with whom one co-resides. It refers quite generally to a group of kin, i.e., persons related by blood, marriage, or adoption. A household is composed of one or more people who occupy a housing unit, whether they all are related or not. Therefore, family and household both is not the same thing, not all households contain families. In Indian context, family and household overlap each other as people share the same dwelling unit are related to each other, with a few unrelated people. Throughout this study, household and family, both of the words have been used interchangeably.
**Household member:** The study adopts a household oriented approach to understand migration process, in which migration has been considered as source of livelihood of the household at the area of origin. Therefore, in this study, migrants ‘for economic or employment related reasons’ can’t be seen as permanently separated from the household. In some surveys, like IHDS (India Human Development Survey) – 1, migrants have been defined as non-resident household members.

**Out-migrants:** In this study, the definition of out-migrants varies according to the nature of the data and the purpose of the study. Out-migrants have been defined with three typologies.

*Out-migrants in Census data:* In census data, out-migrants are considered as those people, whose place of birth (POB) or place of last residence (POLR) was in the state (Bihar) but enumerated out of the boundary of the state.

*Out-migrants in NSS data:* In NSS data out-migrants are defined as those former members of the household, residing currently at different place.

*Out-migrants as the member of the household:* This is the operational definition of the out-migrants adopted in this study. It includes those migrants who migrated for employment related reasons, other economic reasons (business, transfer of service or contract) or connected to the household by sending remittances.

**Household or family structure:** The measurement of household and family structure centers around the notion of departures from what is presumed to be the simplest, or rudimentary
form, the nuclear group of an adult couple and their children. More complex structures are seen as the result of additions of other kins or the addition of unrelated persons such as servants. A rough obverse notion of complexity centres around the extent to which adults other than husband-wife couples tend to maintain separate households. Parish and Schwartz (1972) used a measure of the "number of marital units per household", where the number of marital units is defined as the number of married males, and the number of each widowed and divorced males and females. A ratio of greater than one indicates doubling up of marital units and thus household complexity. The same concept has been adopted in another study (Schmink, 1984). In this study, the root couple (or any surviving male or female from the root couple) will not be counted as marital unit.

**Activity status of household members:** Household members in households without any migrant are broadly classified into two categories; workers, non-workers. Workers are those male and female members in the household who are engaged in any economic activity either as their usual principal activity (more than 6 months in the reference period of last one year) or with subsidiary capacity (less than 6 months in the reference period of last one year). A person who did not work at all during the reference period was treated as a non-worker (Census of India, 2011). In this study, members of a household are classified into three categories: **Non-workers, workers and out-migrant members.** As mentioned earlier that out-migrants (for employment related reasons, other economic reasons (business, transfer of service or contract) or connected to the household by sending remittances) are also considered as the members of the household.

**Sources of Livelihood:** Sources of livelihood of households have been defined as the activity in which any member of household is engaged as main or marginal worker. Activities are
classified into two broad categories, agricultural and non-agricultural. Hunting, forestry and fishing are also included in agricultural activities, rest of the activities are classified as non-agricultural activities. Therefore, the source of livelihood for a household, who don’t have any migrant for economic purpose, can be classified into two broad categories; agricultural and non-agricultural. In the household with at least one migrant member for economic purpose, migration is also considered as source of livelihood. If all available workers of a household is engaged in just one sector of these two or have no working member except migrant for economic purpose, the livelihood of the household is considered as *specialised*. If the member of the household is engaged in more than one sector including migration for employment purpose, the household is considered as diversified.

### 2.4. Methods for chapter five

#### 2.4.1. Lorenz Curve

This method is used to show the inequality in income and expenditure in the household. The Lorenz Curve is a tool used to represent income distributions as proposed by Lorenz (1905); it tells us about the extent of total income in the hands of a given percentage of population. This method is conceptually very similar to the method by quintiles. However, instead of ending up with income shares, the Lorenz Curve relates the cumulative proportion of income to the cumulative proportion of individuals.

