ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DETERMINANTS OF WOMEN UNEMPLOYMENT IN ETHIOPIA: A MULTILEVEL MODEL APPROACH

MESFIN MULU

A Thesis Submitted to

The Department of Statistics

Presented in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Statistics

Addis Ababa University

Addis Ababa, Ethiopia

June 2012

Addis Ababa University

School of Graduate Studies

This is to certify that the thesis prepared by Mesfin Mulu, entitled: ***Determinants of Women Unemployment in Ethiopia, A Multilevel Model Approach*** and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Statistics complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Examiner ______________________ Signature ______________Date __________

Examiner _____________________ Signature ______________Date __________

Advisor ______________________ Signature ______________Date __________

Chair of Department or Graduate Program Coordinator

ABSTRACT

Determinants of Women Unemployment in Ethiopia, a Multilevel Model Approach

Mesfin Mulu

Addis Ababa University, 2012

Employment of women in economic activities has several beneficial effects for women and their families. Women unemployment represents a growing concern worldwide. The main objective of this study was to identify and describe the determinants of women unemployment in Ethiopia. The study is made based on the 2011 Ethiopian Demographic and Socioeconomic Survey (DHS) which was conducted by Central Statistical Agency (CSA) of Ethiopia. To analyze the data, descriptive, standard logistic regression analysis and multilevel model were used. The descriptive result revealed that about 64.42% of the women were unemployed while 35.58% were employed. The logistic regression analysis was performed to investigate the effect of each predicator variable on the unemployment status of women. Accordingly, region, place of residence, age, marital status, exposure to any mass media, husband's/partner's occupation, sex of household head, economic status of the household, educational level, and presence of child of age 5 years and less in the household were found to be the significant determinants for women unemployment. Moreover, multilevel modeling was used to analyze nested sources of variability in hierarchical data, taking account of the variability associated within each level of the hierarchy. The estimates of the multilevel model show that variables that are reported to be significant in logistic regression analysis were also found to be significant. The effect of these significant variables is the same for each region in Ethiopia but the effect of mass media is not the same for each region in Ethiopia.

Acknowledgement

First, and foremost, I thank God for giving me the opportunity to pursue my graduate study at the Department of Statistics, Addis Ababa University.

I owe the deepest gratitude to Mekonnen Tadesse (Assistant Prof.), my thesis advisor for his invaluable and constructive comments and suggestions throughout my study. Without his professional assistance and guidance this study would not be realized.

My special thanks also go to all staff members of the Department of Statistics and to my colleagues for all the encouragements and support that they provided me during my study years.

I would also like to thank my employer, Haramaya University, for the sponsorship that allowed me to pursue my postgraduate education.

Finally and most importantly, my heart-felt thanks go to all members of my family for all the encouragement and support throughout my study.

Table of Content

Content **Page No.**

List of Abbreviations

- **UNDP –United Nation Development Program**
- **ILO-International Labor Organization**
- **CSA-Central Statistical Agency**
- **LFS-Labor Force Survey**
- **FLFP-Female Labour Force Participation**
- **GLIMMIX- Generalized Linear Mixed Model**
- **AIC-Akaike Information Criteria**
- **SC-Schwartz Criteria**
- **BIC-Bayesian Information Criteria**
- **EDHS-Ethiopian Demographic and Health Survey**
- **S.E-Standard Error**
- **OR-Odd Ratio**

List of Tables

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the Study

Employment of women in economic activities has several beneficial effects for women and their families. Women unemployment represents a growing concern worldwide. According to the Platform for Action adopted at the Fourth World Conference on Women in Beijing in September 1995, "More than one billion people in the world today, the great majority of whom are women, live in unacceptable conditions of poverty, mostly in the developing countries" (United Nations, 1996, p. 37). "Women now account for a growing percentage of the world‟s poor." and a publication of the United Nations Development Programme states: "70% of the world‟s poor are „women‟ (UNDP, 1995, p. 4). Women workers everywhere invariably have much higher rates of joblessness and much lower earnings than male workers. In many contexts, women are also concentrated in low-skill informal work or in hazardous forms of work that are ill-suited to their age and experience. Employment outcomes are typically worst for former women laborers and other early school-leavers, groups with least opportunity to accumulate the human capital needed for gainful employment.

Poor people have severely limited access to and control over key assets, including land and physical and human capital. Lacking production and human capital donation, the poor have low income and low consumption. Most poor people are also inadequately educated and generally less healthy than the rest of the population. Many depend for their livelihoods on low-productivity subsistence agriculture or on the informal sector, where returns to labour and capital are generally low. Workers in the informal sector have low incomes, limited protection and frequent spells of unemployment. These factors, coupled with lack of access to institutions that shape policies, prevent the poor from acquiring the capabilities for decent work. In sum, as UNECA (2005) notes, strengthening the link between economic growth, employment and poverty reduction in Africa requires, first, policies to increase the employment intensity of growth, and second, enabling the poor to integrate into the growth process and find decent work.

In most developing countries in general, and in sub-Sahara Africa in particular, the worst-affected groups in Africa"s job crisis are women, young people, the disabled and the elderly (ILO, 2010). According to ILO reports, women workers dominate the informal sector, concentrated in activities such as unpaid agricultural work, food processing, street selling, insignificant cross-border trading, marketing of processed and semi-processed agricultural products and household domestic duties. Only a small but growing percentage of women work in the formal sector - for example in teaching, nursing, mining services, manufacturing and lower-level clerical jobs. Women"s share in wage employment in the non-agricultural sector varies from 28.2 percent in Morocco to 43.1 percent in South Africa since 2010. In 2008, unemployment among women ranged between 15 percent in North Africa and 8.2 percent in sub-Saharan Africa. Women"s unemployment problems arise from a variety of factors including cultural prejudices, educational discrepancies between men and women and a lack of marketable skills.

The gender dimensions of employment can also be appreciated through an analysis of the share of employed people in the working-age population (those aged 15 years and older) or the employment to population ratio. In most countries this ratio is lower for women than for men. The unemployment rate of female and male in South Africa was 15 and 8.1 percent respectively. Similarly, the unemployment rate of female and male in sub-Saharan Africa in 2008 was 8.2 and 7.2 percent respectively (ILO, 2009a). So, this rate indicates that unemployment is more of a problem of females than males in Africa.

Ethiopia remains one of the world"s poorest countries with a per capita income of just US$102 in 2003, or approximately US$800 at purchasing power parity. The proportion of the population living below the poverty line of less than one dollar a day at purchasing power parity is estimated to be 2 percent for 2000, while 78 percent of the population lived on less than two dollars a day (again, at purchasing power parity). The economy is predominantly agricultural, as the agricultural sector accounts for about 80 percent of employment, and about 50 percent of GDP.

As part of Africa, a high level of un- and underemployment is one of the critical socio-economic problems facing Ethiopia. Women are more likely to be employed in jobs of low quality, underemployed, working long hours for low wages, engaged in dangerous work or receive only short term and/or informal employment arrangements. The inadequate employment situation of women has a number of socio-economic and political consequences. Unemployment and underemployment reflect the failure to make use of an important factor of production, labour, for fostering economic growth (ILO, 2010).

In Ethiopia unemployment rates have increased more for women than for men over the five years prior to 2004. In 1999, the women unemployment rate among women was 17.3

per cent, compared to 6.8 percent among men. The urban female women unemployment rate was 43.7 percent compared to 29.4 percent for urban male women (Berhanu et al., 2005).

According to Ethiopian labor force survey report, the unemployment rate of female and male at country level were 12.5 percent and 4.3 percent respectively (LFS, 1999). Similarly, the 2005 Ethiopian LFS reveals that unemployment rate of female and male were 7.8 and 2.5 percent respectively. According to the result of the two surveys, unemployment is more of a problem of women than that of men. This raises an interesting question on what the determining factors of women unemployment in Ethiopia are.

The prevailing situation calls for intervention in view of maximizing the number and magnitude of women at the workplace. To this effect, there is a need to identify major factors that affect the amount of participation of women at the workplace. In light of the above, the objective of this study is to investigate the determinants of women unemployment in Ethiopia. The investigation is conducted through a review of relevant literature, on some demographic and socioeconomic determinants of women unemployment.

1.2. STATEMENT OF THE PROBLEM

In general, unemployment has a significant impact on poverty, homelessness and affects family cohesion. It causes hopelessness and other social evils such as crime, violence, break up of families, alcoholism and prostitution.

The persistence of high unemployment rates in recent years has become a major problem in many African countries in general, and Ethiopia in particular. The size of the unemployed population and rates of unemployment in Ethiopia is differing by sex. The 1999 labour force survey of Ethiopia indicates that the rate of unemployment in Ethiopia was found to be 8.0 percent of which unemployment rate of male and women are 4.3 and 12.5 percent respectively. Similarly, the result labour force survey in March 2005 reveals that unemployment rate of the country was 5 percent of which male and women were 2.5 and 7.8 percent respectively. Thus, unemployment rate for women is higher than for men, implying that women are the most affected. Women unemployment is a major socio-economic problem and has the potential to cause social discontent.

Due to population pressure, the number of women looking for work is expected to increase from year to year in Ethiopia. Failure to address women employment issues will have serious consequences for the economy and society. While these general facts are clear, the specific factors affecting women employment in Ethiopia have received little research attention. There is, therefore, limited empirical basis for formulating policies and programs promoting women employment. So, this study attempts to examine the determinants of women unemployment in Ethiopia and to provide empirical information for policy makers.

1.3. OBJECTIVE OF THE STUDY

General objective:

The general objective of this study is to identify and describe the Demographic and Socio-economic determinants of women unemployment in Ethiopia.

Specific objective:

1. To identify the factors associated with women unemployment in Ethiopia.
2. To examine the extent of the variation in women unemployment status within and between regions of Ethiopia.
3. To identify the factors that may explain the variation in unemployment between regions of Ethiopia.
4. To provide relevant recommendations for policy makers and program managers.

1.4. SIGNIFICANCE OF THE STUDY

This study has the purpose of identifying the major contributing risk factors that limit the number and magnitude of women employment at the workplace in Ethiopia. Understanding the different factors in female labor supply is crucial in targeting effective policy interventions to enable all women to participate in the job market, and to achieve the policy goal of increased female labor force participation (FLFP). So, this study intended to create awareness for governmental and non-governmental organizations to take intervention measures and set appropriate plans to improve the existing level of participation of women at workplace.

The findings or results obtained from this research could also be useful in many ways. The findings could also be helpful for policy making, monitoring and evaluation activities of the government and different concerned agencies.

The study also aims to provide additional information for policy makers to develop effective counter measures that could maximize the number and magnitude of women at the workplace and distribution of manpower for surveillance.

Finally, it also helps to carry out further research to refine the conceptual and methodology of the present study.

1.5. LIMITATION OF THE STUDY

Some of the limitations of this study are:

1. The definition of “unemployed women” used in the study may underestimate the real female workforce given in ILO definition.
2. The study did not incorporate the duration of unemployment because of lack of data.
3. The study focused on identifying factors that are expected to influence women unemployment in Ethiopia. However, due to lack of data the study could not incorporate some of the most influential factors such as microeconomic and macroeconomic variables like investment, inflation and political issues.
4. The study did not make a comparative analysis of unemployment between rural and urban households, and also between women and men in Ethiopia separately.

1.6. ORGANIZATION OF THE PAPER

The paper is organized as follows. The first chapter provides a brief background of the study, statement of the problem, objective, its significance and limitation of the study. The second chapter is a review of literature related to unemployment in Ethiopia and the rest of the world. The third chapter describes the source of data, variables of the study and the methodology used for analysis. The fourth chapter presents the statistical data analysis while the last chapter provides the discussions, conclusion and recommendations made.

