CHAPTER III
DATA BASE AND RESEARCH METHODOLOGY

The design of any research requires considerable understanding of the research methods and the data analysis. Therefore, research design provides the framework for the collection and analysis of data. Thereby, present chapter elucidates the detailed description of the population, sample, sampling design and research instruments applied to collect the data and statistical tools used to analyze the data. Therefore, the present chapter was divided into four sections. Section I, deals with the sample and sampling design. Section II, explains the database, variables and hypothesis of the present study. Section III, describes the methodology applied in the present study and finally Section IV, explains the limitations of the present study.

SECTION I
3.1 SAMPLING DESIGN AND SAMPLE SIZE

3.1.1 Sampling Design

The study entitled, “Health Insurance for the Urban Informal Sector in Punjab: An Empirical Analysis” is a cross-sectional in design. The present study involved multistage stratified random sampling technique to select the districts, occupations and the clusters of the informal sector workers from the Punjab. The first stage involved the selection of the districts from Punjab. It was planned to give a true representation of the three cultural belts of Punjab i.e., Majha (Amritsar, Gurdaspur and Tarn Taran), Doaba (Hoshiarpur, Jalandhar, Kapurthala and Nawanshahar) and Malwa (Barnala, Bathinda, Faridkot, Fatehgarh Sahib, Ferozepur, Ludhiana, Mansa, Moga, Mohali, Muktsar, Patiala, Rup Nagar and Sangrur). Thus, Amritsar, Jalandhar and Ludhiana were selected from Majha, Doaba and Malwa respectively as these districts have the highest proportion of urban population in their respective belts (Business Standard, 2011; Planning Commission, 2013; Statistical Abstract Punjab, 2015). Thereafter, an effort was made to select the three occupations from the informal sector. The operational definition of urban informal worker was adapted from the works of Papola (1980), Fields (1990), Tornberg et al.
(1996), Loewenson (1998), Forastieri (1999), Gumber (2000), Canagarajah & Sethuraman (2001), Betcherman (2002), Phoon et al. (2003), Chattopadhyay (2005), Tajgman, (2006), Cuevas et al. (2009), Mac Kellar (2009), Sandra (1999), Bocquier et al. (2010), Puerta (2010), Acharya et al. (2013), Ametepeh et al. (2013) and Sowah et al. (2013). According to the operational definition an urban informal sector worker is the worker who works without labor contract, tends to work on small scale basis and ill-equipped in terms of education and skills, possesses few resources for significant physical investment such as premises, equipment and machinery. The workers employed in the informal sector are not registered with the authorities and do not participate in the official tax system and social security system to meet the regulatory requirements. Normally, informal sector workers operate on small scale basis with a small level of organization and little or no division between the labor and the capital. Three occupational groups such as construction worker, vendor (Rehri-wala) and non-registered shopkeeper were selected for the present study. These occupational groups were selected on the basis of the facts that (a) These occupational groups are commonly found in all the urban districts of Punjab and (b) The occupational group representatives and the working environment of these groups allows conducting a survey on the demand and capacity to pay for the health insurance.

3.1.2 Sample Size

The sample of the present study was selected from the three urban districts of Punjab: Amritsar, Jalandhar and Ludhiana and from each district seven major cluster of the construction workers, vendors and non-registered shopkeepers were identified. A sample of 10 workers was selected from each cluster and total 630 workers employed in informal sector were surveyed for the present study. The cluster of the vendors and the non-registered shopkeepers were included on the basis of the list obtained from the municipal corporation offices of Amritsar, Jalandhar and Ludhiana. An attempt was made to find those clusters which involve more concentration of vendors and non-registered shopkeepers. There was no list available related to the cluster of the construction workers in Amritsar, Jalandhar and Ludhiana. Thus, clusters of the construction workers were decided on the basis of discussion with the experts and snowball sampling. From each cluster, the respondents were selected randomly. In order to give a better insight of the
informal sector, only those workers who had worked in their present occupation from the past five years were included. The cluster of the vendors selected from Amritsar included Bus Stand, Putlighar Chowk, Majitha Road, Ram Bagh, Shaheedan Road, Hakima Gate and Railway Station. The cluster of the construction workers selected from Amritsar included Ratan Singh Chowk, Hall Bazar, Sultan Wind Gate, Bakar Mandi, Vijay Nagar, Putlighar Chowk and Majitha Road. The cluster of the shopkeeper selected from Amritsar constituted Hakima Gate, Majitha Road, Batala Road, Sultan Wind Road, Ram Tirath Road, Tehsil Pura and Chheharta Road. The cluster of the vendors selected from Jalandhar included Bus Stand, Jyoti Chowk, B.N.C Chowk, Kapurthala Chowk, P.N.B Chowk, Railway Station and Ambedker Chowk. The cluster of the construction worker included from Jalandhar were Rama Mandi, Ladowali Road, Jyoti Chowk, Ravidass Chowk, Patel Chowk, Football Chowk, Doaba Chowk, Nakodar Chowk and Dana Mandi. The clusters of the shopkeeper from Jalandhar were Masand Chowk, Basti Guzan Chowk, Mai Hira Gate, Sodal Mandir Chowk, Preet Nagar, Basti Bawa Khel Road and Lama Pind Chowk. The cluster of the vendors from Ludhiana included Pakhowal Road, Railway Station, Gill Road, Janak Puri, Durgi Road, Chaura Bazar and Basti Jodhewal. The cluster of the construction workers from Ludhiana were Clock Tower, Pakhowal Road, Old Sabzi Mandi, Partap Chowk, Focal Point, Gill Road and Samrala Chowk. The cluster of the shopkeepers from Ludhiana included Basti Jodhewal, Lajpat Rai Road, Karimpura Chowk, Udam Singh Nagar, Agar Nagar, Balmiki Nagar and Pakhowal Road.

