Chapter 7

Determinants of employment and employability in Manipur

7.1 Introduction

Having studied the socio economic structure of Manipur in chapter 6, in this chapter an attempt is made to analyze the determinants of employment and employability. There are number of variables which play a significant role in reasoning out the probability of a respondent's employment status. Further analysis will help clarifying the picture of employment, in general and that of young educated youth in Manipur.

In the earlier chapter it has been found that employment status of a respondent was statistically significant with gender, area of residence, educational qualification of the respondent. For variables like head of the family's educational qualification, type of family, religion were not found to be significant statistically. These variables individually and collectively determine the employment status and their employability. The socio-economic background of the respondents too play a significant role. The employability of the respondent is a combination of factors pertaining to the individual, their family background, social status and their networks in addition to the economic and geographical factors of the region under consideration. Though all these factors play a role, aggregating them is difficult. In this chapter an attempt is made to combine them.

The chapter is organized as following. Second section of the chapter proposes various hypothesis and role of potential determinants of employment and also wage differential for those who are employed. The second section presents the data and methodology followed in the present chapter and the third section presents the results and discussion. The last section ends with the summary of the main findings of the chapter.

7.2 The determinants and proposed hypothesis

The determinants of employment can be broadly classified into social factors, economic factors of the respondent and the region in which analysis is taken up. In addition to these the factors specific to the respondent and their family play an important role. In the

following paragraph the relevant variables and their expected impact are presented. These variables impact is tested in the succeeding sections.

The major Hypothesis being tested are

i. Employment status is positively influenced by the social status of the respondent's family.
ii. Employment status of the respondents is positively influenced by their age and education.
iii. Salary of the respondent is positively influenced by the age, education and experience.
iv. Employability of the respondents is dependent on their educational level.
v. Employability of the respondent is dependent on the age of the respondent.
vi. Employability status is dependent on the employability score of the respondent.

To test these hypothesis the author makes use of multiple regression and correlation analysis as well as logistic regression or logit model. The study makes use of Principal component analysis (PCA) to generate Employability Index. In addition, χ^2, t and F statistics are used for testing various hypothesis.

7.2.1 The social background of the respondents and the employment status

The respondent's social background is a significant set of factors determining the employment status. These factors reflect the capability of individuals to seek job in the market. They reflect the networking, waiting capacity and many other dimensions. The following factors are analyzed to understand the employment scenario in Manipur.

(a) The family size of the respondent

Family size of a household has an important role as a determinant in well being (Lanjouw and Ravallion,1995; Meenakshi and Ray, 2002). In the earlier stages of family formation as the household size increases the number of dependent children increases and leading to high dependency ratio. The burden of the family reduces once the children start working. (Lanjouw and Ravallion, 1995). If the head of the family is low income earner it gets harder to maintain the lifestyle so this makes the elder children of the family

especially male to start working at earlier age foregoing their education. So the urgency of getting a job for a respondent is higher, larger the family size.

(b) Age and Gender of the head of the family

Age and gender play an important role in choices of occupation, earning capacity, cohesiveness of members and management of the household economy (Gaiha,1998). Household with lower age of the head of the family (i.e. in the working age) are likely to have a good earning capacity and earning if they are employed, so an urgency of getting a job for the respondent might be lower in that case. If the head of the family is retired and aged then the urgency of getting a job can be higher for the respondent. Along with age, gender plays an equally important role as determinant of employment. In Manipur, predominant role of male members are well accepted in economic as well as socio-cultural circle and the family headship passes on to the eldest son and so. In exceptional cases where male head has expired or left the family the female takes the responsibility of being the head of the family. In Manipur, very few women participate in industrial activities, with the exception of Imphal (capital) where the participation of women in non-household activities was the highest. In rural areas, women are confined to lowly paid labour intensive jobs like paddy sowing, while more skilled jobs like ploughing and harvesting are handled by men. (MDONER,2005). The chances of female to be employed and earning high income is also less likely so the urgency of getting employed for the respondent can be higher in that case.

(c) Education of the head of the family

Education is widely accepted as one of the most important component of human development and income earnings and escaping the danger of falling into poverty (Thompson & Mcdowell,1994; Rodriguez & Smith,1994; Gaiha,1998). In fact, higher education makes people get more rewarding job and higher pay which can increase the productivity of the members as well as the family (Belinkisi, Gungor and Tapsin, 2015). So having a head of the family with high education has a greater chance of getting a good job with better wages. The young in the family can pursue higher education for themselves and wait for better well suited job.

(d) Occupation of the head of the family

The role of occupation of the head of the family is a significant attribute of the household and the respondents urgency of job. The choice occupation of a person is dependent on many factors such as the place of residence, wealth, educational background etc. These factors can influence the choice occupation of the person into high paying highly productive job or low paying low productive job (Estudillo et al. 2013). If the occupation of the head of the family is falling into the highly productive and high earning job then it is likely that the respondent is not in the urgency of getting a job so he or she can wait for the well suited job. The indicator for the type of occupation that is used is the informal sector job and the formal sector job. The formal sector job provided by the public as well as the private companies, being more secured and higher pay than the informal sector job which are of lower pay and less secured (ILO, 1973).

(e) Location of residence of the respondent

Geographical location of a region is one of the contributing factor in well being of individuals or family (Jalan and Ravallion, 1997). Urban areas everywhere in the world are not only preferred for the provision of basic health amenities, proper education, better connectivity and transportation but also better choice for industrial and service sector hubs which eventually provide a good job market. Manipur is a state with lesser industrial sector, however service sector is booming and the job market is larger compared to rural areas (MDONER,2010), therefore the chances of getting a job is definitely higher in urban areas.

7.2.2 Individual characteristics of the respondents

The respondent's individual characteristics (as stated in chapter 6, 6.3.2) mainly reflect their age, education, skill, aptitude, gender and their influence on employability and job search. Employment status changes overtime but is impacted by the employability of the individual.

(a) Age of the respondent

The respondents whose ages are in between 15 to 25 are less likely to find a well paid secured job and the chances of the getting a job is less because of many factors influencing the demand of labour in the market skill and experience are some of the important factors which are acquired overtime. Ages above 25 are likely to get a job more easily if adequately equipped. The chances are that they might end up being self employed to run the family with age comes out of need or self drive.

(b) Gender

In Manipur especially the female labour force participation is as low as 854 per 1000 persons and the male participation is 994 per 1000 persons. In rural areas the females are less educated and they are not considered as bread winner for the family (MDONER,2001), therefore they end up being a housewife and out of labour force. However if they are highly educated then they look for well paid secured and well suited job in the process, wait for suitable opportunity. This case is similar in urban areas too however the level of education is higher in most cases.

For male members it is likely that the urgency of finding a job is much higher as they enter working age if the family's economic condition is poor. So larger proportion of the male respondents were employed.