**The Lorenz Curve is obtained as follows:** The $x$-axis records the cumulative proportion of population ranked by income level. Its range is therefore (0,1). The $y$-axis records the cumulative proportion of income for a given proportion of population, i.e. the income share
calculated by taking the cumulated income of a given share of the population, divided by the total income $Y$, as follows:

$$l\left( \frac{k}{P} \right) = \frac{\sum_{i=1}^{k} y_i}{Y}$$

where:

$k=1,\ldots,n$ is the position of each individual in the income distribution;

$i=1,\ldots,k$ is the position of each individual in the income distribution;

$P$ is the total number of individuals in the distribution;

is the income of the $i^\text{th}$ individual in the distribution $iy$;

is the cumulated income up to the $k^\text{th}$ individual. $\Sigma = kiiy1$

It is apparent that ranges between 0, for $k=0$, and $Y$, for $k=n$, therefore $\Sigma = kiiy1$ and it ranges between 0 and 1

\[2.4.2. \textbf{Gini Coefficient}\]

This method is used to show inequality in income and expenditure in the household. The Gini coefficient is a measure of inequality of a distribution. It is defined as a ratio with values between 0 and 1: the numerator is the area between the Lorenz curve of the distribution and the uniform distribution line; the denominator is the area under the uniform distribution line.

It was developed by the Italian statistician Corrado Gini who published in his 1912 a paper on "Variabilità e mutabilità" ("Variability and Mutability"). The Gini index is the Gini coefficient expressed as a percentage, and is equal to the Gini coefficient multiplied by 100.
The Gini coefficient is equal to half of the relative mean difference. The Gini coefficient is often used to measure income inequality. Here, 0 corresponds to perfect income equality (i.e. everyone has the same income) and 1 corresponds to perfect income inequality (i.e. one person has all the income, while everyone else has zero income).

2.4.3. Principal Component Analysis (PCA) to create Relative Wealth Index for Households

One common method, frequently used in the analysis of household surveys in, is to create an index based on the “count” of household assets. The rationale has been that since there is no other way of weighting the various assets, assuming an equal weight of each was a reasonable way to proceed. This approach, however, can lead to inaccurate results since two individuals with very different economic resources and therefore standards of living can be assigned the same wealth score. Instead, a more appropriate methodology is required, one in which the distribution of household assets weights more heavily luxury assets. In order to make those weights non-arbitrary and replicable, it is calculated systematically, based on the Principal Component Analysis (PCA) method described below.

Filmer and Pritchett (2001) noted that asset-based measures depict an individual or a household’s long-run economic status and therefore do not necessarily account for short-term fluctuations in economic well-being or economic shocks. Thus, although the expected income variable should be correlated with the wealth measure here estimated, it should be noted that the two might tap different dimensions of economic well-being, as previous studies have found (Gasparini et. al., 2008). Following Filmer and Pritchett, many other studies, especially in the fields of economics and public policy, have implemented and recommend the use of
PCA for estimating wealth effects. The estimation of relative wealth using PCA is based on the first principal component. Formally, the wealth index for household $i$ is the linear combination,

$$y_i = \alpha_1 \left( \frac{x_1 - \bar{x}_1}{s_1} \right) + \alpha_2 \left( \frac{x_2 - \bar{x}_2}{s_2} \right) + \cdots + \alpha_k \left( \frac{x_k - \bar{x}_k}{s_k} \right)$$

Where, $k x$ and $k s$ are the mean and standard deviation of asset $k x$, and $\alpha$ represents the weight for each variable $k x$ for the first principal component. By definition the first principal component variable across households or individuals has a mean of zero and a variance of $\lambda$, which corresponds to the largest eigenvalue of the correlation matrix of $x$. The first principal component $y$ yields a wealth index that assigns a larger weight to assets that vary the most across households so that an asset found in all households is given a weight of zero (McKenzie 2005). The first principal component or wealth index can take positive as well as negative values.