SECTION II

3.2 DATA BASE, VARIABLES AND HYPOTHESIS

3.2.1 Data Collection

This section describes the methods and procedures used for the collection of data. Commensurate with the broad objectives of the study, data was obtained from the secondary as well as the primary sources defined as under:

(1) Secondary Data

Secondary data was obtained to examine the development, growth pattern and trends of health insurance in India. The secondary data was collected from the various annual reports published by Insurance Regulatory and Development Authority of India.
(IRDAI), Handbook on Indian Insurance Statistics (2011-12) published by (IRDAI), various reports of the Insurance Information Bureau of India (IIB), Statistical Abstract of Punjab, Reports of Ministry of Labour and Employment and various annual reports of public and private general insurers. In addition, secondary was also collected from the websites of World Health Organization, Planning Commission of India and Public Health Foundation of India.

(2) Primary Data

The primary data for the present study was obtained with the help of a well structured questionnaire. The questionnaire has been designed on the basis of the constructs defined in the literature. The questionnaire was tested on the pilot basis with the 50 respondents from Amritsar, Jalandhar and Ludhiana. On the basis of the responses given by the respondents in the pilot survey a few changes were made to the preliminary draft of the questionnaire and the final questionnaire (Appendix-I) included five sections defined as under:

(a) Household Characteristics;
(b) Health Insecurities of the Household in the Urban Informal Sector and Socio-Economic Effect of Ill Health on Urban Informal Sector;
(c) Awareness of Health Insurance among Urban Informal Sector;
(d) Demand and Capacity Pay for Health Insurance; and
(e) Problems/ Barriers in the Enrollment of Health Insurance in the Urban Informal Sector.

Although, scale adopted in the present study was identified from the literature but validation of the scale was done with the help of the exploratory factor analysis (EFA) and Cronbach Alpha. The interviews were scheduled over a time span of six months from March 2014 to August 2014. However, the verbal consent was taken from the respondents before commencing in-depth interviews. The confidentiality of their identity and responses given was assured in order to repose full confidence during the interview.

3.2.2 Cost Calculations

Based on the information and details collected from the respondents, outpatient care, inpatient care, chronic disease, direct cost and indirect cost of the healthcare expenditure were calculated as per the following terms:
1) **Outpatient Care:** Outpatient care is a process which does not require admission to a hospital and may be performed outside the premises of a hospital (McIntyre et al., 2006; Gopalan & Das, 2009; Blue Health Intelligence, 2016).

2) **Inpatient Care:** Inpatient care involves the treatment of more serious ailments which requires one or more days of overnight stay in the hospital (John et al., 2009; Kumar et al., 2012).

3) **Chronic Disease:** According to U.S. National Centre for Health Statistics “Chronic disease is the one which lasts for three months or more” (National Health Council, 2014).

4) **Direct Cost:** Direct medical costs include consultation fees, money spent on investigations and drugs (Sauerborn et al., 1996; Segel, 2006; Sam & Philip, 2009; Patrick et al., 2013; Deshmukh et al., 2014).

5) **Indirect Cost:** Indirect costs were measured as the loss of productive working time both for the patient and for the healthy members of the household who have to care for the patient (Russell, 2004; World Economic Forum, 2011; Patrick et al., 2013).

### 3.2.3 Hypothesis

To achieve the specific objectives of the present study, the following hypothesis were formulated:

1) The hypothesis formulated to examine the awareness of health insurance was defined as under:

   **Hₒ₁**: There is no significant association of the demographic variables and awareness of health insurance.

   **Hₒ₂**: There is no significant association of the economic variables and awareness of health insurance.

   **Hₒ₃**: There is no significant association of the regional variables and awareness of health insurance.

   **Hₒ₄**: There is no significant association of the healthcare variables and awareness of health insurance.

   **Hₒ₅**: There is no significant association of the insurance variables and awareness of health insurance.
These hypotheses were tested with the Probit multivariate regression. The explained variable was measured as if aware of health insurance =1 and 0 otherwise, while the description of the explanatory variables was given in Appendix-II.

(2) The hypothesis formulated to examine the demand and capacity to pay for health insurance was defined as under:

\( H_{01} \) : There is no significant association of the demographic variables and willingness to pay for health insurance.

\( H_{01} \) : There is no significant association of the economic variables and willingness to pay for health insurance.

\( H_{03} \) : There is no significant association of the regional variables and willingness to pay for health insurance.

\( H_{04} \) : There is no significant association of the healthcare variables and willingness to pay for health insurance.

\( H_{05} \) : There is no significant association of the insurance variables and willingness to pay for health insurance.

To test the hypothesis, Probit multivariate regression was applied and the explained variable was defined as if willing to pay =1 and 0 otherwise, while description of the explanatory variables was mentioned in Appendix-III.

SECTION III
3.3 METHODOLOGY

The analysis of data was carried out by using percentages, graphs, mean, median, annual growth rate, compound annual growth rate (CAGR), weighted average score (WAS), head count (H), mean gap (G), mean positive gap (MPG), Probit regression, marginal effects, factor analysis, Data Envelopment Analysis (DEA) and Malmquist Productivity Index.