(c) Education of the respondent

Higher educated respondent have a greater chance of getting a high paying job as well as more secured job. In India where the educated unemployment is high, there is a greater chance that they are still unemployed looking for the well suited job. The respondent who has lesser years of schooling and lower qualification is more likely to end up with low paying jobs or in informal sectors. The factors influencing the years of education and schooling are determined by family's economic as well as social background and this determines the employment status of the individual.

The variables selected either social background or individual characteristics of respondents as determinants of employment are already discussed in chapter 6.

7.3 Methodology

There are number of factors determining employment status of the respondents. As mentioned earlier these could be classified as social-economic, individual, family, economic, geographical factors. The employed individual and unemployed individual may be living in similar conditions with the outcomes of income and employment status that are diverse in nature.

The variables under consideration can be analyzed individually e.g. for gender, education, age etc. But in the market, the outcomes are effect of all the variables and their interactions. To analyze the outcomes, appropriate methodologies must be used. Two techniques are prominently used in economic literature. (I) Regression analysis (a) Multiple regression (b) logistic regression (II) Principle component analysis.

Regression analysis is the most widely used technique in economic literature. The appropriate function form is a multivariate function. The independent variables being determinants of employment, this can lead to the problem of multicollinearity. If the dependent variable is income of the respondent then the cases taken for analysis will be only employed. However the respondents can be both employed and unemployed. This is a binary dependent variable. To analyze this data set, one can use logistic function. As the employment status is determined by many variables each with its own merit, none of them can be dropped. To avoid the problem of multicollinearity and extract maximum information, the PCA is made use of. Using this method Employability Index is generated which is used for further analysis.

To examine the determinants of employment, the one of the techniques that has been used is the Principal Component Analysis (PCA). PCA is a multivariate technique that analyzes a data in which observations are described by several inter-correlated dependent variables. PCA's goal is to extract the important information from the variables, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables (Abdi and Williams, 2010).

Suppose we have a sample of n sets of data, with measurements of k variables X_1, X_2,….X_k for each set. We then develop a set of new variables (P_i) called principal components, which are linear combinations of the X's.

$P_1 = a_{11}X_1 + a_{12}X_2 + ….. + a_{1k}X_k$

$P_2 = a_{21}X_1 + a_{22}X_2 + ….. + a_{2k}X_k$

* *

* *

* *

$P_k = a_{k1}X_1 + a_{k2}X_2 + ….. + a_{kk}X_k$

Where a_{ij}'s are called loading, which are estimated from the data so that the constructed principal components satisfy two conditions: (i) the principal components are uncorrelated (or orthogonal) and (ii) the first principal component (P_1) explains the largest possible variations, the second principal component (P_2) explains the next largest variations and so on.

The computations involved in the principal component analysis are to obtain estimates for the loadings, a_{ij}'s with which we develop the orthogonal variables or principal components. Next is to carry out some test of significance to decide whether the estimates, a_{ij}'s, are statistically significant. Finally, it is to decide how many principal components to be retained. The maximum number of principal components is equal to the number of explanatory variables, k. However usually only few are retained in the analysis.

A simple approach for finding the factor loadings is given below:

i) Compute the simple correlation coefficients between the k explanatory variables. These are arranges in the form of correlation matrix.

ii) Determine the sum of each column (or row) of the correlation matrix, i,e., $\Sigma^k_1 r_{xixj}$.

iii) Determine the sum total of the column (or rows) sums, i.e. $\Sigma^k_1\Sigma^k_1 r_{xixj}$ and obtain its square root, i.e., obtain $\sqrt{\Sigma^k_1\Sigma^k_1 r_{xixj}}$.

iv) Then the loadings a_{ij} for the first principal component, P_1, are obtained as follows:

$$a_{ij} = \Sigma^k_1 r_{xixj} / \sqrt{\Sigma^k_1\Sigma^k_1 r_{xixj}}. \quad \text{……..(1)}$$

The eigen value (or latent root or characteristic root) of each of the principal components is the sum of the squares of the loadings of that principal component and is denoted by λ_i. i refers to the ith principal component. Thus it follows

$$\lambda_1 = a^2_{11} + a^2_{22} + \ldots + a^2_{1k}$$

and in general, we can write

$$\lambda_m = a^2_{m1} + a^2_{m2} + \ldots + a^2_{mk}$$

$$= \Sigma^k a^2_{mj}$$

Where λ_m is the eigen value (latent root) of the nth principal component and m refers to the order of the construction of the principal component.

As the principal component analysis tries to extract the maximum possible variance for P_1, its eigen value (λ_1) is always greater than the eigen value (λ_2) is always greater than (λ_3) and so on.

In principal component analysis the maximum number of principal components which can be extracted is equal to the number of explanatory variables and therefore a problem arises for meaningful interpretation, i.e. how many have been suggested to solve this problem, there are two approaches which are commonly used. It is given as follows:

i) Eigen value specification: those eigen values whose values are greater than 1 are retained while the rest are not taken. Greater the variables more its preference.

ii) Scree plot test: The rule is to draw the graph of eigen values against the order of the principal components. (Shenoy and Pant,1994 for further details)

Several tests have been suggested for testing the significance of factor loadings, a_{ij} of these the most commonly accepted which will be taken in this work is the one which consider the factor loadings values greater than 0.30 as significant, provided the sample size is not less than 50.

The goals of PCA are to

(i) extract the most important information from the data;

(ii) compress the size of the data set by keeping only the important information;

(iii) simplify the description of the data set;

(iv) analyze the structure of the observations and the variables.

In order to achieve these goals, PCA computes new variables called *principal components* which are obtained as linear combinations of the original variables. The first

principal component should have the largest possible variance (i.e., inertia and therefore this component will 'explain' or 'extract' the largest part of the inertia of the data). The second component should be computed under the constraint of being orthogonal to the first component and to have the largest possible inertia and so on the other consecutive components are computed likewise. The values of these new variables are called *factor scores*, and these factors scores is used in interpreting geometrically as the *projections* of the observations onto the principal components. (Kutsoyiannis,1977, 2nd edition)

7.4 Empirical Findings: the determinants of employment/ employability

In the earlier chapter, socio economic structure of the respondent were analyzed by using different techniques like mean, standard deviation and chi square. Revisiting the significant variables the PCA is performed on the following variables:

i) Head of the family's occupation ii) Years of education for the head of the family iii) Female headed household iv)Family size v) Family monthly income vi) Area of residence vii) Years of education of the respondent viii) Age of the respondent ix) Head of the family's age x) Gender of the respondent. The summary information of these variables is given in table 7.1.