2. Literature Review

2.1. Concepts and Definition of Unemployment

In defining unemployment; authors like Adebayo (1999.), Dantwala (1971), Falae (1971), Encyclopedia Americana (1995), Englama (2001), and Onah (2001) share different but related views. For instance, Adebayo (1999) defined unemployment as a state in which people who can work are without jobs and are seeking for pay or profit. This definition gives rise to the problem of measurement, especially when we are interested in knowing the average rate of unemployment in the economy over a period of time. Falae (1971) considered such a definition too broad because some categories of people who are without work should not really be regarded as unemployed in any meaningful sense. Falae (1971) therefore pointed to the labour code prescription of lower and upper limits for the labour force in Nigeria and submitted that anyone who is unable to work is not counted as unemployed, even though he or she would love to work. According to the Encyclopedia Americana (1995) unemployment literally applies to all persons without work and actively looking for work. Englama (2001) points out that the unemployment rate in an economy is the number of people unemployed expressed as a percentage of the total labour force. The total labour force is defined as the number of people employed plus the number of people unemployed.

The international accepted measurement of unemployment is based on the following three criteria that must be satisfied simultaneously; “without work”, “currently available for work” and “seeking work“‟ (ILO, 1983). The standard definition of unemployment that is based on the "seeking work" criterion refers to take specific steps in specified period to seek paid employment or self employment.

2.2. Determinants of Women Unemployment in the World

Unemployment has a significant impact on poverty, homelessness and affects family cohesion. It causes hopelessness and other social evils such as crime, violence, break up of families, alcoholism and prostitution.

By understanding the negative impact of women unemployment for economy growth, many scholars from various disciplines began to investigate/identify the determinants of unemployment in different areas of the world. In the remainder of this section the most prominent empirical studies regarding are reviewed.

There are various studies that investigated the determinants of unemployment. Some studies analyzed the determinants of unemployment from a microeconomic perspective, while others investigated the macroeconomic determinants of unemployment in both developed and developing countries. There are also different theoretical models that are relevant for the investigation of the determinants of unemployment. Monternsen (1970) and Lippman and McCall (1976) presented a commonly chosen framework job search model. This model states that when people become unemployed, the expected duration of their unemployment depends on probability of receiving job offers and accepting the offers. The job offer is determined by factors such as education, skill, experience and local demand condition, all of which make a specific person attractive to employers. This model assumes that the probability that an individual accepts offer of employment depends on the individual‟s minimum acceptable wage. The minimum acceptable wage is called reservation wage and is determined by cost of looking for a job, unemployment income, expected distribution of wage offers and probability of receiving subsequent job offers.

Foley (1997) used information contained in a nationally representative longitudinal survey to analyze unemployment duration in Russia. The analysis was done by using a competing-risk, discrete-time waiting model augmented to incorporate unobserved heterogeneity. This was done to analyze if there is evidence of duration dependence in unemployment and the role of demographic characteristics, alternative income support, and local demand conditions in explaining unemployment duration for working-age individuals. The results indicate that married women are found to experience significantly longer unemployment compared to their male counterpart. Older individuals expect to be unemployed longer than younger individuals. Highly skilled or educated individuals have very low unemployment rate compared to those without education or low skilled individuals.

Michael (1985) discusses three independent inquiries into the factors of women unemployment during the period 1950-1980. He uses cross-sectional differentials in female labor force participation by characteristics including age, educational attainment, marital status, and (among married women) the proportion with younger children. He finds that the changes in women‟s unemployment over the three decades were not uniform in terms of age, marital status, or educational composition. The new employees in the 1950s were predominantly older, married, and relatively less educated, while in the 1970s they were younger, less likely to be married, and far better educated. However, although the compositional shift in the population by marital status and the presence of young children seem to have had almost no influence on overall female labor force participation, the increase in age and educational attainment has contributed about one-quarter of the rise in labor force participation. Finally, the differences in labor force participation among groups as defined by age, marital status, presence of young children, and education are far less pronounced in 1980 than they were in 1950.

Similarly, many scholars tried to identify the causes of unemployment in Latin America in different countries. For instance, Psacharopoulos and Tzannatos (1992) published a collection of studies evaluating women's employment and pay in this region. They use the empirical results obtained for each of the countries studies to draw conclusions about the general characteristics and trends in women labour force participation in the region. Similar techniques and comparable specifications are used for all the countries, to allow for an easier comparison of the results. After analyzing the results of individual studies, the authors found that most agree that the probability of a woman working for pay is greater (1) as they enter adulthood and up to the age of 40 to 45 years (after controlling for fertility); (2) if they reside in urban areas; (3) the higher their education level; (4) the more general (rather than technical/vocational) their education; (5) the lower their family responsibilities (in terms of young children present in the household); (6) if they live in a female-headed household; and (7) the lower other income and family wealth.

Yang (1992) studied female labor force participation in Costa Rica. The author finds the major factors that influence women‟s labor market activity are educational attainment, marital status, fertility, other household income, and age. Education has a powerful positive effect on the probability of female labor force participation: more educated women are more likely to participate in the market and are more likely to be employed.

Using the results of probit estimates for female labor force participation, the author predicted the probability of labor force participation for each characteristic holding other characteristics constant at their means. The author found that high school graduates have the highest probability, 54.2 percent; married women are less likely to participate than unmarried women, 17.7 percent versus 40.4 percent; the more children a woman has, the less likely she is to participate in the labor market. A female head of household has a higher likelihood of participating, 34.1 percent, compared to 22.7 percent probability for a woman who is not head of household; women who live in rural areas are less likely to participate in market activities.

Elster and Kamlet (1990) model unemployment of married women from a sociological perspective. They examined whether traditional economic variables have a differential influence across social groups (defined in their paper by broad occupational and age classifications). They also studied whether "income aspirations" have a differential influence across these same social groups. Data for the study were drawn from the U.S. Bureau of the Census, 1980 and public-use microdata sample for the Pittsburg Standard Metropolitan Statistical Area (SMSA), 1983. Results from their logit analysis indicate that individuals‟ responses to particular influences such as education, age, past marital history, fertility, and income aspirations, differ across social groups. It follows from this that such differences influence married women‟s labor force participation behavior.

Franz (1985) analyzed female labor force participation in Germany. He used the Tobit procedure, which allows estimating labor supply functions including both hours worked and labor force participation in a cross-section analysis based on individual data. He found that it is necessary to distinguish among women by marital status: while labor force participation of young single women decreased substantially, married women have a higher labor force participation rate. The Tobit estimates showed that labor supply increases with higher education and with vocational education, and also if the husband is self-employed. The income of the husband had a negative impact on the woman‟s supply of labor. The presence of children reduces labor supply: the younger the children, the more labor supply decreases. In general, the author found out that labor force participation of married women increases slightly until the age of 28, and then it declines monotonically. Finally, he found foreign-born women work more than German women do.

Scott (1992) studied female labor force participation in Bolivia. She used data from the second round of the 1989 Integrated Household Survey (SIH), a biannual survey carried out by the National Statistical Institute of Bolivia (INE). The results revealed that 44 percent of the sampled women work for pay. However, the definition of "employed women" used in the study may have underestimated the real female workforce because unpaid workers in a family business were not counted. In general, women who have lower levels of education than men are more heavily concentrated in the informal sector (World Bank, 1989). Probit estimates of the labor force participation function shows the greatest likelihood of working for pay among women aged 35 to 44 but the probability declines among older women. Unmarried women and heads of household are more likely to work than are married women. Women high school students are less likely to participate in the labor market than those who are not. In contrast, attending, or having completed a technical school, teacher‟s college, or university degree has a highly significant, positive effect on the probability of labor force participation. Pregnancy has the expected negative impact: women who were pregnant in a given year had a lower probability of participating in the labor market than women who had not been pregnant. She also reported that language skills also have a significant impact on labor force participation: bilingual women participate at a higher rate than women who speak only Spanish.

Mahmood et al (2011) studied the basic causes of unemployment among the educated segments in Peshawar Division of Pakistan based on a sample of 442 individuals belonging to Peshawar Division who have at least first degree or are capable of any professional/technical job whether they are employed or unemployed. The paper is an attempt to determine important factors effecting unemployment among the educated segments. They used Logistic regression for their study. The final model concludes that high growth of population (HP), lack of resource (LR), HP*RL (role of attitude in getting high level jobs), HP*NEJ (Non coordination between education and job opportunity), NEJ*RB (red ribbon)*RL are important determinants of unemployment rate in Peshawar Division. Their analysis shows that 69.6% of the males and 30.4% the females are educated and unemployed and thus the percentage of overall employment is comparatively low than developed countries. The backward elimination procedure with the initial model fitted by Brown method have revealed that high growth of population (HP) and lack of resources (LR) are the main effects and HP*RL, HP*NEJ, NEJ*RB*RL are the interaction effects that are significantly causing unemployment among the educated segments.

Tansel and Tasci (2004) identified the determinants of unemployment duration for men and women in Turkey. They used the results of the Household Labor Force Surveys of

2000 and 2001 to construct a cross-section of durations of unemployment spells and analyze the determinants of probability of leaving unemployment or the hazard rate. They also examined the effects of the personal and household characteristics and the local labor market conditions. They used Non-Parametric and parametric estimation methods controlling for the unobserved heterogeneity. The analyses were carried out for men and women separately. Their results indicated that women are experiencing higher unemployment durations than men. Age has a negative and education has a positive effect on the hazard rate. The effect of the local unemployment rate is large and negative; duration dependence of the exit rate from unemployment is different for men and women; for men, there is slight U-shaped duration dependence, while for women there is no duration dependence.

2.3. Determinants of Women Unemployment in Africa

Unemployment is one of the most serious problems facing the African continent. In accordance with IMF/World Bank conditions, most of the African countries applying structural adjustment measures have retrenched large number of public-sector workers.

Kingdon and Knight (2001) studied the unemployment in South Africa using the probit model. The study was conducted using two national household surveys for the mid-1990s. The results indicate that unemployment in South Africa is determined by among others, race, education, age, gender, home ownership, location.

Similarly, Bhorat (2007) studied the unemployment in South Africa. He analyzed a number of labour economic and social choice theories and identified factors or common variables that determine the chance of somebody to be employed or not. A number of variables from economic and social theories that determine unemployment are.

- Shortage of highly educated individuals in many middle and low income countries
- Choices in how to utilize hours in the day
- Gender and culture
- High wage
- Composition of the household
- Marital status
- Wealth of the family or household.

Eita (2010) has investigated the causes of unemployment in Namibia for the period 1971 to 2007. The analysis is carried out through an extensive review of the relevant literature, microeconomic and macroeconomic models of unemployment. The unemployment model (with macroeconomic variables) is estimated using the Engle-Granger two-step econometric procedure. The results revealed that there is a negative relationship between unemployment and inflation in Namibia. Unemployment responds positively if actual output is below potential output, and if wages increase. An increase in investment causes unemployment to decrease significantly. The results provide evidence that the Phillips curve holds for Namibia and unemployment can be reduced by increasing aggregate demand. It is important to increase output up to the country"s potential, and there is a need for wage flexibility (workers need to reduce their wage demands) in order to decrease unemployment in Namibia. Increasing investment will reduce unemployment significantly.