3.3.1 Mean

Mean \(( \bar{X} )\) is known as the most popular and widely used measure of central tendency. It is obtained by taking the sum of all observations comprising a given set of data and dividing the sum by the total number of observations. This method remains
essentially the same whether the data refer to a sample or a finite population (Hooda, 2013).

\[
\overline{X} = \frac{\sum X}{N}
\]

Where,

\[
\sum X = \text{Sum of observations of a series}
\]

\[
N = \text{Number of observations}
\]

3.3.2 Median

The median is generally defined as the middle value in the series when the data is arranged in the ascending or the descending order of the magnitude. The first half of the observations lies above the median and other half lie below the median. The median is the middle value if number of observation is odd and when the numbers of observations are even, the median is the mean of the two middle values. The following formula is used when the number of observations is even (Gupta, 2014):

\[
\text{Median} = \text{Size of} \frac{(N+1)}{2} \text{item}
\]

Where,

\[
N = \text{Number of observations}
\]

3.3.3 Annual Growth Rate

The percent change from one period to another period was calculated with the following formula (Mawson, 2002).

\[
PR = \left(\frac{V_{\text{present}} - V_{\text{past}}}{V_{\text{past}}}\right) \times 100
\]

Where:

\[
PR = \text{Percent rate}
\]

\[
V_{\text{present}} = \text{Present value}
\]

\[
V_{\text{past}} = \text{Past value}
\]

3.3.4 Compound Annual Growth Rate (CAGR)

The compound annual growth rate was worked out by using the following equation (Bajpai, 2015).

\[
Y_t = ab^t e^{ut}
\]
Transforming the above equation in log linear form:

\[ \log Y_t = \log a + t \log b + u_t \]

Where:

- \( Y_t \) is the value of dependent variables in years
- \( t \) is the trend variable
- \( u_t \) is disturbance term
- \( a \) and \( b \) are constants.

From the estimated value of regression co-efficient ‘b’, the compound rate of growth ‘r’ was calculated as follows:

\[ r = \text{Antilog} \left( \hat{\beta} - 1 \right) \]

Where, \( \hat{\beta} \) estimated value of b

### 3.3.5 Weighted Average Score (WAS)

The weighted average score was calculated to identify the important socio-economic implications of illness (outpatient care, inpatient care and chronic disease) on the household. The informal sector workers were asked to respond to the various socio-economic implications on five-point likert scale to measure the extent of the agreement or disagreement of the consequences. The weighted average score were calculated by assigning weight as Strongly Agree (5), Agree (4), Neutral (3), Disagree (2) and Strongly Disagree (1). In the present study weighted average score was obtained on the basis of the frequency of rating of each statement and the average weighted scores for each statement were calculated as under (Grela, 2013):

\[ \text{WAS} = \frac{\sum w_i x_i}{\sum w_i} \]

Where,

- \( x_1, x_2, x_3 \ldots \ldots x_n \) are the \( n \) number of respondents
- \( w_1, w_2, w_3 \ldots \ldots w_n \) are the weights.

### 3.3.6 Head Count (H)

The incidence of catastrophic healthcare expenditure was calculated from the fraction of a sample with the health care costs as a share of total (or non-food) expenditure exceeding the chosen threshold level (Berki, 1986; Russell, 2004; Galarraga
et al., 2008; Bredenkam et al., 2011). To estimate headcount (H) data was required at the household (HH) containing information on both the health care payments and living standards (consumption expenditure, say I). The living standard was estimated by the “ability/capacity to pay” variable (Y) was defined as under:

\[ Y = I - D(I), \]

Where, D (I) represents necessary or the non-discretionary expenditure on items such as food and I represents the consumption expenditure. The present study measured Y=non-food expenditure as the household capacity to pay. The sample of HH is said to incur catastrophic payments on healthcare when the fraction H/I or H/Y exceeds a pre-specified threshold level, say Z. Generally, financial catastrophic healthcare expenditure occurs with the health care payments at or exceeding 40 percent of the household capacity to pay in any year (Su et al., 2006; Gopalan & Das, 2009; Bredenkam et al., 2011; Ghosh, 2011). However, for a comparative picture threshold levels like 5 percent, 10 percent, 20 percent, 30 percent and 40 percent were calculated. This sample of individuals represents the catastrophic payment headcount. An indicator E, was defined as \( E = 1 \) if \( H/I_i > Z \) and zero otherwise. The catastrophic healthcare payment headcount was defined as under:

\[ H = \frac{1}{N} \sum_{i=1}^{n} E_i \]

Where,

\[ N = \text{Number of observations} \]

3.3.7 Mean Gap (G)

The catastrophic healthcare payment gap examined the average degree by which payments as a proportion of income exceeds the threshold level, Z (Pradhan & Prescott, 2002; Ranson, 2002; Xu et al., 2003; Gopalan & Das, 2009; Mondal et al., 2010; Ghosh, 2011). The excess or overshoot was defined as the amount by which the payment fraction \( (H/I_i) \) exceeds the catastrophic threshold Z.

\[ O_i = E_i \left( \frac{H_i}{I_i} - Z \right) \]

Where,

\( O_i \) is the overshoot at a particular threshold level.