Table: 7.1 Summary statistics of the variables considered

Variable used in PCA	Mean	Standard deviation
Head of the family's job	0.4	0.491
Years of education of the head	10.74	4.21
Gender of the respondent	0.62	0.486
Age of the respondent	27.62	5.967
Years of education	13.87	2.84
Area of residence of the respondent	0.3	0.457
Head of the family's age	57.2	20.48
Family size of the respondent	5.35	1.67
Family monthly income	24391.14	11261.62
Female headed household	0.06	0.236

Source: Author's primary data computation

Before applying the technique, first the appropriateness of applying Factor analysis was checked. For this Kaiser-Meiyer-Olkin Index (KMO) and Bartlett's test of Sphericity is done.

The KMO Index was found to be 0.615 which is greater than 0.5, the chi-square was 366.723 with degrees of freedom as 45 and significance level in Bartlett's test of sphericity was less than 0.001 (1% level of significance). The results are statistically significant. The table 7.2 showing KMO as well as Bartlett's test yielded very good results. So, PCA was considered very appropriate to apply.

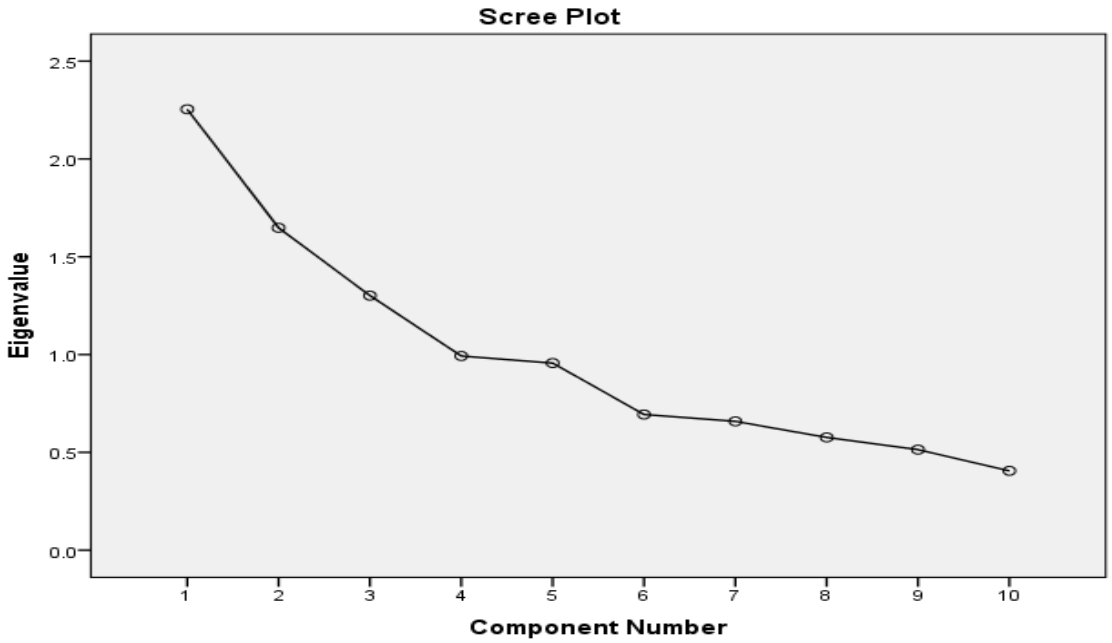
Table 7.2: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.615
Bartlett's Test of Sphericity	**Approx. Chi-Square**	366.723
	df	45
	Sig.	.000

Source: Author's calculation

PCA is undertaken on the set of data having 271 observation covering 10 variables each of which is a determinant of employability. Using the methodology explained in the earlier sub-section principal components (PC) are generated. In total 10 PCA's explain 100% of variables in the data. The Scree-Plot figure shows the eigen values of all the PCA in Figure 7.1. The eigen values are in decreasing order.

Figure: 7.1: Scree Plot of factors and eigenvalues



Note: The scree plot above shows three sharp edges at three factors.

In table 7.4.2 it is found that the first three PCA together explain 52 % of variations in the data under consideration.

Table 7.3: Factors and Initial Eigenvalues

Factors	**Initial eigenvalue**	**% Variance Explained**	**Cumulative %**
1	2.254	22.541	22.541
2	1.648	16.478	39.019
3	1.301	13.011	52.030

Therefore only three PCAs are of significance and are used for analysis. For further analysis final eigen values are generated after rotation in table 7.4

Table 7.4: Final Eigenvalues after rotation

Factors	Final eigenvalue	% Variance Explained	Cumulative %
1	1.746	17.458	17.458
2	1.731	17.312	34.770
3	1.726	17.260	52.030

Source: Author's computation from the survey

The cumulative variations explained here too is 52%. For the 10 variables under consideration the communalities, i.e. the explained variances varies from 0.173 for family size to 0.639 for female headed household.

The extraction has been done for all ten variables as indicated in the table and it is found to be more than 0.5 for 8 out of 10 variables.

Communalities

The initial and final communalities were obtained as follows:

The communality is the amount of variance in a variable explained by different factors. Initially the number of factors extracted is the same as the number of variables. So, the communality in each variable is unity.

Table 7.5: Initial and final communalities

Variables	Initial	Extraction
Head of the family's job (formal/informal)	1.000	0.489
Years of education of the head of the family	1.000	0.591
Gender of the respondent	1.000	0.577
Age of the respondent	1.000	0.499
Area of residence of the respondent	1.000	0.524
Years of education of the respondent	1.000	0.576
Head of the family's age	1.000	0.629
Family size	1.000	0.173
Family monthly income	1.000	0.506
Female headed household	1.000	0.639

Extraction Method: Principal Component Analysis

Factor matrix and factor loadings

Theoretically where there are 10 variables one can generate 10 PCAs. These 10 PCAs are orthogonal to one another, i.e. the correlation between them is zero. The first principal component extracts the maximum information. The second principal component is generated from these unexplained variation after drawing the first PC. The third PC after extracting PC_1 and PC_2 - thus the orthogenity. However each PCA uses unused variations from all the variables, hence one can generate weights of factor loads for each variable for all the PC. The factor loads for PC_1, PC_2 and PC_3 are presented in table 7.4.6. These factor loads are used for Factors which are linear combination of factor loads and standard normal variables. In table 7.4.6, the factor loads of rotated components matrix are presented.Factor matrix and factor loadings after rotation were as follows

Table 7.6: Component (Factor Loads) matrix of the variables for Employment Index

	Principal Component		
Variables	1	2	3
Head of the family's job (formal/informal)	0.479	0.498	-0.111
Years of education of the father	0.602	0.356	-0.321
Gender of the respondent	-0.364	0.118	0.656
Age of the respondent	0.450	-0.194	0.509
Area of residence of the respondent	0.577	0.222	0.376
Years of education of the respondent	0.711	-0.123	-0.236
Head of the family's age	0.363	-0.703	-0.054
Family size of the respondent	0.042	-0.238	0.338
Family monthly income	0.517	0.239	0.426
Female headed household	-0.298	0.742	0.025