Benefo and Pillai (2003) studied the determinants of female non-family work in Africa. They used demographic and health survey data from two African countries, Ghana and Zimbabwe for comparison purpose. The traditional kinship-oriented family organization in Africa, along with high fertility, has long been seen as factors that constrain women‟s participation in the labor-force, particularly in seeking formal sector employment. The authors found that education emerges as the most important determinant of non-family work. Even if female education levels increase, single women may not gain easy entry into the informal economy managed by kinship based social networks. A large proportion of these educated women may not find jobs if the formal economy does not expand. The study argues that a variety of factors, encompassing education, urbanization and family organization govern non-family work. To seriously address women‟s participation in formal sectoral employment, policy needs to target all aspects of the problem in a comprehensive manner. Family planning policies, for instance, attempt to reduce family size in order to enable women to enter the labor market, under the assumption that it is women with wider kin relationships that are handicapped. What this research shows is that the non-conjugal relationships become more important in affording access to non-family resources in the informal sector. Continuing focus on women‟s education is likely to improve women‟s formal sectoral employment. Unfortunately, in Africa a number of school and family related problems result in higher dropout rates for girls than for boys.

Bakare (2011) studied the determinants of urban unemployment in Nigeria. He uses time series secondary data and parsimonious error correction mechanism to test the relationship between the level of unemployment and demand for labour, supply of labour, population, inflation, capacity utilization, gross capital formation and nominal wage rate. His empirical investigations showed that the rising nominal wages and the accelerated growth of population which affected the supply side through a high and rapid increase in labour force relative to the absorptive capacity of the economy appear to be the main determinant of high unemployment in Nigeria.

Ledezma et al (2003) analyzed women in the Venezuelan labor market, focusing on their labor force participation and their income. They used aggregate data from the National Census since 1950. Their results showed that Venezuela is similar to other Latin American countries where older women, “cohabitors” and those with the lowest level of education increased their labor force participation, as a strategy to cope with reduced family income.

Cox and Psacharopoulos (1992) and Winter (1992) investigated the labor market behavior of Venezuelan women. Both studies used the same data source, the Household Survey data, for 1987 and 1989, respectively, and the same methodology. Both studies estimated a probit equation for a sample of working and nonworking women. Like other reviewed literatures, education has powerful effects on labor force participation in the two papers. The probit coefficients show that the probability rises steadily with each successive level of education. Cox and Psacharopoulos found that living in a rural area reduces the probability of participating in the labor force by 13 percent, considerably more than Winter who found only a 6 percent difference. Other results are related to specific variables used by each of the researchers. Cox and Psacharopoulos found that being a wife or partner reduces the probability of labor force participation by 22 percent, implying that family responsibilities compete for time spent in the market. Being head of household raises the probability by 23 percent. An increase in the income of other family members equivalent to 15 US$ per month reduces the probability of working by 4 percentage points. Winter finds that being a mother of young children significantly increases the probability that a woman will withdraw from the labor force. And, finally, age is also an important factor. The probability of women working increases steadily starting in the mid-twenties and peaks between the ages of 41 and 45. Low labor market participation rates among women in their early twenties are consistent with the high enrollment of women in this age group in higher education (44 percent). Moreover, many in their early twenties may be having babies.

2.4. Determinant of Women Unemployment in Ethiopia

Two empirical papers investigated the determinants of unemployment duration in urban Ethiopia: Serneels (2004) and Seife (2004). Both studies used the same data source and the same methodology. Not surprisingly, results of the papers are quite similar. The authors found that education has powerful effects on labor force participation as other literatures suggest.

Serneels (2004) studied determinants of unemployment duration in urban Ethiopia and the course of hazard, or the probability of leaving unemployment. He uses non-parametric as well as parametric and semi parametric estimation methods and control for unobserved heterogeneity. The authors found that age has a large negative effect as expected while education has a positive effect. Those with a father working in the public sector are more likely to leave unemployment early. This can be interpreted as an information effect, hiring practices, or a household welfare effect. His result also indicates that unemployment duration has no negative effect on the probability of leaving unemployment for the vast majority of unemployed. From a theoretical point of view, this can be explained by the presence of segmentation in the labour market.

Seife (2004) investigated unemployment duration in developing countries in the context of urban Ethiopia. The author used parametric and semi-parametric models to analyze the determinants of unemployment duration in a developing country context and data from a nationally representative urban household survey in Ethiopia. The study revealed that mean unemployment duration in urban Ethiopia is very long 3 years for completed spells and 4.7 years for incomplete spells. The author's econometric evidence shows that the hazard rate employment is significantly affected by age, marital status, and highest level of education attained, location and support mechanism while unemployed. Ethnic background and gender are not found to be important determinants. Appropriate tests show that the results are not driven by unobservable heterogeneities. The nonparametric hazard function and the baseline hazard retrieved from the semi-parametric estimation reveal a unique shape with alternating signs of duration dependence across a range of years.

Serneels (2007) investigated the nature of unemployment among young men in urban Ethiopia and found that it is concentrated among relatively well-educated first-time job seekers who aspire to a public sector job and spend on average close to four years in unemployment. This is consistent with a segmented labor market model where youngsters queue in unemployment for a good job, as confirmed by an empirical test of the theoretical prediction. The authors concluded that a negative (causal) relationship between household welfare and both the incidence and duration of unemployment, indicating that unemployment is concentrated among the relatively worse off urban households, which from a national perspective represent the middle classes. Job search through social networks is only effective after one has become unemployed, suggesting that networks provide insurance only after exposure to the risk. Generally, most variables have the same effect on both the incidence and duration of unemployment also confirming the queuing theory.

Chapter 3

3. Data and Methodology

3.1. Data Source

The dataset used in this study has been taken from the Ethiopia Demographic and Health Survey (EDHS) conducted by central Statistics Agency (CSA) in 2011. The 2011 EDHS is a nationally representative survey of women aged 15−49 from 17,817 households from 624 clusters throughout Ethiopia, 187 in urban areas and 437 in the rural areas. The survey utilized a multistage cluster sample based on the 1994 Population and Housing Census sample frame and was designed to obtain and provide information on the basic indicators of the health and demographic variables of interest for the following domains: Ethiopia as a whole, urban and rural areas of Ethiopia (each as a separate domain), and all geographic areas (nine regions namely: Tigray, Affar, Amhara, Oromiya, Somali, Benishangul-Gumuz, Southern Nations, Nationalities and Peoples (SNNP), Gambela and Harari regional states and two city administrations :Addis Ababa and Dire Dawa).

This multistage 2011 EDHS dataset is of hierarchical structure. The hierarchy for this study follows individuals/women as level-1, and regions as level-2. This means that individuals are nested in regions. From among the 17,817 households, 17,385 women were identified as eligible for the individual interview. Interviews were completed with 16,515 women, yielding a response rate of 95 percent. Thus, the analysis presented in this study on women about unemployment status is based on 16,386 women in Ethiopia.

3.2. Variables of the study

The dependent and independent variables that were considered to affect the status of employment of women were selected based on experiences from the available similar studies and the available data on the subject.

3.2.1. The Response Variable

The response variable of this study is unemployment status of women in Ethiopia. According to ILO's definition, those persons who are simultaneously "without work", "currently available for work" and "seeking work"" are considered as unemployed. For the purpose of our study, the response variable, unemployment status of women, is classified as unemployed women (those women who were not working for wage during the period of the survey) and otherwise employed women. Therefore, the outcome for the i^{th} woman is represented by a random variable Y_i with two possible values coded as 1 and 0. In view of this, the outcome of the i^{th} woman, Y_i was measured as a dichotomous variable.

$$Y_i = \begin{cases} 1, if\ the\ i^{th}\ women\ is\ unemployed \\ 0, otherwise \end{cases}$$

3.2.2. Explanatory Variables/Factors

Based on the reviewed literatures, some of the common predictors that are expected to influence the unemployment status of women in Ethiopia were recorded as given below for the purpose of the analysis.

TABLE3.1: Demographic and Socioeconomic Variables

Variables	Description	Values/Categories‟
REGION	Region	1=Tigray 2=Affar 3=Amhara 4=Oromiya 5=Somali 6=Ben-Gumuz 7= SNNP 8 = Gambela 9 =Harari 10= Addis Ababa, 11= Dire Dawa
PLRESIDE	Place of Residence	0= Rural 1= Urban
MARITAL	Marital Status	0=Married 1=Others
EDUCATION	Women‟s Education Level	0=No Education 1=Primary 2=Secondary+
MASS MEDIA	Exposure to any Mass Media	0=Not at All 1= Less a Weak 2=At Least Once a Week
HEDUC	Partner‟s Education Level	0=No Education 1=Primary 2=Secondary+
ECOSTATUS	Economic Status Of Household	0=Poor 1=Medium 2= Rich
ADDICTION	Drug Addiction	0= No 1=Yes
AGE	Age of Women	0=Less Than 25 1=Between 25-34 2=Above 34
HHSIZE	Household Size	0=Less Than or equal to 5 1=6-10 2=Above 10
PRCHILD	Presence of Child under 5 years of age	0=No 1=Yes
SEXHHEAD	Sex of Head of Household	0=Female 1= Male
PREGNANCY	Pregnancy Of Women	0=No or Unsure 1=Yes
MIGRATION	Migration Status	0= Visitors 1= Usual Residence
HOCCU	Occupation of Husband/Partners	0= Not Working 1= Agric Employee 2=Non-Agric Employee
DECISION	Decision Autonomy	0=No, 1=Yes

3.3. The Methodology

3.3.1. Logistic Regression Analysis

Logistic regression is a popular modeling approach when the dependent variable is dichotomous or polytomous. This model allows one to predict the log odds of outcomes of a dependent variable from a set of variables that may be continuous, discrete, categorical, or a mix of any of these. Hosmer and Lemeshow (2000) have described logistic regression focusing on its theoretical and applied aspect.

Often the outcome variable in social data is, in general not continuous, instead is binary. In such a case, binary logistic regression is a useful way of describing the relationship between one or more independent variables and a binary outcome variable that has only two possible values. Indeed, a generalized linear model is used for binary logistic regression. The most attractive feature of a logistic regression model is that it neither assumes linearity in the relationship between the covariates and the outcome variable, nor does it require normally distributed variables. It also does not assume homoscedasticity and in general has less stringent requirements than linear regression models. Thus logistic regression is used in a wide range of applications leading to binary dependent data analysis (Hilbe, 2009; Agresti, 2002).

Logistic model, as compared to its competitor, the probit model, is less sensitive to outliers and easy to correct a bias (Copas, 1988). In instances where the independent variables are categorical or a mix of continuous and categorical, logistic analysis is preferred to discriminant analysis (Agresti, 2007). The assumptions required for statistical tests in logistic regression are far less restrictive than those for ordinary least squares regression. There is no formal requirement for multivariate normality, homoscedasticity, or linearity of the independent variables within each category of the response variable. However, the assumptions that apply to logistic regression model include: meaningful coding, inclusion of all relevant and exclusion of all irrelevant variables in the regression model and low error in the explanatory variables.

3.3.1.1. Model

In the terminology of logistic regression analysis the odds of a success is defined to be the ratio of the probability of a success to the probability of a failure.