\( H \) is the head count

\( I \) represents the consumption expenditure

\( Z \) represents threshold level
While, the catastrophic payment gap was measured by:

\[ G = \frac{1}{N} \sum_{i=1}^{n} O_i \]

Where,

- \( N \) = Number of observations

However, H only examines the incidence of catastrophic healthcare expenditure while, G measures the intensity of the occurrence.

### 3.3.8 Mean Positive Gap (MPG)

The mean positive gap (MPG) also known as the mean positive overshoot (MPO) highlighted the mean out-of-pocket payments for health care in excess of the threshold over all households exceeding the threshold (Jogelkar, 2008). Thus, MPG is commonly used as an indicator of mean overshoot among households with the catastrophic health care payments at a particular threshold level. It amplifies the intensity of the catastrophic health care payments (Wyszewianski, 1986; Russell, 2004; Xu et al., 2007). The mean positive gap was calculated as under:

\[ MPG = \frac{G}{H} \]

Where,

- \( G \) is the mean gap,
- \( H \) is the head count.

### 3.3.9 Probit Regression Model

Probit regression assumes only two values 0 and 1 for the variable \( Y \) (dependent variable), there is a latent, unobserved continuous variable \( Y^* \) that determines the value of \( Y \) (Chambers & Cox, 1967; Naglar, 1994; Bendig & Arun, 2011; Giesbert et al., 2011; Greene, 2014). \( Y^* \) has been specified as under:

\[ Y_i^* = \beta_0 + B_1 x_{i1} + B_2 x_{i2} + \ldots + B_k x_{ik} + u_i \]  

(1)

Where,

- \( Y_i = 1 \) if \( Y_i^* > 0 \)
- \( Y_i = 0 \), otherwise.

Where, \( x_1, x_2, \ldots, x_k \) (independent variables) represent the matrix of random variables, and \( u \) represents a random disturbance term.

Now from equation (1)
Pr \((Y_i = 1) = Pr (\beta_0 + B_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i > 0) \) \hspace{1cm} (2)

Rearranging terms,

\[
Pr \((Y_i = 1) = Pr (u_i > - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}))
\]

\[
= 1 - Pr(u_i < - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}))
\]

\[
= 1 - F (- (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})) \hspace{1cm} (3)
\]

Where, \(F\) is the cumulative density function of the variable \(u\) and based on the assumption that \(u\) is normally distributed, we have

\[
Pr \((Y_i = 1) = 1 - \Phi (- (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}))
\]

\[
= 1 - \Phi (- X_i \beta)
\]

\[
= \Phi (X_i \beta) \hspace{1cm} (4)
\]

Where, \(\Phi\) represents the cumulative normal distribution function.

The estimation of the coefficients \((\beta_\prime s)\) can be made with the help of maximum likelihood techniques. The use of maximum likelihood techniques also facilitates the estimation of the corresponding standard errors that are asymptotically efficient. However, these estimates cannot be interpreted in the same manner as the normal regression coefficients are estimated (Hair et al., 2010). These coefficient give the impact of the independent variables on the latent variable \(Y^*\), not \(Y\) itself. To transfer \(Y^*\) into a probability estimate for \(Y\) we compute the cumulative normal of \(Y^*\). Because of this transformation there is no linear relationship between the coefficients and \(Pr \((Y_i = 1)\). Hence, the change in \(Pr \((Y_i = 1)\) caused by a given change in \(x_{ji}\) will depend upon the value of all of the other \(x_s\) and their corresponding coefficients, or more precisely on the value of the sum \(X_i \beta\), as well as the change in \(x_{ji}\). Before proceeding to the data analysis it is important to realize that the above need not dissuade us from attempting to identify true interactive effects when using the Probit model. Our estimates of \(Pr \((Y_i=1)\) are based on a transformation of an underlying model generating the latent variable \(Y^*\). \(Y^*\) is the variable of interest, though it is unobserved. Since the model generating \(Y^*\) is linear, and thus not contaminated by assumed interactive effects, we can use this underlying model to perform tests for substantive interactive relationships among independent variables.
Therefore it is required to add explicit interactive terms to the model specification (Naglar, 1994; Giesbert et al., 2011; Greene, 2014). However, Brooks (2008) mentioned a method to make the interpretation of coefficients of the Probit regression in more meaningful form namely marginal effects. The marginal effect tends to explain a one unit change of any variable on the probability of \( Y_i = 1 \), but it is different for every person. Therefore, the model coefficients are scaled at its mean and, subsequently, can be interpreted as the marginal effect a one unit change of the independent variable from the sample mean ceteris paribus has on the probability of \( Y = 1 \). Interpreting the marginal effects coefficients, it needs to be distinct for discrete and continuous variables \( x \). For continuous variables, the coefficient provides the percent change an infinitesimal alteration of \( x \) has on the probability that \( Y = 1 \). For discrete variables, however, the coefficient denotes the change in probability that \( Y = 1 \) if the discrete variable switches from 0 to 1. In order to identify the variables which were statistically significant, \( z \)-statistic was calculated. While, \( z \)-statistic is a standardized value which can be calculated as the raw score of \( x \) minus the population mean divided by the population standard deviation (Naglar, 1994; Monheit & Vistnes, 2008).

### 3.3.10 Factor Analysis

Factor analysis is a general name denoting a class of procedures chiefly applied for data reduction and summarization. Most of the times in research, there may be a large number of correlated variables which must be reduced to a manageable level. Factor analysis is generally used to examine and represent the relationship among the set of interrelated variables. In factor analysis, a given set of \( n \) variables grouped into \( p \) number of groups called ‘Factors’ which are less in number than the set of original variables. The variables within a group (factor/strata) are of the same nature or are of complementary with respect to the phenomenon under study but between two groups ‘Factors’ variables are independent. Thus factors \( F_i \) and \( F_j \) are orthogonal.