Source: Author's computation from the primary survey

Table 7.7: Rotated component matrix[a] of the variables

	Principal Component (Factor Load)		
Variables	**PC_1**	**PC_2**	**PC_3**
Head of the family's job (formal/informal)	0.545	-0.275	0.341
Years of education of the father	0.726	-0.085	0.239
Gender of the respondent	-0.654	-0.284	0.262
Age of the respondent	-0.114	0.317	0.620
Area of residence of the respondent	0.190	-0.012	0.698
Years of education of the respondent	0.592	0.393	0.267
Head of the family's age	0.058	0.790	0.046
Family size of the respondent	-0.279	0.216	0.221
Family monthly income	0.121	-0.053	0.699
Female headed household	0.016	-0.799	-0.015

Note: Extraction method is done with the help of Principal component analysis and Rotation method is done using Varimax with Kaiser Normalization. a denotes the rotation converging in 6 iterations.
Source: Author's computation from the primary survey

After the extraction method of principal component analysis, the factor components which are greater than 0.5 have not been suppressed and been taken as the variables of the factors.

Thus, Factor 1 comprised variables are:

i) Head of the family's job (formal/informal) ii) Years of education of the head of the family iii) Gender of the respondent iv) Years of education of the respondent

Factor 2 comprised variables:

i) Head of the family's age ii) Female headed household

Factor 3 comprised variables:

i) Age of the respondent ii) Area of residence of the respondent iii) Family monthly income

Labeling the factors

The factor 1 can be termed as social background of the respondent, factor 2 can be termed as characteristics of head of the family and factor 3 as individual's economic background and efficiency.

Thus the factor 1 social and educational background of the respondent constitute 17.46% followed by factor 2 the characteristic of the head of the family with 17.31 % and the factor 3 individual's economic background and efficiency with 17.26 %.

The table has mean and median for all the respondents corresponding to the selected variables. It also has component score for each variable which explains that for example if head of the family's job is either formal or informal (1 being formal and 0 being informal), if it is 1 i.e. being formal then the component index shows 0.545 units more likely for the respondent to be employed. Likewise area of the residence of the respondent shows that if the respondent is from urban which is 1 then it has 0.190 units more than the ones which are from rural.

Similarly for continuous variables like the age of the respondent, increasing a unit of the age of the respondent will decrease the employability index by 0.114.

7.4.1 The Relationship between the employment status and the employability

The following are the various methods considered to find the relationship between the employment status of the respondents and their employability. The employability is

calculated from the principal component analysis of the selected variables. The employability scores index (E.I.) is derived from the first component of principal component analysis as follows:

$E.I._1 = F_1V_{11} + F_2V_{21} + F_3V_{31} + F_4V_{41} + \ldots\ldots\ldots F_nV_{n1}$

$E.I._2 = F_1V_{12} + F_2V_{22} + F_3V_{32} + F_4V_{42} + \ldots\ldots\ldots F_nV_{n2}$

…… ……. …….

…. …… …….

$E.I._{271} = F_1V_{1271} + F_2V_{2271} + F_3V_{3271} + F_4V_{4271} + \ldots\ldots\ldots F_nV_{n271}$

Where E.I. is the employability scores or index, F are the factor loads and the V standard normal variables for every respondent. Using this methodology E.I. (score) for all 271 respondents have been estimated.

The range of E.I. which derived using factor loads of PC_1 is -2.46 to 2.79. As E.I. scores index is generated using the information from all 10 variables under consideration, it is a good representative of Employability of the respondents. Thus E.I. itself is a variable which can be used for further analysis.

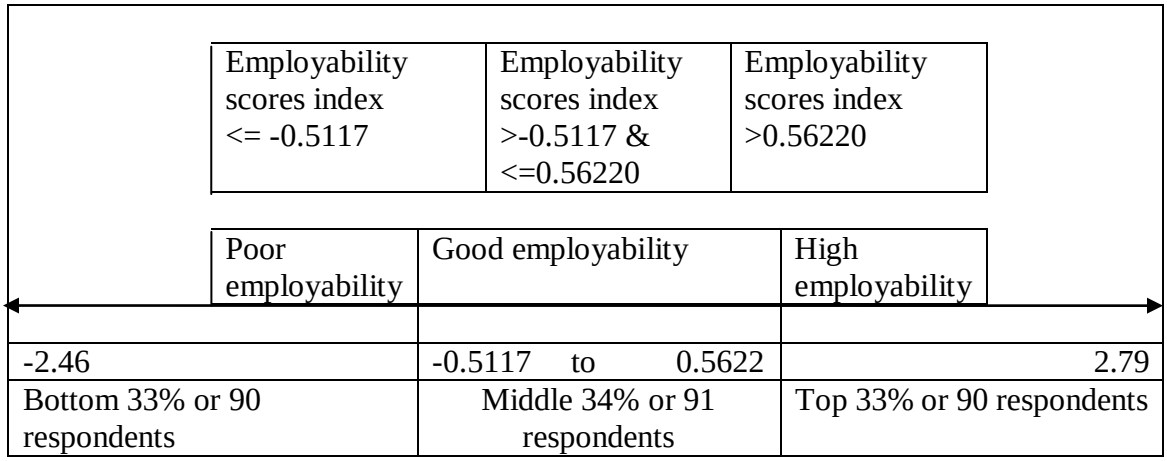
The first method of determining the cut off scores

The employment score has been divided into three parts where the first one is categorized as poor employability which is score of lowest 33% of the respondents, 34% for the middle section is termed as good employability and the top 33% is termed as the high employability section. The cutoff is taken using the methodology of Henry et al, (2000) in their work on 'Assessing the relative poverty of microfinance member household'.

The respondents employment score index which has scores less than -0.511 are termed to be of poor employability and those whose employment score are greater than 0.5622 are considered to have high employability. Those respondents who have employability

scores in the range -0.5117 to +0.56220 have been classified as 'Good Employability' category.

Fig 7.2: Determining the cut off scores for the employability



Employability scores is an Index which combines various variables that potentially determine employment of Individuals. However at a given point of time, the actual employment status is factual. One may come across individuals with high E.I. unemployed and individuals with low E.I employed. To statistically test the dependence or independence of the E.I. score and Employment status Chi-square test is conducted. The results are presented in table 7.8.

The table below shows that out of 271 respondents 33.57 percent or 91 respondents had good level of employability, where 90 respondents each were in poor employability as well as high employability. In case of those who were unemployed highest share of proportion which was 45.2 percent were considered to be of high employability. However for those who were already employed largest portion were in the poor level of employability which was 50.4 percent.

Employability Index and Employment status are dependent on one another as the results of cross tab presented in Table 7.8 shows high χ^2 value which is statistically significant at 1% level.