Let Y be an n×1 vector of response variable with $y_i = 1$ if the i^{th} women under study is unemployed and $y_i = 0$ if the i^{th} women is employed, X is an n × (k+1) design matrix of explanatory variables, β is a (k+1) ×1 vector of parameters. The data layout for X is given as follows.

$$X=\begin{bmatrix} 1 & X_{11} & X_{12} & \cdots X_{1k} \\ 1 & X_{21} & X_{22} & \cdots X_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots X_{nk} \end{bmatrix}$$

X without the leading column of 1s, is termed as predictor data matrix. Let the conditional probability that the outcome is present (probability of success) be given by:

$$\pi = P(Y=1/X) = \frac{exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)}{1+exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)} = \frac{exp(X'\beta)}{1+exp(X'\beta)} \ldots\ldots \quad (3.1)$$

Using (3.1), we obtain the odds of success as

$$odds(Y=1)=\frac{\pi}{1-\pi}=\exp(X'\beta)\text{...}(3.2)$$

In logistic regression analysis, it is assumed that the explanatory variables affect the response through a suitable transformation of the probability of the success. This transformation is a suitable link function of π, and is called the logit-link, which is defined as:

$$\log it(\pi)=\log\left(\frac{\pi}{1-\pi}\right)=\log\left(e^{\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k}\right)=\log\left(e^{X'\beta}\right)\text{..........}(3.3)$$

The transformed variable, denoted by logit (π) is the log-odds and is related to the explanatory variables as:

$$\log it(\pi)=\eta(x)=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k=X'\beta \qquad (3.4)$$

where β' = (β_0, β_1, β_2… β_k) are the model parameters and X' = (X_0, $X_{1\ldots}$ X_k) with X_0 =1, are explanatory variables.

The above equations give suitable representations of the success probability, odds, and log-odds. Indeed, these representations facilitate interpretations of parameter estimates. The parameter β_i refers to the effect of X_i on the log odds that Y = 1, controlling the other X"s in the model.

3.3.1.2. Parameter Estimation

The most commonly used method of estimating the parameters of a logistic regression model is the method of Maximum Likelihood (ML) instead of Ordinary Least Squares

(OLS) method. Mainly for this reason the ML method based on Newton-Raphson iteratively reweighted least square algorithm becomes more popular with researchers (Ryan, 1997). The sample likelihood function is, in general defined as the joint probability function of the random variables whose realizations constitute the sample. Specifically, for a sample of size n whose observations are (y_1, y_2 ... y_n), the corresponding random variables are (Y_1, Y_2… Yn). Since the Y_i is a Bernoulli random variable, the probability mass function of Y_i is:

$$f_i(y_i) = \pi_i^{y_i} \ (1-\pi_i)^{1-y_i} \quad \text{……………………………………………………} \quad (3.5)$$

$Y_i = 0$ or 1 and i=1, 2,…, n

Since the observations are assumed to be independent, the likelihood function is obtained as the product of the terms given in expression (3.5) as follows:

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i} \left[1-\pi(x_i)\right]^{1-y_i}$$

The principle of maximum likelihood states that we use as our estimate of β the value which maximizes the likelihood function. However, it is easier mathematically to work with the log likelihood function. The *log likelihood* is :

$$L(\beta) = ln[l(B)] = \sum_{i=1}^{n} y_i \, ln\left(\frac{\pi_i}{1-\pi_i}\right) + \sum_{i=1}^{n} ln(1-\pi_i) \quad \text{………………………………..} \quad (3.6)$$

To find the value of β that maximizes L (β) we differentiate L (β) with respect to β and set the resulting expressions equal to zero. These equations, known as the *likelihood equations, a*re:

$$\sum_{i=1}^{n}\left[y_i - \pi(x_i)\right] = 0$$

and

$$\sum_{i=1}^{n} x_{ij}\left[y_i - \pi(x_i)\right] = 0 \quad \text{......} \quad (3.7)$$

For logistic regression the expressions in equation (3.7) are nonlinear in β and thus require special methods for their solution. These methods are iterative in nature and have been programmed into available logistic regression software. As such, it represents the fitted or predicted value for the logistic regression model. An interesting consequence of equation (3.7) is that

$$\sum y_i = \sum \hat{\pi}_i$$

That is, the sum of the observed values of y is equal to the sum of the predicted (expected) values.

In fact, the maximum likelihood estimates of β in the multiple binary logistic regression models are those values of β that maximize the log-likelihood function given in equ (3.6). No closed form solution exists for the values of $\hat{\beta}$ that maximize the log-likelihood function. Computer-intensive numerical search procedures are therefore required to find the maximum likelihood estimates $\hat{\beta}$ and hence $\hat{\pi}$, because the multiple logistic regression model computes the probability of the selected response as a function of the values of the predictor variables. There are several widely used numerical search procedures, one of these employs iteratively reweighted least squares algorithm. In this study, we shall rely on standard statistical software programs specifically designed for logistic regression to obtain the maximum likelihood estimates of parameters.

3.3.1.3. Statistical Tests of Individual Predictors.

The statistical significance of individual regression coefficients is tested using the Wald and Score chi-square statistic. The Wald statistic is a test which is commonly used to test the significance of the individual logistic regression coefficients for each independent variable (that is, to test the null hypothesis in logistic regression that a particular logit (effect) coefficient is zero i.e. $H_0 : \beta_i = 0$ against $B_i \neq 0$.

The Wald test is based on the behavior of the log-likelihood function at the ML estimate $\hat{\beta}$, having chi-squared form. The standard error of the estimate depends on the curvature of the log-likelihood function at the point where it is maximized, with greater curvature giving smaller SE values. For a dichotomous dependent variable, the Wald statistic is:

$$W = \left[\frac{\hat{\beta}}{SE\left(\hat{\beta}\right)} \right]^2 \qquad \text{(3.8)}$$

Under the null hypothesis for large sample size, this statistic has an approximate chi-square distribution with one degree of freedom.

The score test or Lagrange Multiplier is based on the behavior of the log-likelihood function at the null value for β of 0. It uses the size of the derivative (slope) of the log-likelihood function evaluated at the null hypothesis value of the parameter. The derivative at β = 0 tends to be larger in absolute value when $\hat{\beta}$ is further from that null value.

3.3.1.4. Overall Model Evaluation

A logistic model is said to provide a better fit to the data if it demonstrates an improvement over the intercept-only model (also called the null model). An intercept-only model serves as a good baseline because it contains no predictors. Consequently, all explanatory variables are added to the model. An improvement over this baseline is examined by using three inferential statistical tests: **the likelihood ratio, score, and Wald tests**.

The Likelihood Ratio (LR) test is performed by estimating two models and comparing the fit of one model to the fit of the other. Removing predictor variables from a model will almost always make the model fit less well (i.e., a model will have a lower log likelihood), but it is necessary to test whether the observed difference in model fit is statistically significant. The likelihood ratio test does this by comparing the log likelihoods of the two models, if this difference is statistically significant, then the less restrictive model (the one with more variables) is said to fit the data significantly better than the more restrictive model. If one has the log likelihoods from the models, the likelihood ratio test is fairly easy to calculate. The likelihood ratio test is performed to test the overall significance of all coefficients in the model on the basis of test statistic:

$G = [(-2\ln L_o)-(-2\ln L_1)]$

where, L_0 is the likelihood of the null model and L_1 is the likelihood of the saturated model. The statistic G is distributed chi-squared with degrees of freedom equal to the difference in the number of degrees of freedom between the two models and plays the same role in logistic regression as the numerator of the partial F-test does in linear regression.

3.3.1.5. Hosmer-Lemeshow Goodness-of-Fit test

The Hosmer-Lemeshow test is one of the recommended tests for overall fit of a binary logistic regression model. This goodness-of-fit statistic is used to assess the fit of a logistic regression model. Hosmer and Lemeshow‟s goodness of fit test divides subjects into deciles based on predicted probabilities and then computes a chi-square from observed and expected frequencies. Using this grouping strategy, the Hosmer-Lemeshow goodness-of-fit statistic, $\hat{C}$ is obtained by calculating the Pearson chi-square statistic from the g×2 Table of observed and estimated expected frequencies. A formula defining the calculation of $\hat{C}$ is as follows:

$$\hat{C} = \frac{\sum_{k=1}^{g}(O_k - E_k)^2}{V_k} \qquad (3.9)$$

where $E_k = n P_k$, $V_k = nP_k(1-P_k)$, g is the number of group, O_k is observed number of events in the k^{th} group, E_k is expected number of events in the k^{th} group, and V_k is a variance correction factor for the k^{th} group. This statistic has an approximate chi-square distribution with (g-2) degrees of freedom. If the calculated value of the Hosmer-

Lemeshow goodness-of-fit test statistic is greater than 0.05, we will not reject the null hypothesis that there is no difference between observed and model-predicted values, implying that the model estimates are adequate to fit the data at an acceptable level.

3.3.1.6. Validations of Predicted Probabilities.

The resultant predicted probabilities using logistic model can be revalidated with the actual outcome to determine if high probabilities are indeed associated with events and low probabilities with nonevents. The degree to which predicted probabilities agree with actual outcomes is expressed as either a measure of association or a classification Table. The predictive ability of the model was assessed using the following four measure of association.

1. **Somers' D** - Somer's D is used to determine the strength and direction of relation between pairs of variables. Its values range from -1.0 (all pairs disagree) to 1.0 (all pairs agree). It is defined as

   $$\text{Somers D}=\frac{n_c - n_d}{t} \text{ --(3.10)}$$

   where n_c is the number of pairs that are concordant, n_d the number of pairs that are discordant, and t is the total number of pairs with different responses.

2. **Gamma** - The Goodman-Kruskal"s Gamma method does not penalize for ties on either variable. Its values range from -1.0 (no association) to 1.0 (perfect association). Because it does not penalize for ties, its value will generally be greater than the values for Somer's D. Its value is calculated as follows.

$$\text{Gamma}=\frac{nc-nd}{nc+nd} \text{ ---(3.11)}$$

3. **Tau-a**: Kendall's Tau-a is a modification of Somer's D that takes into account the difference between the number of possible paired observations and the number of paired observations with a different response. It is defined to be the ratio of the difference between the number of concordant pairs and the number of discordant pairs to the number of possible pairs.

$$\text{Tau-a} = \frac{2(nc-nd)}{N(N-1)} \text{ --(3.12)}$$

Usually this value is much smaller than Somer's D since there would be many paired observations with the same response.

4. **C**: it is equivalent to the well known measure ROC. **C** ranges from 0.5 to 1, where 0.5 corresponds to the model randomly predicting the response, and a 1 corresponds to the model perfectly discriminating the response.

$$C = \frac{[nc+0.5(t-nc-nd)]}{t} \text{ ---(3.13)}$$

where N is the total number of observation, i.e. number of women, in the study, t is the number of women pairs having different response values, n_c is the number of pairs which are concordant and n_d is the number of discordant. A pair with different response is said to be concordant (discordant) if the larger response has a higher (lower) predicted event probability than the smaller response).

3.3.1.7. Model Diagnostics

Regression model building is often an iterative and interactive process. The first model we try may prove to be inadequate. Regression diagnostics are used to detect problems with the model and suggest improvements.

There are three ways that an observation can be considered as unusual, namely outlier, influence and leverage. In logistic regression, observations whose values deviate from the expected range and produce extremely large residuals and may indicate a sample peculiarity are called **outliers**. These outliers can unduly influence the results of the analysis and lead to incorrect inferences. An observation is said to be **influential** if removing the observation substantially changes the estimate of coefficients. Influence can be thought of as the product of leverage and outliers. An observation with an extreme value on a predictor variable is called a point with high leverage. **Leverage** is a measure of how far an independent variable deviates from its mean. In fact, the leverage indicates the geometric extremeness of an observation in the multi-dimensional covariate space. These leverage points can have an unusually large effect on the estimate of logistic regression coefficients (Cook, 1998).

To identify if an observation is outlier or influential, the following rules of thumbs were employed in this study.

- **Residuals: S**tandardized, Standard, deviance and Pearson residuals are obtained using SAS. Observations with values larger than 3 in absolute value are considered as outliers (Agresti, 2007).

- **Leverage Values (Hat Diag):** Measure of how far an observation is from the others in terms of the levels of the independent variables (not the dependent variable). Observations with values larger than one are considered to be potentially highly influential (Belsley et al., 1980).