The Factor Analysis used in the present study is given as under:

\[
X = LF + U
\]

Where, \( X \) is vector of all the original variables.

\[
X' = [X_1, X_2, X_3, \ldots \ldots, X_n]
\]

\( F \) is vector of ‘Factors’ derived
\[ F' = [F_1, F_2, F_3 \ldots \ldots \ldots F_p] \]

U is vector of error terms.

\[ U' = [E_1, E_2, E_3 \ldots \ldots \ldots E_n] \]

L is matrix of Factor Loading (Loading Coefficient Matrix)

\[
L = \begin{pmatrix}
a_{11} & a_{12} & a_{13} & \ldots & \ldots & a_{1p} \\
a_{21} & a_{22} & a_{23} & \ldots & \ldots & a_{2p} \\
a_{31} & a_{32} & a_{33} & \ldots & \ldots & a_{3p} \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
a_{n1} & a_{n2} & a_{n3} & \ldots & \ldots & a_{np} \\
\end{pmatrix}
\]

The coefficient (Factor loading) \( a_{ij} \) belongs to \( i^{\text{th}} \) variable and \( j^{\text{th}} \) Factor which is similar to simple correlation coefficient and shows the extent to which variable \( X_i \) is related to \( F_j \) Factor. “A salient loading is one which is sufficiently high to assume that a relationship exists between the variable and the factor. In addition, it usually means that relationship is high enough so that the variable can aid in interpreting the factor and vice-versa” (Gorsuch, 1974).

The sum of the square of the factor loadings of \( X_i \) original variables under the derived \( p \) Factors is called the communalities for \( X_i \) variables.

\[
(a_{i1})^2 + (a_{i2})^2 + (a_{i3})^2 + \ldots \ldots \ldots (a_{ip})^2 = (C_i)^2
\]

Communality in factor analysis is similar to the \( R^2 \) in the regression analysis and it explains the extent to which the derived factors explain the \( i^{\text{th}} \) variables. The derived communality value generally should be equal to or larger (more than 60 per cent) to be sure that each variable has been explained well. The communality of a variable is that proportion of its variance which can be accounted for by the common factors. (Lindeman et al., 1980).

The Principal Component Analysis (factor analysis) produces components (factors) in descending order of their importance and factor loadings which explained the relative importance of different variables in explaining variance in the phenomenon. Most of the studies using ‘Factor Analysis’ adopted ‘First Principal Component’ as guiding principle for determining individual indicator weights (Gorsuch, 1974; Lindeman et al.,
The present study an effort was made to study all the ‘Principal Components’ (factors derived) to determine relative weights of selected variables so as to reflect maximum possible variations. The method for determining the relative weights for the variables is explained below:

\[ W_i = F_{ik} \lambda_k \]

- \( W_i \): is weight of \( i^{th} \) variable.
- \( F_{ik} \): is factor loading of \( i^{th} \) variable and \( k^{th} \) factor which reflects the highest correlation between variable \( (X_i) \) and factor \( (F_k) \);
- \( \lambda_k \): is variation explained by \( k^{th} \) factor.

The present study identified 36 correlated variables (statements) which can restrict the enrollment of the informal sector workers towards the health insurance. The responses of the respondents were obtained on the basis of five-point likert scale such as strongly agree (5), Agree (4), Neutral (3), Disagree (2) and Strongly Disagree (1). Before running the factor analysis, the reliability of the scale items was measured with the help of Cronbach’s alpha, which assesses the internal consistency of entire scale. Thereafter, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett’s test of sphericity were performed. The higher values of KMO (between 0.5 and 1.0) indicates that factor analysis is appropriate, however values below 0.5 imply that factor analysis may not be appropriate (Malhotra, 2007; Chawla & Sondhi; 2011). An Eigen value greater than 1 criterion was employed to determine the number of factors. To determine the minimum loading necessary to include an item in its respective constructs, Hair et al. (2010) suggested that variables with loading greater than 0.30 is considered significant, loading greater than 0.40 more important, and loading 0.50 or greater are very significant. In the present study the factor loading of 0.30 or greater was considered. In order to obtain more interpretable results varimax rotation was used to rotate the solution. This caused the loadings to be distributed among the selected factors making it easier to interpret results (Malhotra, 2007; Chawla & Sondhi; 2011).