Table 7.8: Employment status and the level of employability index of the respondents

Employment status		**Employability**		
	Poor	**Good**	**High**	**Total**
Unemployed	27 (18.50) (30.0)	53 (36.30) (58.2)	66 (45.20) (73.3)	146 (100) (53.9)
Employed	63 (50.4) (70.0)	38 (30.4) (41.8)	24 (19.2) (26.7)	125 (100) (46.1)
Total	90 (33.21) (100)	91 (33.57) (100)	90 (33.21) (100)	271 (100) (100)

Value of χ^2= 35.056 sig. (2 tailed)= 0.000 d.f. =2 N=271, the figures inside parentheses are in percentage

Source: Author's calculation from the primary survey

The second method of determining the cut off scores

In this method the total sample is divided into three parts based on their employment score index.

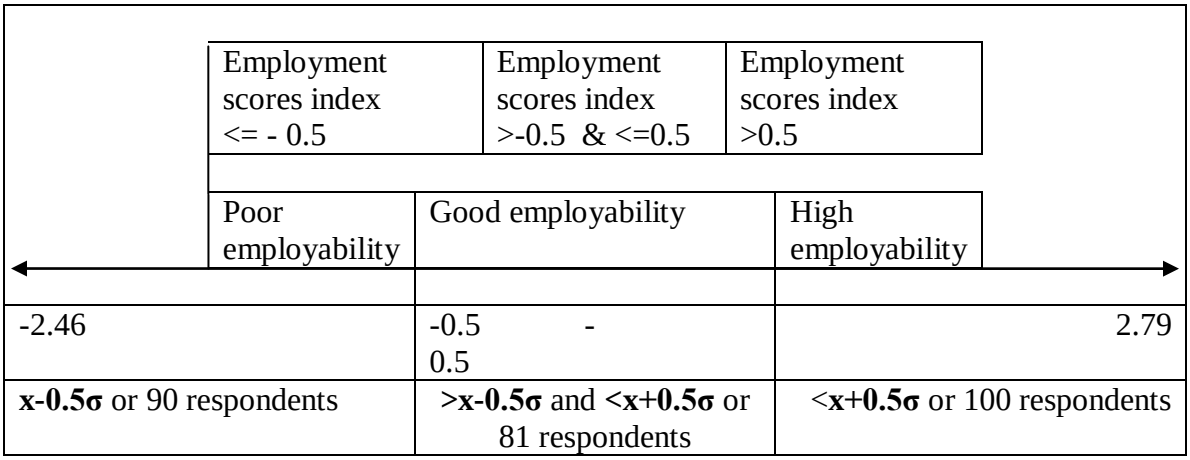
The range of the cutoff scores is $x \pm 0.5\sigma$ where x is the mean and σ is the standard deviation. Since it is standardized normal distribution the $x = 0$ and $\sigma=1$. Therefore the lower range of the cut off are those which are less than -0.5 and the higher range of cut off are those which are greater than 0.5 and the rest in the middle. Given below figure 7.4.3 explains the cut off.

The respondents employment score index which has scores less than -0.5 are termed to be of poor employability and those whose employment score are greater than 0.5 are considered to have high employability. Those respondents who have employability scores in the range -0.5001 to +0.5 have been classified as 'Good Employability' category.

Employability scores is an Index which combines various variables that potentially determine employment of Individuals. However at a given point of time, the actual employment status is factual. One may come across individuals with high E.I. unemployed and individuals with low E.I employed. To statistically test the dependence

or independence of the E.I. score and Employment status Chi-square test is conducted. The results are presented in table 7.9.

Fig 7.3: Determining the cut off scores for the employability



The table below shows that out of 271 respondents 29.88 percent or 81 respondents had good level of employability, where 90 respondents were in poor employability and 100 respondents in high employability. In case of those who were unemployed highest share of proportion which was 50.68 percent were considered to be of high employability. However for those who were already employed largest portion were in the poor level of employability which was 50.4 percent.

Employability Index and Employment status are dependent on one another as the results of cross tabulation presented in Table 7.9 shows the second method based of the figure 7.3. It shows high χ^2 value which is statistically significant at 1% level.

Table 7.9: Employment status and the level of employability of the respondents

Employment status		**Employability**		
	Poor	**Good**	**High**	**Total**
Unemployed	27 (18.50) (30.0)	45 (30.82) (55.56)	74(50.68) (74.0)	146 (100) (53.9)
Employed	63 (50.4) (70.0)	36(28.8) (44.44)	26 (20.8) (26.0)	125 (100) (46.1)
Total	90 (33.2) (100)	81 (29.88) (100)	100 (36.90) (100)	271 (100) (100)

Value of χ^2= 37.035 sig. (2 tailed) = 0.000 d.f. =2 N=271, the figures inside parentheses are percentage

Source: Author's calculation from the primary survey

The third method of determining the cut off scores

In this method the total sample is divided into five parts based on their employment score index. Each part is divided equally having 20 percent weightage in ascending order of the factor score index. The least, low, mid, high and very high are the groups that are equally divided.

Given below figure 7.4 explains the cut off. The respondents employment score index which has scores less than 0.88841 are termed to be of least employability or for those whose employment score are greater than 0.88841 and less than 0.29254 are considered to have low employability or those whose employment score are greater than 0.29254 but less than 0.35818 are considered to be mid employability or if the employment score are greater than 0.35818 and lower than 0.91962 then it is of high employability or if the employment score is greater than 0.91962 then it is considered very high. Those respondents who have employability scores in the range -0.5001 to +0.5 have been classified as 'Good Employability' category. The results are presented in table 7.10

Fig 7.4: Dividing into 5 equal parts, with each having 20 percent shares.

Employabilty score index (E.I)						
	E.I<= -0.88841	E.I.> -0.8884 &<= -0.29254	E.I.> 0.29254 & <=0.35818	E.I.> 0.35818 & <=0.91962	E.I.> 0.91962	
←	Least	Low	Mid	High	Very High	→
-2.46			>0.29254 & <=0.35818			2.279
	20% or 54 respondents	20% or 54 respondents	20% or 54 respondents	20% or 54 respondents	20% or 55 respondents	

The table 7.10 shows the employability score is high among the larger portion of unemployed whereas it is low among the employed respondents. The reason being that the employability score is generated from combining different variables leading respondent's employment score to be high. Therefore a situation like a person having high score of employability being unemployed can exist.