- **DFBETAS:** Measure of how much an observation has affected the estimate of a regression coefficient (there is one DFBETA for each regression coefficient, including the intercept). Values larger than $\frac{2}{\sqrt{n}}$ in absolute value are considered highly influential.

- **Cook's D**: Measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. In logistic regression, a case is identified as influential if its Cook's distance is greater than 1.0 (Hosmer-Lemeshow, 2000).

3.3.2. Multilevel Logistic Regression Model

A multilevel logistic regression model also referred to in the literature as a hierarchical model, can account for lack of independence across levels of nested data (e.g., women nested within regions). Standards logistic regression assumes that all experimental units (in this case, women) are independent in the sense that any variables affecting the dependent variable have the same effect in all regions. Multilevel modeling relaxes this assumption and allows these variables" effects to vary across regions.

In many applications, data have hierarchical or clustered structures, such as medical and health services research where patients are clustered within hospitals, in educational studies where students are nested within schools, or work status of women studies where women are nested within regions. These studies often involve the analysis of data with complex patterns of variability, such as multilevel, nested sources of variability. For example, in access of jobs of regions, there may be variability between the regions as well as between women who are nested within the regions.

Hierarchical models are statistical models that can be used to analyze nested sources of variability in hierarchical data, taking account of the variability associated with each level of the hierarchy. These models have also been referred to as multilevel models, mixed models, random coefficient models, and covariance component models (Breslow and Clayton, 1993; Longford, 1993; Snijders and Bosker, 1999; Hox, 2002; Goldstein, 2003).

As mentioned above, multilevel logistic regression analysis is a methodology for analysis of data with complex patterns of variability, with a focus on nested sources of variability on multiple categories. For this study the variation of unemployment status of women within regions means that not only unexplained variation between women but also unexplained variation between regions. This can be expressed by statistical models with random coefficients.

3.3.2.1. Two level model

For simplicity of presentation two-level models are used for this study, i.e., models accounting for women-level and regional-level effects. In this data structure, level-1 is the women level and level-2 is the regional level. Within each level-2 unit there are n_j women in the j^{th} region. We further simplify the presentation by assuming there is a women-level predictor and regional-level factor.

To provide a familiar starting point, we first consider a two-level model for binary outcomes with a single explanatory variable. Conceptually, the basic (two level) multilevel model for a binary response is equivalent to model (3.2) except for the notation in the outcome variable. Suppose we have data consisting of women (level one) grouped into regions (level two). Let Y_{ij} be the binary response for a woman i in region j and X_{ij}, an explanatory variable at the women level. We define the probability of the response equal to one as P_{ij} = Pr (y_{ij} = 1) and let P_{ij} be modeled using a logit link function. The standard assumption is that y_{ij} has a Bernoulli distribution. Then the two-level model can be written as

$$log\left[\frac{P_{ij}}{1-P_{ij}}\right] = \beta_0 + \beta_1 X_{ij} + U_j \text{...} (3.14)$$

where $U_j \sim IID(0, \sigma_u^{\,2})$. U_j is the random effect at level two. Without U_j, Equation (3.14) would be a standard logistic regression model. Conditional on U_j, the Y_{ij} is assumed to be independent. The model (3.14) is often described as follows.

$$log\,it(P_{ij}) = log\left[\frac{P_{ij}}{1-P_{ij}}\right] = \beta_{0j} + \beta_1 X_{ij}$$ [Level 1 model]

and, $\beta_{0j} = \beta_o + U_j$ [Level 2 model]

3.3.2.2. Heterogonous Proportion

For the proper application of multilevel analysis, the first logical step is to test heterogeneity of proportions between groups. The most commonly used test statistic to check for heterogeneity of proportions between groups is the chi-square. To test whether there are indeed systematic differences between the groups, the well-known chi-square test can be used. The test statistic of the chi-squared test for contingency Table is often given in familiar form:

$$\chi^2 = \frac{\sum_i (O-E)^2}{E} \text{--}(3.15)$$

where O is the observed and E is the expected counts in the cell of contingency Table. It can also be written as:

$$\chi^2 = \sum_{j=1}^{g} n_j \frac{(\bar{Y}._j - \hat{P}.)^2}{\hat{P}.(1-\hat{P}.)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.16)$$

where the group average, $\bar{Y}.j = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ij}$ is the proportion of successes in group j which is an estimate for the group-dependent probability P_j. Similarly, the overall average $\hat{P}. = \bar{Y}.. = \frac{1}{n}\sum_{j=1}^{g}\sum_{i=1}^{n_j} Y_{ij}$ here is the overall proportion of successes.

The decision will be based on the chi-square distribution with (g-1) degrees of freedom. This chi-square distribution is an approximation valid if the expected number of success and of failures in each group, $n_j \bar{Y}._j$ and $n_j(1-\bar{Y}._j)$ respectively, all are at least 1 while 80 percent of them are at least 5 (Agresti, 2002).

Estimation of between and within group variance: the theoretical variance between the group dependent probabilities, i.e., the population value of Var (P_j), can be estimated by:

$$\tau^2 = S^2{}_{between} - \frac{S^2{}_{within}}{\tilde{n}} \quad \text{......} \quad (3.17)$$

where $\tilde{n}$ is given by:

$$\tilde{n} = \frac{1}{n-1}\left\{ n - \frac{\sum_{j=1} n^2{}_j}{n} \right\} = \bar{n} - \frac{S^2(n_j)}{g\bar{n}}.$$

and

$$n = \sum_{j=1}^{g} n_j$$

For dichotomous outcome variables, the observed between-groups variance is closely related to the chi-squared test statistic (3.16). They are connected by the formula

$$S^2{}_{between} = \frac{\hat{P}.(1-\hat{P}.)}{\bar{n}(N-1)} \chi^2 \quad \text{......} \quad (3.18)$$

The within-group variance in the dichotomous case is a function of the group averages, via,

$$S^2_{within} = \frac{1}{M-N}\sum_{j=1}^{N} n_j \overline{Y}._j (1-\overline{Y}._j) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.19)}$$

3.3.2.3. The Random Intercept-Only Model

The empty two-level model for a dichotomous outcome variable refers to a population of groups (level-two units, i.e. regions)) and specifies the probability distribution for group-dependent probabilities without taking further explanatory variables into account. This model only contains random groups and random variation within groups. It can be expressed with logit link function as follows.

$$\operatorname{logit}(P_j) = \beta_0 + U_{0j}$$

$$U_{0j} \sim IID\left(0, \sigma^2{}_0\right)$$

where β_0 is the population average of the transformed probabilities and U_{0j} is the random deviation from this average for group j.

3.3.2.4. **The Random Intercept and Fixed Slope Model**

In the random intercept logistic regression model, the intercept is the only random effect meaning that the groups differ with respect to the average value of the response variable. It represents the heterogeneity between groups in the overall response.

The logistic random intercept model expresses the log-odds, i.e. the logit of P_{ij}, as a sum of a linear function of the explanatory variables and a random group-dependent deviation U_{0j}. That is,

$$log\,it(P_{ij}) = log(\frac{P_{ij}}{1-P_{ij}}) = \beta_{0j} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \ldots + \beta_k X_{kij}$$

$$log\,it(P_{ij}) = \beta_{0j} + \sum_{h=1}^{k} \beta_h X_{hij} \quad \ldots\ldots (3.20)$$

where the intercept term β_{0j} is assumed to vary randomly and is given by the sum of an average intercept β_0 and group- dependent deviations U_{0j}. That is:

$$\beta_{0j} = \beta_0 + U_{0j}$$

As a result

$$log\,it(P_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h X_{hij} + U_{0j} \quad \ldots\ldots (3.21)$$

$$U_{0j} \sim IID\left(0, \sigma^2{}_0\right)$$

where β_0 is the log-odds that y = 1 when x = 0 and u = 0, β_h is the effect on log-odds of dependent variable in same group (same value of u), exp (β_h) is an odds ratio, comparing odds for individuals in the same group. U_{0j} is the effect of being in group j on the log-odds that y = 1 also known as a level 2 residual, $\sigma^2{}_0$ is the level 2 (residual) variance, or the between-group variance in the log-odds that y = 1 after accounting for X.

Note that the first part of the left-hand side of (equ 3.21), incorporating the regression coefficients, $\beta_0 + \sum_{h=1}^{k} \beta_h X_{hij}$ is the fixed part of the model, because the coefficients are fixed. The remaining part, U_{oj}, is called the random part of the model.

The region intercepts measure the differences between the regions, controlling for other effects in the model. Equation (3.21) is a mixed model because it has both fixed effects and random effects. It is a logistic mixed model, because the link function is logit, and thus, a member of the family of generalized linear mixed models (GLIMMIX).

3.3.2.5. The Random Coefficient Model

So far, we have allowed the probability of unemployment to vary across regions, but we have assumed that the effects of the explanatory variables are the same for each region. We will now modify this assumption by allowing the difference between explanatory variables within a region to vary across regions. To allow for this effect, we will need to introduce a random coefficient for those explanatory variables. So, a random coefficient model represents heterogeneity in relationship between the response and explanatory variables.

As mentioned above, the response variable in this study, unemployment status was binary. Therefore the statistical model used in this analysis was the two-level random coefficient multilevel regression model. The model, with p level-1 predictors and q level-2 predictors, can be expressed as:

$$log\,it(\pi_{ij}) = log\left[\frac{\pi_{ij}}{1-\pi_{ij}}\right] = \beta_{0j} + \sum_{h=1}^{p}\beta_{hj}X_{hij} + \sum_{h=1}^{q}U_{hj}X_{hj}$$

Where

$\beta_{0j} = \beta_0 + U_{0j}$, i = 1,2,3,..,n_j, j = 1,2,3,……,11

Now the above equation is written as

$$log\,it(\pi_{ij}) = log\left[\frac{\pi_{ij}}{1-\pi_{ij}}\right] = \beta_0 + \sum_{h=1}^{p}\beta_h X_{hij} + U_{oj} + \sum_{h=1}^{q} U_{hj} X_{hj}$$ ------------------- (3.22)

$Y_{ij} \sim$ Binary (π_{ij}), $U_{0j} \sim$ Normal (0, σ^2_0)

The first part of equation (3.25), $\beta_0 + \sum_{h=1}^{p}\beta_h X_{hij}$, is called the fixed part of the model. The second part $U_{0j} + \sum_{h=1}^{q} U_{hj} X_{hj}$ is called the random part.

Y_{ij} is the value of the response variable for the i^{th} woman in the j^{th} region. X_{hij} is the value of individual-level explanatory variable Xp for the i^{th} woman in the j^{th} region. X_{hij} is the value of region-level explanatory variable Xq for the j^{th} region. U_{0j} is the region-level residual of the j^{th} region. It is a random effect that represents the discrepancy between β_0 and the true intercept of the jth region and n_j is the number of women respondents in the jth region.

3.3.2.6. Parameter Estimation

The estimation method of most of the statistical models (e.g., linear regression, logit models, and log-linear models) regularly used are well established. But this is not true for multilevel models for binary data. Parameter estimation for multilevel logistic model is not straightforward like the methods for simple logistic regression model. In this paper we are focused on hierarchical logistic regression models, which can be fitted using the new SAS procedure GLIMMIX. Proc GLIMMIX is developed based on the GLIMMIX macro and provides highly useful tools for fitting generalized linear mixed models, of which the hierarchical logistic model is a special case. Therefore, in this study, the parameters were estimated using maximum likelihood method using Proc GLIMMIX of SAS 9.2.