**3.3.11 Data Envelopment Analysis (DEA)**

Data Envelopment Analysis was applied to examine the efficiency and productivity of the health insurance business of public and private general insurance
companies from the year 2006-07 to 2013-14. The year 2006-07 was been taken as the base year due to the fact that during this year the first standalone health insurance company was introduced in Indian market and this is the exclusive year in which the maximum number of the general insurance companies started their health insurance business. The Data Envelopment Analysis was applied on four public and nine private general insurance companies of which two were the standalone health insurance companies. The public general insurance companies were New India Assurance Company Limited, Oriental Insurance Company Limited, National Insurance Company Limited, and United India Insurance Company Limited. The private general insurance companies were Royal Sundram Alliance Insurance Company Limited, Tata AIG General Insurance Company Limited, Reliance General Insurance Company Limited, IFFCO Tokio General Insurance Company Limited, ICICI Lombard General Insurance Company Limited, Bajaj Allianz General Insurance Company Limited, HDFC ERGO General Insurance Company Limited of which two are standalone health insurance companies (deals exclusively in the health insurance business) i.e. Star Health and Allied Insurance Company Limited and Apollo Munich Health Insurance Company Limited. An input-output model was applied to examine the technical efficiency, pure technical efficiency and scale efficiency of the health insurance business of the public and private general insurance companies. However, improvement space and the improvement direction was calculated for the public and private general insurance companies on the basis of the on the basis of the technical efficiency scores. The result of Data Envelopment Analysis (DEA) and Malmquist Total Factor Productivity Analysis (MTFA) depends heavily on the selection of input and output. From the literature review, it came into notice that researchers had a general agreement on the selection of input and output variables. Following the literature Mansor & Radam (2000), Saad et al. (2006), Qiu & Chen (2006), Sinha & Chatterjee, (2008), Bawa & Ruchita (2011), Mathur & Paul (2014) and Nandi (2014) an indicator of input was taken as the commission paid. However, the selection of output variable as the net premium was consistent with Mansor & Radam (2000), Saad et. al. (2006), Mathur & Paul (2014) and Nandi (2014). Furthermore, the selection of the input and output variables were in accordance with the
assumption of DEA that number of decisions making units (DMUs) should be three times the number of inputs and outputs selected.

The technique of Data Envelopment Analysis (DEA) was developed by Charnes et al. (1978) based on the Farrell’s work (Farrell, 1957). Data Envelopment Analysis is a non-parametric technique to measure the relative efficiency of a set of similar units usually denoted as the decision-making units (DMUs). Initially, DEA was designed to examine the relative efficiency of the not-for-profit organizations such as the hospitals and schools, however, gradually it was extended to cover for-profit organizations as well (Rai, 1996). Using linear programming technique, the various DEA models intended to provide efficiency scores under different orientations and assumptions of returns-to-scale. DEA helps to examine the overall technical efficiency (OTE) and decomposes it into two mutually exclusive and non-additive components, namely, pure technical efficiency (PTE) and scale efficiency (SE). In DEA, there is no need to select a priori functional form relating to inputs and outputs like Cobb-Douglas and Translog production/cost function (Banker, 1984; Cummins & Weiss, 1993). In DEA, technical efficiency (TE) can be viewed from two perspectives. Firstly, input-oriented TE focuses on the possibility of reducing inputs to produce given output levels. Secondly, output-oriented TE considers the possible expansion in outputs for a given set of input quantities (Cummins & Weiss, 1993). A measure of TE for a DMUo can be explained as under:

\[ \theta_0^{output} = \frac{\text{Actual output}_0}{\text{Maximum possible output}_0} \text{ in output-oriented context, or} \]
\[ \theta_0^{input} = \frac{\text{Minimum possible input}_0}{\text{Actual input}_0} \text{ in input-oriented context.} \]

To quantify a measure of the technical efficiency need is to identify the divergence between actual production and production on the boundary of the feasible production set (Banker, 1984; Hwang & Gao, 2003). This set summarizes all technological possibilities of transforming inputs into outputs which are available to the organization. A DMU is technically inefficient if production occurs within the interior of this production set. A measure of scale efficiency (SE) can be obtained by comparing TE measures derived under the assumptions of constant returns-to-scale (CRS) and variable returns-to-scale (VRS). As noted above, the TE measure corresponding to CRS assumption represents overall technical efficiency (OTE) which measures inefficiencies due to the input/output configuration as well as the size of operations. The efficiency
measure corresponding to VRS assumption represents pure technical efficiency (PTE) which measures inefficiencies due to only managerial underperformance. The relationship, SE=OTE/PTE helps to provide scale efficiency and for one output and one-input scenario, the derivation of the concepts of technical, pure technical, and scale efficiency under DEA approach is illustrated in the following figure 3.3.1.

![Figure 3.3.1: Technical Efficiency (TE), Pure Technical Efficiency (PTE) and Scale Efficiency (SE) Measures](source: Kumar & Gualti (2008))

Figure 3.3.1, exhibits two efficient frontiers where one assumes CRS (shown by line OO) and the other assumes VRS (shown by line segment PABCQ). Projecting the inefficient DMU D onto VRS efficient frontier (point E) by minimizing input X while holding output Y constant (i.e., input-orientation), PTE for DMU D is defined as \( \frac{X_E}{X_D} \). Similarly, if the optimization mode changes to that of output maximization, PTE for firm D is \( \frac{Y_D}{Y_H} \). Focusing on the CRS efficient frontier, DMU D is projecting onto point F, where the input-oriented OTE measure is defined by \( \frac{X_F}{X_D} \). The output oriented OTE measure is similarly defined as \( \frac{Y_D}{Y_L} \). However, given that the slope of CRS efficient frontier equals to 1, then \( \frac{X_F}{X_D} = \frac{Y_D}{Y_L} \) i.e., orientation does not change OTE scores. Extending the above illustration to scale efficiency, input and output-oriented scale efficiency measures are defined as \( \frac{X_F}{X_E} \) and \( \frac{Y_H}{Y_L} \) respectively. Increasing returns-to-
scale (IRS) imply that the DMU can gain efficiency by increasing the production of Y (which generally occurs when producing on the PAB of VRS efficient frontier), while decreasing returns-to-scale (DRS) imply that a reduction in scale increases efficiency (which occurs on the portion BCQ of VRS efficient frontier). If one is producing optimally, then, there is no efficiency gain by changing the scale of production. This occurs when firm operate at the point B where the two frontiers are tangent i.e., OTE=PTE.