Table 7.10: Employment status and the level of employability of the respondents

Employment status	**Employability**					
	Least	**Low**	**Mid**	**High**	**Very high**	**Total**
Unemployed	18	16	33	36	43	146
Employed	36	38	21	18	12	125
Total	54	54	54	54	55	271

Value of χ^2= 39.714 sig.= 0.0000 d.f = 4 N=271

Source: Author's calculation from the survey

7.4.2 Age and the employability score index

The employability score of the respondent is dependent on the age of the respondent. The respondents age and employability score are cross classified and have been presented in table 7.11. The cross tabulation is used for testing for the test of dependence. Chi-square test is concluded and it is found that chi-square value is 35.817 and is found to be statistically significant at 1 percent level. It can be inferred that age of the respondent and employability score are dependent.

Table 7.11: Age range and the employment scores index

Employability Index				**Total**
Age range	**Poor**	**Good**	**High**	
18 to 25 years	37	37	22	96
26 to 35 years	35	46	69	150
36 to 45 years	18	7	0	25
Total	90	90	91	271

Value of χ^2= 35.817 sig. = 0.000 d.f = 4 N=271

Source: Author's calculation from the survey

7.5 Determinants of the employment: Regression Analysis

This section analyzes the relationship between employment status, employability factor score index and the other important determinants of employment. Multiple regression as well as logistic regression are used where ever necessary. Multiple regression technique helps in building suitable models and the coefficients are partial regression coefficients. Only the important variables are incorporated in the function. Family income is reflective of the economic and social status. It also captures the earning capacity of its members and their physical and human capital. Multiple regression analysis with dependent variable as family income with variables such as i) number of family members, (ii) education of the head of the family, (iii) education of the respondent, (iv) age of head of the household

as independent variables is conducted. Subsequently regression equations are fitted with salary of respondents as dependent variable.

Multiple regression

The following is the brief description of multiple regression:

$Y= f(X_1, X_2, X_3…..X_n)$

In the linear form, it is

$Y= \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 +…… \beta_nX_n + \varepsilon_i$

where, X_i are the independent variables and α is the constant and β are the partial regression coefficient of the respective independent variables where n is the number of independent variables taken.

The first multiple regression analysis is

Family income = $f(X_1, X_2,X_3,X_4)$

(Where X_1 is the education of the head of household, X_2 is the education of the respondent, the X_3 is the age of the head of the household in years and X_4 is the number of working adults)

The results is as given below

Family income = 11017.29 + 494.771X_1 + 259.322 X_2 + 32.571 X_3 + 507.541 X_4
(2.647) (2.922)* (1.014) (0.975) (1.058)
R^2 is 0.058 and F value is 4.123

The education of the age of the head of the household showed that for every 1 year increase in the age of education of the head of the household there will be Rs 494.771 increase in Family's income and it is statistically significant at 1 % level of significance. The other variables are not significant, but all of them exhibit positive sign of the coefficient. The model is a good fit as the F-statistic is significant at 1% level. Though R^2

appears to be low, but for sample size of 271, it is not insignificant. F statistic is the confirmation test for the model.

Of the 271 respondents only 125 are working. The respondents receive wages/salary and can be used as dependent variable in the Multiple regression analysis. An attempt is made to fit multiple regression with relevant explanatory variables. The following regression equations are fitted.

1. Salary of the respondent = f (Education of respondent, Age of respondent, Age of respondent^2)
2. Salary of the respondent = f(E.I., Age of respondent)
3. Salary of respondent= f(E.I., Age of respondent, Age of respondent^2)

The result is shown below

1) Salary of the Respondent = -15640 + 828.771X_1 +985.979X_2 – 12.694X_3
(-1.157) (3.444)* (1.091) (-0.912)
R^2 is 0.148 and F-value is 7.025

The variables show that the level of education of the respondents in years is the significant variable and is statistically significant at 1%. It implies that the salary of a respondent is positively related to the years of education pursued which can mean that a person who is more qualified gets better salary in the market. Education is a proxy to human capital. The age of the respondents even though showing a positive relation with the salary is not statistically significant. This implies that age which is proxy to experience is a positive factor influencing earnings.

The next equation includes the employability score/index derived with the help of PCA. The factor score-1 has been taken as it explains the larger share of the variance than any other factor.

Given below is the employed respondents and the relationship between the age and the employability scores/index. A Multiple regression is used to analyse the data.

Salary of the respondent = f (X_1,X_2)

where X_1 is the employability score index and X_2 is the age of the respondents.

2) Salary of the respondent = 21339.436 + 7736.932X_1 – 445.794X_2
(10.316) (15.934) (-6.088)****
R^2 = 0.688, F value is 134.258

The result show that salary of the respondent as dependent variable regressed with the age and the employment score index and it is statistically significant at 1% level of significance. The age showed inverse relationship which meant that as age increases the salary reduce significantly, other things remaining constant. It may be because of the young respondents getting more better pay due to their potential. The Employability score explains variations in salary significantly. The model is a good fit as the F-statistic is significant at 1% level and R^2 is high.

To check the growth of the salary with years of age of the employed respondents another regression had been fitted, the results are as given below. The Age^2 is added to the regression to check its significance.

Salary of the respondent = f (X_1, X_2, X_3) where X_1 represents the employability score index, X_2 is the age and the X_3 is the age^2.

3) Salary of the respondent= -4274.652 + 2357.038X_1 + 860.321X_2 -8.028X_3
(-0.300) (3.095) (0.938) (-0.560)**
R^2 = 0.692, F value is 90.664

The result showed that age of the respondents mattered in terms of salary and it had positive relationship. The growth rate of salary along with the age of the respondents showed increasing at decreasing rate as the coefficient is negative, however it is not statistically significant. The t values are not significant, the reason may be due to the age and Age^2 variables being highly correlated and hence the problem of multicollinearity. The model is a good fit as the R^2 has improved and F-statistic is highly significant.

After analyzing the major factors influencing the probability of employment further analysis is done using technique of the Logit Model. Many statistical tests require the dependent (response) variable to be continuous so a different set of tests are needed when the dependent variable is categorical. One of the most commonly used tests for categorical variables is the Chi-squared test which looks at whether or not there is a relationship between two categorical variables but this doesn't make an allowance for the

potential influence of other explanatory variables on that relationship. For continuous outcome variables like the ones analysed earlier, Multiple regression can be used. However the multiple regression, linear probability model suffers from two serious weakness. First is the problem of heteroscedasticity of the error term and the other is that the predicted value of the probability of happening the dependent event might overflow the zero and one interval, which is not possible theorectically (Gujarati, 2004).

Therefore the most reliable choice depends on the techniques that can solve this problem and limit probability to 0-1. The model that satisfies the conditions is the logit model. Empirical evidence has proven that this model gives good results. In the current excessive Logit model is used to analyse the determinants of employment.