3.3.2.7. Significance Testing in Multilevel Regression

As with ordinary least squares regression or logistic regression, we can consider significance tests for individual estimates, such as intercepts, slopes, and their variances, as well as whether the full model accounts for a significant amount of variance in the dependent variable. In between, there is also the possibility of determining whether of subset of predictors contribute significantly (Newsom, 2011).

3.3.2.7.1. Significance Testing for Fixed Effects

The fixed effects in multilevel regression are typically tested in a familiar way, by creating a ratio of the intercept or slope estimate to the estimate of the standard error (Newsom, 2011). The usual null hypothesis test is whether the coefficient, either intercept or slope, is significantly different from zero (i.e., is the population value zero or not). This kind of ratio, usually distributed as a *z* or *t*, is used in many statistical tests (Bryk and Raudenbush, 2002). For this study, the significance of fixed effect is tested using the statistic:

$$t = \frac{\hat{\beta}_h}{S.E(\hat{\beta}_h)} \qquad (3.23)$$

where $\hat{\beta}_h$ is either the intercept or slope coefficient and S.E. ($\hat{\beta}_h$) is the standard error estimate. The degrees of freedom used in this analysis are between-within. The DDFM=BETWITHIN option divides the residual degrees of freedom into between-subject and within-subject portions. PROC GLIMMIX then determines whether a fixed effect changes within any subject. If the GLIMMIX procedure does not process the data by subjects, the DDFM=BETWITHIN option has no effect. If so, it assigns within-subject degrees of freedom to the effect; otherwise, it assigns the between-subject degrees of freedom to the effect. If there are multiple within-subject effects containing classification variables, the within-subject degrees of freedom are partitioned into components corresponding to the subject-by-effect interactions.

3.3.2.7.2. Significance Testing for Random Effects

Random effects tests examine hypotheses about whether the variance of intercept or slopes (or their covariances) is significantly different from zero. The tests of variances and covariances are made using a Wald z-test and chi-square test. The Wald test for variances is simply a ratio of the variance estimate divided by the standard error estimate. Significance tests of variances (but not covariances) using this approach should be interpreted after dividing the P-Value from the output in half (i.e., as a one-tailed test; Snijders & Bosker, 1999).

3.3.2.8. Goodness of Fit Test

It is useful to be able to judge whether a model is a good fit to the data. For this study, test of goodness of fit were employed using the deviance. The maximum likelihood procedure produces a statistic called the deviance, which indicates how well the model fits the data. The test compares the deviance (-2 log likelihood) of two models by subtracting the smaller deviance (model with more parameters) from the larger deviance (model with lower parameters). The difference is a chi-square test with the number of degrees of freedom equal to the number of different parameters in the two models. Similarly, the overall model evaluation is also examined using Akaike Information Criteria (AIC) and Schwartz Information Criteria (BIC). The smaller the value, the better of the model will be.

3.3.2.9. Model Diagnostic

It is of interest to obtain the residual values from the estimated multilevel model. Plots are a good way to examine the residuals. But in multilevel logistic regression, many different residual plots can be used. For this study, the fitted model was checked for possible presence of outliers and influential values in a similar fashion with standard logistic model. But additionally, the presence of outliers and influential observation were examined for level two. The value of standardized residuals greater than 3 in absolute value is considered as an outlier for both level one and level two. Leverage and influence value greater than one is considered as an influential observation for both level one and level two.

CHAPTER 4

4. STATISTICAL DATA ANLAYSIS

The data analysis is done using SAS 9.2 and MLwiN 2.25 statistical (software) packages. The results of the analysis are divided into the following sections: descriptive analysis results, results of logistic analysis and results of multilevel analysis. These results and their discussions are presented in the following sections.

4.1. DESCRIPTIVE ANALYSIS

The major demographic and socioeconomic background characteristics of the respondents with unemployment are presented in Table 4.1 below. The total number of women covered in the present study is 16386. Among these, 10,557 (64.42%) were unemployed where as 5,829(35.58%) were employed at the date of the survey.

Table 4.1 Distribution of Demographic and Socioeconomic factors

Variables	Unemployment			
	No	Yes	Total	Percentage of Unemployment
Region				
Tigray	501	1223	1724	70.94
Affar	133	1085	1218	89.08
Amhara	612	1469	2081	70.59
Oromiya	864	1270	2134	59.51
Somali	143	740	883	83.81
Benishangul-Gumuz	540	715	1255	56.97
SNNP	821	1212	2033	59.62
Gambela	442	688	1130	60.88
Harari	432	663	1095	60.55
Addis Ababa	921	818	1739	47.04
Dire Dawa	420	674	1094	61.61
Age				
Less than 25	2004	4772	6776	70.43
Between 25-34	2112	3130	5242	59.71
Above 34(ref.)	1713	2655	4368	60.78
Place of Residence				
Rural	3231	7813	11044	70.74
Urban(ref.)	2598	2744	5342	51.37
Educational Level				
No education	2443	5752	8195	70.19
Primary	2137	3665	5793	63.27
Secondary and above (ref.)	1249	1149	2398	47.91
Migration Status				
Visitor	199	349	548	63.69
Usual residence (ref.)	5630	10208	15838	64.45
Household size				
Less or equal to 5	3215	5015	8230	60.94
Between 6-10	2427	5123	7550	67.85
Above 10(ref.)	187	419	606	69.14
Sex of Household Head				
Female	2037	2562	4599	55.71
Male(ref.)	3792	7995	11787	67.83
Mass Media				
Not at all	1474	4252	5726	74.26
Less a week	1187	2262	3449	65.58
At least once a week(ref.)	3168	4043	7211	56.07
Economic Status HH				
Poor	1490	4518	6008	75.20

Medium	772	1502	2274	66.05
Rich (ref.)	3567	4537	8104	55.98
Pregnancy				
No	5492	9610	15102	63.64
Yes (ref.)	337	947	1284	73.75
Marital Status				
Married	2947	6467	9414	68.69
Others(ref)	2882	4090	6972	58.67
Husband Education				
No education	1652	4193	5845	71.74
Primary	3202	5396	5845	92.32
Secondary and Above (ref.)	975	968	1943	49.82
Husband Occupation				
Not working	1697	3023	4720	64.05
Agric employee	2257	5487	7744	70.85
Non Agric –employee(ref.)	1875	2047	3922	52.19
Decision				
No	2880	5252	8132	64.58
Yes (ref.)	2949	5305	8254	64.27
Presence of child (5 years or less)				
No	3358	5216	8574	60.84
Yes	2471	5341	7812	68.37
Drug Addiction				
No	5628	10268	15896	64.59
Yes(ref)	201	289	490	58.97

Source: Own calculation based on the 2011 EDHS report

The proportion of unemployment status of women varied from one region to the other in Ethiopia. For example, the highest percentage of unemployed women was observed in Affar (89.08%) followed by Somali (83.81%) where as the lowest percentage was recorded in Addis Ababa (47.04%) and followed by Benishangul-Gumuz (56.97%). Hence, there appears to be some variation in the proportion of unemployment status amongst women in different regions.

The proportion of unemployed women, observed in Table 4.1 also differs with their age groups. For instance, higher proportion of unemployment was observed for women under

25 years of age (70.43%) and the lowest proportion of unemployment of women was found in the age group between 25-34 years (59.70%).

Similarly, the proportion of unemployed women, observed in Table 4.1 differs by type of place of residence: urban and rural. Accordingly, higher numbers of unemployed women (70.74 %) resided in rural areas, and relatively small number of unemployed women (51.37 %t) resided in urban areas.

Table 4.1 also reveals that the unemployment status of women varies by their educational status. The highest percentage of women unemployment was observed in women who have no education (70.19%) as opposed to the lowest percentage of unemployment which was recorded for women who have secondary and above education level (47.91%). Similarly, we see that nature of unemployment of women varies by their exposure to mass media and head of household. The highest percentage of women unemployment was observed for women who have no any exposure to mass media (74.26%) as opposed to the lowest percentage of unemployment which was recorded for women with an exposure of at least once a week (56.07%). With regard to head of household head, higher percentage of women unemployment (67.83%) reside in male headed households and relatively small percentage of women unemployment (55.71%) reside in female headed households.

Unemployment status of women also varies according to household size and migration status. A higher percentage of women unemployment was observed in households of size above ten (69.14%) as opposed to the lowest percentage of unemployment was observed for households in which household size is less or equal to five (60.94%). About 64.45%

unemployed women have permanent residence or non migrants while 63.69% of unemployed women are migrants. This figure indicates that the proportion of unemployed women who have permanent residence and visitors are almost similar.

Table 4.1 also shows that the proportion of unemployment status of women vary by households‟ economic status, marital status and nature of pregnancy. The highest percentage of unemployment was observed among women from poor households (75.2%) as opposed to the lowest percentage of women unemployment recorded for women residing in rich households (55.98%). With regard to marital status, the percentage of unemployment was observed among married women (68.70%) is higher than the percentage of unemployment recorded for unmarried women (58.66%). Similarly the percentage of unemployed women was observed among pregnant women (73.75%) than non pregnant women (63.64%).

Table 4.1 shows that the proportion of unemployment status of women varies by husband/partners educational level and employment status. The highest percentage of women-unemployment was observed among those women whose husband/partners have primary education (92.32%) as opposed to the lowest percentage of women unemployment which was recorded from women whose husband/partners have secondary and above education level (49.32%). With regard to husband/partners employment status, the highest percentage of women-unemployment was observed among those women whose husband/partners are working in the agricultural sector (70.85%) and lowest percentage of women unemployment was recorded in those women their husband/partners are non- agricultural employee (52.19%).

With regard to the presence of child of less than or equal to five years of age in household, the greater number of unemployed women was observed in those households who have women of less than or equal to five years (68.37%) and a lower percentage was observed in those households with no child of less than or equal to 5 years (60.83%) in the household.

Table 4.1 shows that the proportion of unemployment status of women varies by their nature of addiction and power of decision making in the household. A higher proportion of unemployed women were observed in those women with no drug addictions like cigarette, tobacco, chat, etc (64.59%) and a lower percentage was observed in those addicted women (58.98%). Regarding to decision power in the household, the percentage of unemployed women who were making decisions in the household (64.27%) is almost the same as among those women who were not making decisions in the household (64.58%).

4.2. DETERMINANTS OF WOMEN UNEMPLOYMENT IN ETHIOPIA: LOGISTIC REGRESSION ANALYSIS

In this section, the logistic regression analysis results obtained by using stepwise inclusion of variables, overall model evaluation, statistical tests of individual predictors, goodness-of-fit statistics, and validations of predicted probabilities are presented.

The Initial Log Likelihood Function, (-2 Log Likelihood or -2LL) is a statistical measure like total sums of squares in regression. If the independent variables have a relationship to the dependent variable, we will improve our ability to predict the dependent variable accurately, and the log likelihood value will decrease. The initial –2LL value is 21332.017 at step 0, before any variables have been added to the model.

The statistical significance of individual regression coefficients is tested using the Wald and score chi-square statistic. In this section, we identify the statistically significant predictor variables and determine the direction of relationship with and contribution to the dependent variable.