The graphical representation (Figure 3.3.1) of the technical, pure technical and scale efficiency measures can be reframed in terms of linear programming models that can be used to examine the efficiency of individual DMUs using actual data on input and output variables. There were several mathematical programming models which were developed and proposed in the literature (Banker, 1984; Charnes et al., 1994; Cooper et al., 2007). Essentially, each of these models seeks to establish which of n DMUs determine the best practice or efficient frontier. The geometry of this frontier is prescribed by the specific DEA model employed in the present study. In the present study CCR model, named after Charnes, Cooper, and Rhodes (1978) and BCC model named after Banker, Charnes and Cooper (1984) were used to obtain the efficiency measures under the CRS and VRS assumptions, respectively. The efficiency measures obtained from the CCR model are commonly known as the overall technical efficiency (OTE) scores and are confounded by scale efficiencies. The efficiency obtained with the help of the BCC model is popularly known as pure technical efficiency (PTE) scores and devoid of scale efficiency effects. The scale efficiency (SE) for each DMU can be obtained by a ratio of OTE score to PTE score (SE=OTE/PTE). The formal notations of input-oriented DEA models for measuring TE scores for DMUo, applied in the present study under the different scale assumptions were defined as follows:

\[ \text{(i)} \]
\[ \min \theta_0, \lambda_1, \lambda_2, \ldots, \lambda_n, S_j^-, S_j^+ \]
\[ \text{Subject to} \]
\[ \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta_o x_{io} \]
\[ \sum_{i=1}^{m} S_j^- + \sum_{j=1}^{n} S_j^+ \]
(iii) \[ \sum_{j=1}^{n} \lambda_{j} y_{rj} - s_{i}^{+} = y_{ro} \]

(iv) \[ s_{i}^{-}, s_{i}^{+} \geq 0 \quad (i = 1, \ldots, m; r = 1, \ldots, s) \]

(v) \[ \lambda_{j} \geq 0, \quad \text{if, constant returns to scale} \]

(vi) \[ \sum_{j=1}^{n} \lambda_{j} = 1, \quad \text{if, variable returns to scale} \]

Where, 
- \( X_{io} \) = amount of input i used by DMUo,
- \( Y_{ro} \) = amount of output i used by DMUo,
- \( m \) = the number of outputs,
- \( s \) = the number of DMUs, and
- \( n \) = the number of DMUs, and
- \( \varepsilon \) = a small positive number.

The solution to the equation (i) is interpreted as the largest contraction of the DMU o’s input that can be carried out, given that DMU o will stay within the reference technology. The restrictions (ii) and (iii) form the convex reference technology in the DEA. The restriction (iv) restricts the input slack \((S_{i}^{-})\) and output slack \((S_{i}^{+})\) variables to be non-negative. The restriction v) limits the intensity variables to be non-negative. The model involving (i)–(v) is known as envelopment form of CCR model and provides Farrell’s input-oriented TE measure under the assumption of constant returns-to-scale. The measure of efficiency provided by CCR model is known as the overall technical efficiency (OTE) and denoted as \( \theta_{0}^{CCR} \). The last restriction imposed variable returns-to-scale assumption on the reference technology. The model involving (i)–(iv) and (vi) is known as BCC model and provides Farrell’s input-oriented TE measure under the assumption of variable returns-to-scale. The measure of efficiency provided by BCC model is known as the pure technical efficiency (PTE) and denoted as \( \theta_{0}^{BCC} \). The ratio \( \theta_{0}^{CCR} / \theta_{0}^{BCC} \) provides a measure of scale efficiency (SE). The all of the aforementioned efficiency measures are bounded between one and zero. The measure of scale efficiency (SE) does not indicate whether the DMU in question is operating in the area of increasing or decreasing returns-to-scale. The nature of returns-to-scale can be determined from the
magnitude of optimal \( \sum_{j=1}^n \lambda_j^* \) in the CCR model (Banker, 1984; Seiford & Zhu, 1999) listed following three cases:

(i) If \( \sum_{j=1}^n \lambda_j^* = 1 \) in any alternate optima, then CRS prevail on DMU o;

(ii) If \( \sum_{j=1}^n \lambda_j^* < 1 \) in any alternate optima, then IRS prevail on DMU o; and

(iii) If \( \sum_{j=1}^n \lambda_j^* > 1 \) in any alternate optima, then DRS prevail on DMU o

The CCR and BCC models need to be solved n times, once for each DMU to obtain the optimal values for

\[ \theta_o, \lambda_1, \lambda_2, ..., \lambda_n, s_i^-, s_i^+ \] (i.e., \( \theta_o^*, \lambda_1^*, \lambda_2^*, ..., \lambda_n^*, s_i^{-*}, s_i^{+*} \)). The interpretation of the results of above models can be explained as under:

(a) If \( \theta_0^* = 1 \), then DMU under evaluation is a frontier point, i.e., there are no other DMUs that are operating more efficiently than this DMU. Otherwise, if \( \theta_0^* < 1 \), then the DMU under evaluation is inefficient, i.e., this DMU can either increase its output levels or decrease its input levels.