Logit Model (Logistic Regression)

To analyze the qualitative variables such as employment status of the respondents logit model is used. The employment status is categorized as 0 or 1. It is a dummy variable and is considered as dependent variables. An educated person may be employed or unemployed, but a group of educated persons would exhibit higher probability of employment. Similarly an individual with a specific age in the working age group may or may not be employed but as age progresses the probability of employment would increase. In the real work if a set of factors explain a happening of event, the same set factors explain the probability of event not happening. To capture the impact of the explanatory factors on the event happening or otherwise is best done using logistic function or the logit model. Logit model is also known as logistic regression which is useful when dependent variable takes value between 0 and 1. When dependent variable value is 0 or 1 neither ordinary least square method nor weighted least square is helpful. The dependent variable is 1, if a respondent is employed and 0 if a respondent is unemployed. Instead of t statistic Z test is used to study the statistical significance of the independent variables. In logit model, the conventional goodness of fit, R^2 is not meaningful, therefore Mcfadden R^2 is used. The important thing in logit model is the expected sign of the regression coefficient and their statistical significance. Hence in logit model instead of using F test as in linear regression model, likelihood ratio (LR) statistic is used. LR statistic follows the chi-square distribution with degrees of freedom equal to

number of explanatory variables. For meaningful interpretation of logit function result the odd ratio is used which is obtained by taking the antilog of various slope coefficients. Odd ratio shows the ratio of probability in the model.

The logit model is used to examine the factors which can be affecting the employment status of the respondents. Many variables are included as they are probable explanators but some of the variables may not be significant. Usually, the variables that are considered to be important theorectically or realistically (social) are included in the function. Their explanation may vary with data set, i.e. economic or social strata.

The following is the brief description of logit model:

$$P_i = \frac{1}{1+ e^{-(\alpha+\Sigma\beta_i X_i+\varepsilon_i)}}$$

or

$$\text{Log}\left(\frac{P_i}{1-P_i}\right) = (\alpha + \Sigma\beta_i X_i + \varepsilon_i)$$

Where

P_i is the probability of the ith respondent getting employed

X_i is the independent variable for the ith respondent.

α, β are the parameters of the model to be estimated

ε_i is the stochastic error term assumed to be independently and identically distributed with zero mean and unit constant variance.

Various models are fitted and analyzed with employment status being the dependent variable which will predict the probability of the employability.

7.6 Results and Discussion

Logistic regression models are fitted with both quantifiable and dummy explanatory variables. All together nine variables have been used. The list is presented below. In addition the first three principle components generated using the method of Principle component analysis (PCA) is obtained from rotated matrix using Varimax method. The PCA have also been used as explanatory variables. The PCA methodology and the results are presented in the earlier section of this chapter. The models fitted explain specific relation between the employability status and the variable incorporated.

The list of variables used in the models are:

i) Head of the family's job (informal/formal) = X_1
ii) Years of education of the head of the family= X_2
iii) Gender of the respondent= X_3
iv) Age of the respondent = X_4
v) Area of residence of the respondent =X_5
vi) Years of education of the respondent = X_6
vii) Female headed household = X_7
viii) Head of the family's age = X_8
ix) Family monthly income = X_9

Table 7.12 presents the model 1 and 2 and the logit regression results fitted to find the relationship with respect to employment status as the dependent variable and independent variables. Here it is analyzed to check their significance level with respect to employment status.

The model 1 consist of employment status as the dependent variable and years of education of the head of the family (X_2), Age of the respondent (X_4) and Years of education of the respondent (X_6) as independent variables. The analysis found that X_4 and X_6 are statistically significant which implies the employment status of a respondent dependent on the age of the respondent as well as the years of the education of the respondent. X_2 was not significant which implies years of the education of the head of the

family did not have a significant impact on the employment status of the respondent and the coefficient is negative.

The model 2 consist of employment status as the dependent variable and the independent variables are the head of the family's job (X_1), Age of the respondent (X_4), Years of education of the respondent (X_6) and Female headed household (X_7) with high odds ratio of 3.96. The analysis found that X_4, X_6 and X_7 are statistically significant which implies that employment status of being employed was positively correlated with age, years of education of the respondent and whether the respondent belonged to female headed household though significant with negative sign. However X_1 was not significant which meant job whether informal or formal of the head of the family of the respondent had nothing to do with the employment status and the respondents characteristics are more relevant.

LR statistic is statistically significant for both the models. Mcfadden R^2 is found to be 0.1421 for model 1 and 0.1593 for model 2.

Model 1: ES = f(X_2, X_4, X_6)

Model 2: ES = f(X_1, X_4,X_6,X_7)

Where ES is the employment status

Table 7.12: **The employment status and the variables contributing to employment factors@ (Dependent Variable: Employment Status)**

	Model 1		Model 2	
Variables	**Coefficient**	**Odds ratio**	**Coefficient**	**Odds ratio**
Constant	-0.335876		-0.984956	
X_1	-		-0.221968	1.248
X_2	-0.001979	1.001	-	
X_3	-		-	
X_4	0.098519***	1.103	0.176159***	1.192
X_5	-		-	
X_6	-0.175489***	1.191	-0.285405***	1.33
X_7	-		1.377249**	3.963
X_8	-		-	
X_9	-		-	
N	271		271	
Log-likelihood	-160.4431		-157.2175	
LR Statistic	53.17055***		59.62189***	
McFadden R-square	0.142146		0.159393	

Note: *Significant at 10%, ** Significant at 5%, *** Significant at 1%

@ Computed from primary survey data by the researcher

Table 7.13 shows the result of the employment status and all the nine variables. The table shows that only 5 variables out of 9 were found to be statistically significant. The significant variables are X_4 (age of the respondent) with a positive coefficient which implies higher the age greater the possibility of a person to be employed (approximately 1.172 times for every one year increase) with an odds ratio of 1.172, X_6 (years of education of the respondent) negative correlation which implies that higher the age of education lesser the chance of the person to be employed (approximately 1.3 times less for every 1 year more) and X_9 (family monthly income) positive relation which implies higher the family income has the higher chance of a person being employed all at 1% level of significance. X_3 (gender of the respondent) is positively correlated which implies that the male persons have higher chance to be employed because of possibility of being the bread earner of the family and X_7 (female headed household) is also significant positively at 5% level of significance implying that a person from female headed family have a higher chance of working because the female headed earner may not have suffice the situation of the younger members to remain unemployed. The odds ratio is 5.54.

The variables X_1(head of the family's job), X_2 (years of education of the head of the family), X_5 (Area of residence of the respondents) and X_8 (head of the family's age) were not significant. The LR statistic shows that the model is significant overall.