Table 4.2: Test of Significance of Independent Variables Using Wald Test

Analysis of Maximum Likelihood Estimates								
Parameter	**DF**	**Estimate**	**S.E**	**Wald Chi-Square**	**Pr > Chisq**	**OR**	**95% Wald CL**	
Intercept	1	-0.3022	0.1261	5.7439	0.0165	-	-	-
REGION	**10**	**-**	**-**	**567.6441**	**<.0001**	**-**	**-**	**-**
Tigray	1	0.0246	0.0893	0.0760	0.7828	1.025	0.860	1.221
Affar	1	1.1303	0.1176	92.3640	<.0001	3.097	2.459	3.900
Amhara	1	-0.1657	0.0875	3.5856	0.0583	0.847	0.714	1.006
Oromiya	1	-0.5852	0.0850	47.3855	<.0001	0.557	0.472	0.658
Somali	1	0.7688	0.1168	43.3311	<.0001	2.157	1.716	2.712
Benishangul-Gumuz	1	-0.8814	0.0936	88.7197	<.0001	0.414	0.345	0.498
SNNP	1	-0.6048	0.0860	49.4342	<.0001	0.546	0.461	0.647
Gambela	1	-0.5337	0.0959	30.9731	<.0001	0.586	0.486	0.708
Harari	1	-0.0345	0.0934	0.1365	0.7118	0.966	0.804	1.160
Addis Ababa	1	-0.2515	0.0838	9.0046	0.0027	0.778	0.660	0.916
EDUCATION	**2**	**-**	**-**	**21.2887**	**<.0001**	**-**	**-**	**-**
No education	1	0.2857	0.0646	19.5562	<.0001	1.331	1.172	1.510
Primary	1	0.2263	0.0558	16.4798	<.0001	1.254	1.124	1.399
MASSMEDIA	**2**	**-**	**-**	**69.7083**	**<.0001**	**-**	**-**	**-**
Not at all	1	0.4075	0.0493	68.4462	<.0001	1.503	1.365	1.655

less a week	1	0.1261	0.0479	6.9145	0.0086	1.134	1.033	1.246
ECOSTATUS	**2**	**-**	**-**	**27.9906**	**<.0001**	**-**	**-**	**-**
Poor	1	0.2275	0.0521	19.0472	<.0001	1.255	1.134	1.390
Middle	1	-0.0224	0.0608	0.1364	0.7119	0.978	0.868	1.101
HOCCU	**2**	**-**	**-**	**67.8850**	**<.0001**	**-**	**-**	**-**
Not working	1	0.4736	0.0644	54.1522	<.0001	1.606	1.415	1.822
Agric employee	1	0.3334	0.0553	36.3784	<.0001	1.396	1.252	1.555
AGE	**2**	**-**	**-**	**259.5166**	**<.0001**	**-**	**-**	**-**
less than 25	1	0.7468	0.0547	186.4317	<.0001	2.110	1.896	2.349
between 25- 34	1	0.00141	0.0477	0.0009	0.9764	1.001	0.912	1.100
PLRESIDE	**1**	**-**	**-**	**37.9193**	**<.0001**	**-**	**-**	**-**
Rural	1	0.3723	0.0605	37.9193	<.0001	1.451	1.289	1.634
PRCHILD	**1**	**-**	**-**	**11.2237**	**0.0008**	**-**	**-**	**-**
N0	1	-0.1516	0.0453	11.2237	0.0008	0.859	0.786	0.939
SEXHHEAD	**1**	**-**	**-**	**52.0603**	**<.0001**	**-**	**-**	**-**
Female	1	-0.3049	0.0423	52.0603	<.0001	0.737	0.679	0.801
PREGNANCY	**1**	**-**	**-**	**11.3157**	**0.0008**	**-**	**-**	**-**
No or unsure	1	-0.2411	0.0717	11.3157	0.0008	0.786	0.683	0.904
MARITAL	**1**	**-**	**-**	**64.1294**	**<.0001**	**-**	**-**	**-**
Married	1	0.4180	0.0522	64.1294	<.0001	1.519	1.371	1.683

A negative sign in column labeled „Estimate" indicates an inverse relationship of explanatory variable with the log odds of the dependent variable. In contrast a positive coefficient column labeled „Estimate" indicates a positive relationship to the log odds of the dependent variable.

The statistical significance of individual regression coefficients is tested using the Wald and score chi-square statistic (see in the appendix for this result). According to Table 4.2

regions, place of residence, age, marital status, exposure to any mass media, husband/partners occupation, sex of household head, economic status of the household, educational level, presence of a child aged less or equal to 5 years in the household, and pregnancy were found to be significant predictors of unemployment at 5% level of significance. Thus, the estimated model is given by:

$$\log it(\hat{p}) = \beta_0 + \sum_{i=0}^{2} B_{1i} AW_i + \sum_{j=1}^{15} B_{2j} RW_j + \sum_{k=0}^{2} B_{3k} EW_k + \sum_{l=0}^{1} B_{4l} SHH_l + \sum_{m=0}^{1} B_{5m} PRW_m + \sum_{n=0}^{2} B_{6n} EMW_n$$
$$+ \sum_{0=0}^{2} B_{7o} ESH_o + \sum_{p=0}^{1} B_{8p} PW_p + \sum_{q=0}^{1} B_{9q} MW_q + \sum_{r=0}^{2} B_{10r} OH_r + \sum_{s=0}^{1} B_{11s} PCH_s$$

Where:

$\hat{p}$ = predicted probability of unemployment, B_0 = constant, AW_i = women age of level i, RW_j= women‟s region of level j, EW_k= educational background of women at level k, SHH_l = Sex Head of household at level l, $PRWd_m$=place of residence of women at level m, EMW_n = exposure of women to mass media at level n, ESH_o = economic status of household at level o, PW_p= nature of pregnancy of women at level p, MW_q = marital status of women at level q, OH_r = occupation of husband/partners at level r and PCH_s = Presence of child in the household at level s.

The value of explanatory variable for each category is taken as 1 if this variable falls in the corresponding category. For example,

AW_i = 1 for women age of level = i and AW_i = 0 for other levels.

RW_j = 1 women‟s region of level = j and RW_j =0 for other levels.

$EW_k = 1$ educational background of level = k and $EW_k = 0$ for other levels.

Similarly, each of the other variables takes value 1 if it falls within the corresponding level of category. Based on the above result, the regression equation consisting of the significant variables is given by:

$$\log it(\hat{p}) = -0.3022 + 0.7468\mathbf{AW_1} + 0.00141\mathbf{AW_2} + 0.0246\,\mathbf{RW_1} + 1.1303\mathbf{RW_2} + \ldots$$

$$- \ 0.2515\mathbf{RW_{14}} + 0.2857\mathbf{EW_1} + 0.2263\mathbf{EW2} + 0.3723\mathbf{PRW_1} + 0.4075\mathbf{EMW1}$$

$$+ \ \mathbf{0}.1261\mathbf{EMW_2} + 0.2275\mathbf{ESH_1} - 0.0224\mathbf{ESH_2} - 0.2411\mathbf{PW1}$$

$$+ \ 0.4180\mathbf{MW_1} + 0.3049\mathbf{SHH_1} + 0.4736\mathbf{OH_1} + 0.3334\mathbf{OH_2}$$

$$-0.1516\mathbf{PCH_1} \ldots\ldots\ldots\ldots\ldots\ldots \text{(EQ1)}$$

As can be seen in Table 4.2 at the final step, all independent variables are added to the logistic regression equation in a stepwise manner. The addition of these variables reduced the initial log likelihood value (-2 Log Likelihood) of 21,332.017 to 19,190.665.

Table 4.3: Result of Model Fit Statistics for Intercept only and Full Model

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	21334.017	19242.665
SC	21341.722	19442.974
-2 Log L	21332.017	19190.665

The difference between these two measures is the model chi-square value or likelihood ratio (2141.3524 = 21,332.017-19,190.665) that is tested for statistical significance. This test is analogous to the F-test for R^2 or change in R^2 value in multiple regressions which tests whether or not the improvement in the model associated with the additional variables is statistically significant.

Table 4.4: Test of Significance of the Relationship between the Dependent and Independent Variables.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	2141.3524	25	<.0001
Score	1967.4172	25	<.0001
Wald	1708.8907	25	<.0001

In Table 4.4 the model Chi-Square value of 2141.3524 has a P-Value of less than 0.01. Similarly, score and Wald tests also have P-values less than 0.05 and are significant. These indicate that all three tests yield similar conclusions, that is, the final model with explanatory variables was more effective than the null model. So, we conclude that there is a significant relationship between the dependent variable and the set of independent variables.

There are several numerical problems that can occur in logistic regression that are not detected by statistical packages: multicollinearity among the independent variables, zero cells for a dummy-coded independent variable because all of the subjects have the same value for the variable, and "complete separation" whereby the two groups in the dependent event variable can be perfectly separated by scores on one or a combination of the independent variables. All of these problems produce large standard errors (over two) for the variables included in the analysis and very often produce very large β coefficients as well.

As can be seen in Table 4.2 the standard errors of the variables included in the model are substantially lower than two. In addition, the β coefficients have low value. Therefore, we conclude that the fit of the derived model is good.

The Hosmer-Lemeshow goodness of fit test was employed and the result is presented in Table 4.5. A good model fit is indicated by a none-significant chi-square value.

Table 4.5: Test of Significance of Hosmer-Lemeshow Goodness of Fit Statistics

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
11.4670	8	0.1766

The better model fit is indicated by a smaller difference in the observed and predicted classification. The Hosmer-Lemeshow Goodness-of-fit test tests the hypotheses:

H_o: the model is a good fit, vs.

H_a: the model is not a good fit

Since the P-value in Table 4 P-value = 0.1766 is larger than 0.05, we do not reject the null hypothesis, and we conclude that the model is a good fit.

Finally, the four measures of association for assessing the predictive ability of the model are presented in Table 4.6.

Table 4.6: Association of Predicted Probabilities and Observed Responses

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	89.8	Somer's D	0.800
Percent Discordant	9.8	Gamma	0.803
Percent Tied	0.4	Tau-a	0.367
Pairs	61536753	c	0.900

The value of the Gamma statistic is 0.803(see Table 4.6) implying that 80.30% fewer prediction errors are by using the estimated probabilities than by chance alone. The value of C statistic is 0.900. This means that for 90% of all possible pairs of women the model correctly assigned a higher probability to those who were unemployed. Here, the values displayed in Table 4.6 indicate that the fitted model has reasonable predictive power, although the tau-a value is very small.

A more appealing way to interpret the regression coefficient in logistic model is using odds ratio. The odds ratio indicate the effect of each explanatory variable directly on the odds of being unemployed rather than on log (odds). Estimates of odds greater than 1.0 indicate that the risk of unemployment is greater than that for the reference category. Estimates less than 1.0 indicate that the risk of unemployment is less than that for the reference category of each variable. So, the final model presented in Table 4.2 is interpreted in terms of odds ratio as follows.

The odd of unemployment of women compared to employment in Tigray is not significantly different from that in Dire Dawa. The odds of unemployment of women in Affar are 3.097 times higher than the odds of unemployment of women in Dire Dawa controlling for other variables in the model. The odds of unemployment of women compared to employment of women are lower by a factor of 0.557 when the respondent (women) is in Oromiya compared to women in dire Dawa controlling for other variables in the model.

The odds of unemployment of women who have age less than 25 years were 2.110 (OR=2.110) times higher than the odds of unemployment of women who have age above 34 years old controlling other variables in the model.

Women who had no education were 33.1% more likely to be unemployed (OR=1.331) compared to women with secondary or higher education controlling for other variables in the model, while women with primary education were 25.4% (OR=1.254) more likely to be unemployed compared to women with secondary or higher education controlling for other variables in the model.

Women who live in rural households were 45.1% more likely to be unemployed (OR=1.451) than women who reside in urban households controlling for other variables in the model. The odds of unemployment of women decreased by a factor of 0.737 when the women is head of household compared to women who are not the head of the household controlling for other variables in the model.

Women who had no any exposure of mass media were about 50.3% more likely to be unemployed than that of women who have exposure at least once a week controlling for other variables in the model.

Women who live in poor households were about 25.5% more likely to be unemployed than that of women who live in rich households controlling for other variables in the model.