(b) If left-hand side of the constraint (ii) and (iii) is called the ‘Reference Set’, and the right-hand side represents a specific DMU under evaluation. The non-zero optimal \( \lambda_j^* \) represents the benchmarks for a specific DMU under evaluation. The reference set provides coefficients \( (\lambda_j^*) \) for defining hypothetical efficient DMU.

(c) The efficient targets for inputs and outputs can be obtained as \( \hat{x}_{io} = \theta_0^* x_{io} - s_i^{-*} \) and \( \hat{y}_{ro} = y_{ro} + s_i^{+*} \), respectively. These efficiency targets show how inputs can be decreased and outputs increase to make the DMU under evaluation efficient.

### 3.3.12 Malmquist Productivity Index

Data Envelopment Analysis helps to measure the productivity change of health insurance business of general insurance companies by using Malmquist Index summary. The Malmquist Productivity Index was applied to examine the Efficiency Change (EC),
Technological Change (TC), Pure Technical Efficiency Change (PTEC), Scale Efficiency Change (SEC) and Total Factor Productivity Change (TFPC) of health insurance business of the public and private general insurance companies. Overall, it is provided that TFPC which comprised of EC, TC, PTEC and SEC. The total factor productivity growth measures the change in the production frontier and how the current frontier relates to the firm frontier over the time period. The total factor productivity change involved two major components namely technology change (technical progress) and efficiency change. However, technology change was explained by a shift in the production frontier while efficiency change was based upon a firm’s efficiency relative to the past and the future frontiers. The output-orientated productivity change measures were used as an output distance function which addressed the maximal proportional expansion feasible without altering the input quantities (Coelli et al., 1998). To estimate the technical efficiency changes and technical changes over a period of time, the decomposed Malmquist Productivity Index was used in the present study. Furthermore, Caves et al. (1982) proposed that output-based Malmquist Productivity Index between time periods t and (t + 1) can be defined as:

\[
M_{t,t+1}(y', y^{t+1}, x', x^{t+1}) = \left[ \frac{D' (y^{t+1}, x^{t+1})}{D' (y', x')} \times \frac{D^{t+1} (y^{t+1}, x^{t+1})}{D^{t+1} (y', x')} \right]^{1/2}
\]

Where, D represents the distance function and the value of M is the Malmquist Productivity Index. The first ratio represents the period t of Malmquist Index. It measures productivity change from period t to period (t+1) using period t technology as a benchmark. The second ratio is the period (t + 1) Malmquist Index and measures productivity change from period t to period (t +1) using period (t + 1) technology as a benchmark. A value of M greater than one (M >1) denotes productivity growth, while a value less than one (M < 1) indicates productivity decline and M= 1 corresponds to stagnation. According to Fare et al. (1994) the output-based Malmquist Productivity Index between time periods t and (t + 1) can be decomposed into two components, which is an equivalent of index (1), as Fare et al., (1994), Coelli (1996), Grifell et al., (1996, 1997):
In the figure 3.3.2, the calculation of Malmquist Productivity Index is illustrated, where a single output $y$ is produced using a single input $x$. It was based on constant returns to scale (CRS) technology. The firm A produces at the point A1 in the first period and at the point A2 in the second period. The firm A is technically inefficient in the first period as the point A1 is below the frontier for that period. In the second period, the point A2 is on the frontier and thereby firm A is technically efficient. The technical change includes a time component and involves advances in technology, which is represented by an upward shift in the production frontier from the first period to the second period.

Using equation (2) it is:

$$ M_{t,t+1}(y', y'^{t+1}, x', x'^{t+1}) = \frac{D'^{t+1}(y', x')}{D'(y', x')} \left[ \frac{D'(y', x')}{D'^{t+1}(y', x')} \times \frac{D'(y', x')}{D'^{t+1}(y', x')} \right]^{1/2} $$

(2)

Where, the value of $M$ is the Malmquist Productivity Index between two time periods. It is easy to calculate that the value of $M$ is greater than one ($M > 1$), which implies productivity growth. In equation (3) the terms EFFCH and TECHCH are also greater than one (EFFCH > 1 and TECHCH > 1). Therefore, the firm A experienced the
positive technical efficiency change and technological change from one period to another. To construct the Malmquist Index for adjacent periods, it was required to calculate four different distance functions $D_t(y_t, x_t)$, $D_t(y_{t+1}, x_{t+1})$, $D_{t+1}(y_t, x_t)$ and $D_{t+1}(y_{t+1}, x_{t+1})$. There are many different methods that could be used to measure the distance function, which makes up the Malmquist Productivity Index. These required distance functions can be calculated using either mathematical programming or econometric techniques. The DEAP computer program was used to construct Malmquist Indices.

**SECTION IV**

**3.4 LIMITATIONS OF THE STUDY**

- The present study was restricted to three districts of Punjab i.e. Amritsar, Jalandhar and Ludhiana and the sample size was limited to 630 workers employed in the informal sector;
- Due to cultural, geographical or socio-economic disparities the results are mere an indication and may or may not be applicable to the whole of the urban informal sector of India;
- Dependent variable used in this study was dichotomous variable (a) aware of health insurance or not and (b) willing to pay for the health insurance or not. It is sometimes misleading to assume that people always make dichotomous choices;
- Although, utmost care was taken while selecting input and output variables for the performance evaluation of health insurance business, still the inclusion of some other variables might influence the overall results; and
- The study was confined to some selected general insurance companies dealing with health insurance business, while the inclusion of other would provide more appropriate results.