### 2.4.4. Pearson product-moment correlation coefficient

In statistics, the **Pearson product-moment correlation coefficient** (sometimes referred to as PCC or Pearson's $r$) is a measure of the linear correlation between two variables $X$ and $Y$, giving a value between $+1$ and $-1$ inclusive, where $1$ is total positive correlation, $0$ is no correlation, and $-1$ is total negative correlation. It is widely used in the sciences as a measure of the degree of linear dependence between two variables. It was developed by Karl Pearson.
Pearson's correlation coefficient when applied to a population is commonly represented by the Greek letter $\rho$ (rho) and may be referred to as the population correlation coefficient or the population Pearson correlation coefficient. The formula for $\rho$ is:

$$
\rho_{X,Y} = \frac{\text{COV}(X, Y)}{\sigma_X \sigma_Y}
$$

where:

- COV is the covariance
- $\sigma_X$ is the standard deviation of $X$

### 2.5. Methods for Chapter six

#### 2.5.1. Multinomial logit model

Multinomial logit models are used for nominal scale data having three or more categories, which are not in any intrinsic order. It estimates a series of binary logit models. One group is chosen as the referent group and other the comparison group. If there are $k$ numbers of categories of the outcome variable, it estimates equations for $k-1$ number of groups.

The following multinomial logistic regression model has been used in this study:

$$
Z_1 = \log\left(\frac{P_1}{P_3}\right) = a_1 + \sum_{j} a_j X_j
$$

$$
Z_2 = \log\left(\frac{P_2}{P_3}\right) = a_2 + \sum_{j} b_j X_j
$$
\[ a_i, i = 1,2: \text{constants} \]
\[ b_{ij}: i=1,2, \ldots, n, j=1,2, \ldots, n : \text{multinomial regression coefficient}. \]

P1: Estimated probability of low birth weight.

P2: Estimated probability of normal birth weight.

P3: Estimated probability of high birth weight.

Through this model, the predicted probabilities can also be derived, which are usually easier to understand than the coefficients or the odd ratios. For getting the predicted probabilities, the `prvalue` command is used in STATA. This can be used with either a categorical variable or a continuous one and shows the predicted probabilities for each of the values of the variable specified. Here, it can be seen how the probabilities of the membership to each category of birth weight change as a particular independent variable is varied and the other independent variables are held at their means. This method is used to understand the diversification of income in the household. The dependent variable, diversification of income is a has three categories, income from one source, income from two sources and income from three or more sources. The several socio-economic variables such as place of religion (Hindu and Muslim), caste group (SC/ST, OBC and Others), education level of head (up to primary, middle and secondary, higher secondary and above), size of landholding they possess (No land, 0.01 to 0.2 Hectare, 0.2 Hectare and above), number of young members in the household (15 to 44 years) (No one, one, two and three or more members) complexity of the household (one couple, two couple, three or more).
2.5.2. *Gini Decomposition by Income Source*

The Gini coefficient is widely used to measure inequality in the distribution of income, consumption, and other welfare proxies. Decomposing this measure can help you understand the determinants of inequality. The techniques used more often decompose inequality either by subpopulations or by income source.

Extending the results of Shorrocks (1982), Lerman and Yitzhaki (1985) show that the Gini coefficient for total income inequality, $G$, can be represented as

$$ G = \sum_{k=1}^{K} S_k G_k R_k $$

where $S_k$ represents the share of source $k$ in total income, $G_k$ is the source Gini corresponding to the distribution of income from source $k$, and $R_k$ is the Gini correlation of income from source $k$ with the distribution of total income ($R_k = \text{Cov}\{y_k, F(y)\}/ \text{Cov}\{y_k, F(y_k)\}$, where $F(y)$ and $F(y_k)$ are the cumulative distributions of total income and of income from source $k$).

As noted by Stark, Taylor, and Yitzhaki (1986), the relation among these three terms has a clear and intuitive interpretation; the influence of any income component upon total income inequality depends on
• how important the income source is with respect to total income (Sk);
• how equally or unequally distributed the income source is (Gk); and
• how the income source and the distribution of total income are correlated (Rk).