The description of the model

Model X_3: ES = f(X_1,X_2……..X_9)

Table 7.13: The employment status and the variables contributing to employment factors@ (Dependent Variable: Employment Status)

	Model 3	
Variable	**Coefficients**	**Odds ratio**
Constant	-2.413388	
X1	-0.434726	1.54
X2	-0.016925	1.017
X3	0.699186**	2.011
X4	0.159383***	1.172
X5	0.303081	1.353
X6	-0.264879***	1.303
X7	1.713505**	5.54
X8	0.007355	1.007
X9	0.00000372***	1.000
N	271	
Log-Likehood	-149.06	
LR Statistic	75.93679	
McFadden R-square	0.086	

Note: * Significant at 10%, ** Significant at 5%, *** Significant at 1%

@ Computed from primary survey by the researcher

Table 7.14 gives results of the logistic regression of employment status and each of the first three factor loadings. The factor loadings were derived from the combination of 10 variables. The FAC_1 is termed as "Social Background of the respondent" because the four variables are Head of the family's job (formal/informal), Years of education of the respondent, Years of education of the head of the family and Gender of the respondent which have factor loads greater than 0.5. The FAC_2 is termed as "characteristics of head of the family" which consist of head of the family's age and female headed household which have factor loads greater than 0.5. Similarly FAC_3 represents "Characteristics of head of the family" which consist of age of the respondents, area of residence of the respondent and the family income of the respondent all having factor loads greater than 0.5.

The table 7.14 shows logistic regression of each Factor components separately with respect to employment status. The results show that FAC_1 is highly significant which explained that the Social background of the respondent matters which consist of Educational background and the occupation of the head of the family. The respondent is likely to be employed lower the educational qualification and if the head is working in informal jobs. This explains the need of the respondents to take up a job. FAC_3 is also found to be statistically significant with the employment status implying the Economic background and efficiency level consisting of family income and the age of the respondent matter as urban respondents who are more aged had more probability of being employed as they can be more experienced in the job market. The job openings as well as the need of job can be much higher in urban areas rather than rural as the cost of living is much higher. However the FAC_2 is significant at 10% level, it explains the significance of characteristics of head of the family consisting of female head of the family and the age of the head of the family do not matter significantly. The odds-ratio for each factor is greater than one and highest for FAC_1.

LR statistic is statistically significant in all the models. McFadden R^2 is 0.073 for FAC_1, 0.008 for FAC_2 and 0.058 for FAC_3.

The descriptions of the models are given below:

Model 4: ES = f(FAC_1)

Model 5: ES = f(FAC_2)

Model 6: ES = f(FAC_3)

Table 7.14: Employment status of respondents w.r.t various determinants@ (Dependent variable: Employment Status)

Variable	Model 4	Model 5	Model 6	Odds ratio#
Constant	-0.169397	-0.155284	-0.161237	
FAC_1	-0.681443***	-	-	1.976
FAC_2	-	-0.219643*	-	1.245
FAC_3	-	-	0.599002***	1.82
N	271	271	271	
Log-likelihood	-173.1989	-185.4536	-176.1447	
LR Statistic	27.65912***	3.149659**	21.76735***	
McFadden R-square	0.073944	0.008420	0.058193	

Note: * Significant at 10%, ** Significant at 5%, *** Significant at 1%

@ Computed from primary survey data by the researcher, # values for Model 1, 2 and 3 respectively

Table 7.15 presents the logistic regression done to analyze the relationship with respect to employment status and the factor loadings viz. FAC_1, FAC_2 and FAC_3.

FAC_1 and FAC_3 turned out to be statistically significant at 1% level of significance which implies that Education and Economic background of the head of the family is responsible for the employment status of the respondent. It has negative coefficient which explains that the respondents are likely to be unemployed if the head of the family is more educated and in better economic conditions. It can be explained with the help of using odds ratio, which can also be quantified approximately to (anti log of 0.7482) 2.11 times. Addition to it the respondents social background and the efficiency level shows positive correlation which implies that being in urban and higher aged respondents are more likely to be employed. Similarly it can also be quantified approximately to (anti log of 0.6808) 1.975 times more likely to be employed.

The description of the model is given below:

Model 7: ES = f(FAC_1, FAC_2, FAC_3)

Table 7.15: Employment status of respondents w.r.t. various determinants (Dependent variable: Employment Status)

Variable	Model 7	Odds ratio
Constant	-0.172855	
FAC_1	-0.748247***	2.11
FAC_2	-0.2519*	1.28
FAC_3	0.680820***	1.97
N	271	
Log-likelihood	-159.347	
LR Statistic	55.36280***	
McFadden R-square	0.148006	

Note: * Significant at 10%, ** Significant at 5%, *** Significant at 1%

@ Computed from primary survey by the researcher

7.7 Summary and Conclusion

The main objective of the chapter was to examine the role of various variables as determinant of employment. The first technique used is Principal Component Analysis (PCA) that combines the related variables into factors which can represent collectively as a determinant influencing the employment status of a respondent. The next technique used is the Logistic regression also known as Logit model. It is utilized to see whether the factors which are derived from the PCA had the probability of influencing the employment status of the respondents.

The major findings are as follows:

- The variables taken in the PCA analysis were i) Head of the family's occupation ii) Years of education for the head of the family iii) Female headed household iv)Family size v) Family monthly income vi) Area of residence vii) Years of

education of the respondent viii) Age of the respondent ix) Head of the family's age. (The importance of these variables had been explained earlier in the chapter).

- The PCA concluded with 3 factors viz. FAC_1, FAC_2 and FAC_3 out of the nine variables. The first factor load consist of Head of the family's job (formal/informal), Years of education of the head of the family, Gender of the respondent and Years of education of the respondent representing the Social background of the respondent, the second factor consist of Head of the family's age and Female headed household which represents the characteristics of head of the family and the last factor consist of Age of the respondent, Area of residence of the respondent and the family monthly income collectively representing the Economic background and the efficiency level of the respondent.
- The three factor loads were collectively explained by 52.3 percent of the variations.
- The employability of a respondent and the employment status had inverse relationship such as if the employability of a respondent was high he or she was most likely to be unemployed and vice versa. Therefore it implies that there is mismatch of job demand and the job supply. Especially jobs for the educated were short in supply.
- The age and the employment status showed there is significant relationship between the two, which implies as the person may be in dire need for a job or he or she must have gained years of experience or he or she may have with years gathered more contacts as well.
- The salaried respondents had positive correlation with the age which can imply that higher the experience greater is the amount of salary. Even though the sensitiveness of two variables were low it is statistically significant.
- The logistic regression of the three factors, which was derived from the PCA had been analyzed along with the employment status being the dependent dichotomous variable. The findings were such that the FAC_1 which consist of educational background and occupation of the head of the family turned out to be significant along with the FAC_3 consisting of mainly the age of the respondent, area of residence and family monthly income.

- This meant that the Head of the family's job, Years of education of the head as well as years of the education of the respondent and the gender played a major role in determining the employment status of the respondent.
- The FAC_2 was found to be less significant which consist of age of head of the family and female headed household. It was significant at 10% level of significance.
- However when the logistic regression for each variables with respect to the employment status is analysed the variables gender, age, years of education of the respondent, female headed household and family income are statistically significant.