If an income source represents a large share of total income, it may potentially have a large impact on inequality. However, if income is equally distributed (Gk = 0), it cannot influence inequality, even if its magnitude is large. On the other hand, if this income source is large and unequally distributed (Sk and Gk are large), it may either increase or decrease inequality, depending on which households (individuals), at which points in the income distribution, earn it. If the income source is unequally distributed and flows disproportionately toward those at the top of the income distribution (Rk is positive and large), its contribution to inequality will be positive. However, if it is unequally distributed but targets poor households (individuals), the income source may have an equalizing effect on the income distribution.

Lerman and Yitzhaki (1985) show that by using this particular method of Gini decomposition, you can estimate the effect of small changes in a specific income source on inequality, holding income from all other sources constant. Consider a small change in income from source k equal to ey_k, where e is close to 1 and y_k represents income from source k. It can be shown (see Stark, Taylor, and Yitzhaki [1986]) that the partial derivative of the Gini coefficient with respect to a percent change e in source k is equal to

$$\frac{\partial G}{\partial e} = S_k(G_k R_k - G)$$
where $G$ is the Gini coefficient of total income inequality prior to the income change. The percent change in inequality resulting from a small percent change in income from source $k$ equals the original contribution of source $k$ to income inequality minus source $k$’s share of total income:

$$\frac{\partial G/\partial e}{G} = \frac{S_k G_k R_k}{G} - S_k$$

2.6. Methods for Chapter Eight

2.6.1. Binary logistic regression

To study the determinants of use of remittances, binary logistic regression analysis has been used. Association is seen between first use of remittance and several socio-economic and household level characteristics; binary logistic regression has been used. The dependent variable, first use of remittance is a binary one categorised into food and non-food items, food items being the reference category. The several socio-economic variables such as place of religion (Hindu and Muslim), caste group (SC/ST, OBC and Others), education level of head (up to primary, middle and secondary, higher secondary and above), size of landholding they possess (No land, 0.01 to 0.2 Hectare, 0.2 Hectare and above), number of young members in the household (15 to 44 years), complexity of the household (one couple, two couple, three or more), type of destination of migration (Internal and International). The household level characteristics are indicated by the complexity of households by the number of couples coexisting in a particular family and the age and sex of the household head.
In SPSS statistical output, the parameter estimate is the \( b \) coefficient used to predict the log odds (logit) of the dependent variable. Let \( z \) be the logit for a dependent variable, then the logistic prediction equation is:

\[
Z = \ln \left( \frac{\text{odds(events)}}{\text{prob non event}} \right) = \ln \left( \frac{\text{prob(event)}}{1-\text{prob(event)}} \right) = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k
\]

Where \( b_0 \) is constant and \( k \) is the number of independent (X) variables. In ordinal logistic regression, the threshold coefficient will be different for every order of dependent variables. The coefficient will give the cumulative probability of every order of dependent variables.

The dependent and independent variables have been dichotomised before running the regression analysis for better interpretation of the results, the interest category being coded as 1 and the reference category as 0.

2.6.2. Multinomial logit model

This method is used to understand the determinants of out-migration from household. The dependent variable, type of out-migration from household has three categories, no migration, internal migration (migration within the national boundary), and international migration. The several socio-economic variables such as place of residence, religion (Hindu and Muslim), caste groups (SC/ST, OBC and Others), education levels of head (up to primary, middle and secondary, higher secondary and above), size of landholding (No land, 0.01 to 0.2 hectares, 0.2 hectares and above), number of young members in the household (15 to 44 years) (No
one, one, two and three or more members), and complexity of the household (one couple, two couple, three or more) are used in this study.

2.7. Use of qualitative method

As much as 15 in-depth interviews have been conducted to understand the process of international migration and its impact on the household during the field work. Five respondents were return migrants, five were defacto head of the household and five were relatives of migrants.