The odds of unemployment status for women who are not pregnant is lower by a factor of 0.786 compared to pregnant women controlling for other variables in the model. Conversely, married women were about 51.9% more likely to be unemployed than other women controlling for other variables in the model.

Women who live in a household without a child aged less or equal to five years old were about 14.1% (OR = 0.859) less likely to be unemployed than women who live in households with child aged five years old and less with controlling for other variables in the model.

Using the derived logistic regression model,

$$\begin{aligned} \log it(\hat{p}) = {} & -0.3022 + 0.7468\mathbf{AW}_1 + 0.00141\mathbf{AW}_2 + 0.0246\,\mathbf{RW}_1 + 1.1303\mathbf{RW}_2 + \ldots \\ & - 0.2515\mathbf{RW}_{14} + 0.2857\mathbf{EW}_1 + 0.2263\mathbf{EW2} + 0.3723\mathbf{PRW}_1 + 0.4075\mathbf{EMW1} \\ & + \mathbf{0}.1261\mathbf{EMW}_2 + 0.2275\mathbf{ESH}_1 - 0.0224\mathbf{ESH}_2 - 0.2411\mathbf{PW1} \\ & + 0.4180\mathbf{MW}_1 + 0.3049\mathbf{SHH}_1 + 0.4736\mathbf{OH}_1 + 0.3334\mathbf{OH}_2 \\ & - 0.1516\mathbf{PCH}_1. \end{aligned}$$

It is possible to calculate the predicted value of the probability of unemployment for any combination of values of the explanatory variables. For example, for non pregnant married women who are heads of household from region 7, aged 28, with primary education reside in rural areas, and living in poor households which is headed with agric employed husband without child in the household, whose exposure of mass media is less than a week, the predicted probability of being unemployed is estimated (using the above equation) to be about 14.9%.

4.2.1. **MODEL DIAGNOSTICS**

Regression model building is often an iterative and interactive process. The first fitted model for the data may be inadequate. Regression diagnostics are used to detect problems with the model and suggest improvements. A failure to detect outliers and influential cases can have severe distortion on the validity of the inferences. It would be reasonable to use diagnostics to check if the model is adequate or not. The adequacy of the fitted model was checked for possible presence and treatment of outliers and influential values. The diagnostic test results for detection of outliers and influential values are presented in the Appendix A. The residuals like Studentized, deviance, Pearson and standardized residuals are all less than 3 in absolute value. The residuals less than 3 in absolute value show the absence of an outlier observation. The DFBETAs for model parameters (including the constant term), Cook"s influence statistic and Leverage Values (Hat Diag) were less than 0.0156, 1 and 1 respectively. DFBETAs less than 0.0156 imply an observation has no effect on the estimate of a regression coefficient while a Cook"s distance less than unity showed each observation had no impact on the group of regression coefficients. A value of the leverage statistic less than one shows that no observation is far apart from the others in terms of the levels of the independent variables (not the dependent variable). Thus, from the above goodness of fit tests and diagnostic checking, we can say that our model is adequate.

4.3. DETERMINANTS OF WOMEN UNEMPLOYMENT IN ETHIOPIA: MULTILEVEL LOGISTIC ANALYSIS

The data used in this study have a hierarchical structure. Units at one level are nested within units at the next higher level. Here, the lower level (level-1) units are the individual women, and the higher level (level-2) units are the regions that constitute the groups into which the women are clustered or nested. The nesting structure is women within regions that resulted in a set of 11 regions with a total of 16,386 women. The data used in this study consist of variables describing individuals as well as variables describing regions. Therefore, the statistical model used has to describe the data at both levels, in order to find the effect on unemployment of both the individual woman and the regions.

As mentioned above, multilevel models were developed to analyze hierarchically structured data. These models contain variables measured at different levels of the hierarchy. Unobserved heterogeneity is modeled by introducing random effects. Random intercepts are used to model are used to model unobserved heterogeneity in the overall response; random coefficients model unobserved heterogeneity in the effects of explanatory variables on the response variable. As one of the aims of this study was to model the heterogeneity between regions, a random intercept model was used. This allows the overall probability of unemployment to vary across regions. The chi-square test was applied to assess heterogeneity between regions mean. The test yields $X^2 = 854.127$ with d.f 10 (P<0.0001). Thus, there is evidence of heterogeneity among the regions.

Women in this study were selected from different regions of Ethiopia. Thus there are two kinds of random variability in the data – that between different women in a single region, and that between different regions. The advantages of using a multilevel model include the ability to fully explore the variability at all levels of the data hierarchy, and estimation of correct standard errors in the presence of clustered data. Women from the same region would tend to be more similar compared to women chosen at random from different regions. The model takes into account the correlation structure of the data, enabling correct inferences to be made.

4.3.1. Determinants of Women Unemployment in Ethiopia: Random Intercept-Only Model

We first fit a simple model with no predictors i.e. an intercept-only model that predicts the probability of being unemployed. The functional form of the model is:

$$\log it(P_j) = \beta_0 + U_{0j}$$

The estimates of parameters and standard errors are presented in Table 4.8. The maximum likelihood estimate of the empty model of standard logit model is 0.5939 with standard error 0.0163. The ML estimate from the standard logit model of the ratio of unemployed to employed is exp (0.5939) = 1.811, which is the same as the sample ratio of 10,557 unemployed to 5829 employed. It is in fact the estimated odds-ratio when no predictors have been considered in the model. A crude comparison has been made to understand the multilevel effects. Compared to the odds-ratios obtained by multilevel methods the standard logistic model odds-ratio has been underestimated. It is observed that there is a significant difference between the standard logistic estimate and the multilevel logistic estimate. Therefore, by failing to take into account the women within regions, the standard logistic model has underestimated the odds-ratio by about 15.87% [(0.5939-0.7060)*100/0.7060] compared to multilevel model.

Table 4.8: Result of Parameter Estimate of intercept-only model with random effect using GLIMMIX

Fixed part	Estimates	S.E	d.f	t-value	P-value
Intercept	0.7060	0.1842	10	3.83	0.0033
Random Part	**Estimate**	**S.E**	**Z-value**	**P-value**	
Var(U_{oj}) = σ^2_0	0.3692	0.1602	2.30	0.0106	
Deviance= 20465.93, **AIC =** 20469.93 , **BIC** = 20470.72 **D**eviance-based Chi-square = 866.09					

The variance of the regional level residuals errors, symbolized by σ^2_0 is estimated to be 0.3692. This parameter estimate is larger than the corresponding standard errors and calculation of the Z-test shows that it is significant at $p<0.0106$. The significance of this residual term indicates that there are regional differences in the women unemployment status in Ethiopia.

The deviance-based Chi-square (deviance = 866.09) indicated in Table 4.8 is the difference in deviance between an empty model without random effect (deviance = 21332.02) and an empty model with random effect (deviance =20465.93). This value is compared to chi-square distribution with 1 degree of freedom. The significant of it ($X^2 =$ 866.09, P-value < 0.0001) implies that an empty model with random intercept is better than an empty model without random intercept. The deviance reported in the above Table is a measure of model misfit; when we add explanatory variables to the model, the deviance is expected to go down.

4.3.2. Determinants of Women Unemployment in Ethiopia: Random Intercept and Fixed Effect Model

We will now extend our model to allow for regional effects on the probability of being unemployed. We begin with a random intercept or variance components model that allows the overall probability of unemployment to vary across regions. The results of two-level random intercept and fixed slope model are presented in Table 4.9.

Table 4.9: Result of Parameter Estimate of random intercept and fixed slope multilevel logistic model using Proc GLIMMIX

Effect	Level	Estimate	S.E	DF	t-value	Pr > |t|
Intercept	Intercept	-0.7117	0.2447	10	-2.91	0.0156*
EDUCATION	No education	0.2781	0.06815	20	4.08	0.0006*
EDUCATION	Primary	0.2259	0.05708	20	3.96	0.0008*
EDUCATION	Secondary+	.	.	.	.	.
MASSMEDIA	Not at all	0.4032	0.04942	20	8.16	<.0001*
MASSMEDIA	less a week	0.1243	0.04797	20	2.59	0.0174*
MASSMEDIA	at least once a week	.	.	.	.	.
ECOSTATUS	Poor	0.2293	0.05246	20	4.37	0.0003*
ECOSTATUS	Middle	-0.02567	0.06086	20	-0.42	0.6776
ECOSTATUS	Rich	.	.	.	.	.
HOCCU	Not working	0.4536	0.07096	20	6.39	<.0001*

HOCCU	Agric employee	0.3103	0.05817	20	5.34	<.0001*
HOCCU	Non-Agric employee(ref.)	.	.	.	.	.
AGE	Less than 25	0.7689	0.05609	20	13.71	<.0001*
AGE	Between 25-34	0.01835	0.04865	20	0.38	0.7100
AGE	Above 34(ref.)	.	.	.	.	.
PLRESIDE	Rural	0.3584	0.06060	9	5.91	0.0002*
PLRESIDE	Urban(ref.)	.	.	.	.	.
ADDICTION	No	0.2055	0.1040	10	1.98	0.0763
ADDICTION	Yes (ref.)	.	.	.	.	.
HHSIZE	Less or equal to 5	0.04069	0.09860	20	0.41	0.6842
HHSIZE	Between 6-10	0.1195	0.09769	20	1.22	0.2356
HHSIZE	Above 10(ref.)	.	.	.	.	.
PRCHILD	No	-0.1408	0.04578	10	-3.08	0.0117*
PRCHILD	Yes (ref.)	.	.	.	.	.
MIGRATION	Visitor	-0.00742	0.09727	10	-0.08	0.9407
MIGRATION	Usual(ref.)	.	.	.	.	.
SEXHHEAD	Female	-0.2849	0.04356	10	-6.54	<.0001*
SEXHHEAD	Male(ref.)	.	.	.	.	.
PREGNANCY	No	-0.2440	0.07169	10	-3.40	0.0067*
PREGNANCY	Yes (ref.)	.	.	.	.	.
MARITAL	Married	0.4186	0.05229	10	8.01	<.0001*
MARITAL	Others(ref)	.	.	.	.	.
HEDU	No education	0.06306	0.07366	20	0.86	0.4021
HEDU	Primary	-0.00169	0.06536	20	-0.03	0.9797
HEDU	Secondary+ (ref.)	.	.	.	.	.
DECISSION	No	0.03396	0.03491	10	0.97	0.3536
DECISSION	Yes (ref.)	.	.	.	.	.
Random Effect	**-**	**Estimate**	**S.E**	**Z-value**		**Pr > |Z|**
Var(U_{oj})= σ^2_0	-	0.3103	0.1355	2.29		0.0110*
Deviance =19239.64, AIC = 19,287.64, BIC = 19,297.19 Deviance-based Chi-square = 1,226.29						

*=significant at 5% level of significance

In Table 4.9, the variance component representing variation between regions has decreased from 0.3692 in the empty model with random intercept to 0.3103 in the random intercept and fixed slopes multilevel logistic regression model and the significance of it indicates that there is a significant variation between regions in the unemployment status of women.

The deviance-based Chi-square (deviance = 1226.29) indicated in Table 4.9 is the difference in deviance between the empty model with random intercept (deviance = 20,465.93) and fixed slope model with random intercept (deviance = 19,239.64). This value is compared to chi-square distribution with 23 degree of freedom. The significant of it (X^2 =1,226.29, df = 23, P-value < 0.0001) implies that fixed slope model with random intercept model is better than empty model with random intercept. Therefore, this model is a better fit as compared to the empty model with random intercept.

Moreover, the AIC and BIC value for fixed slope model with random intercept (AIC=19,287.64, and BIC=19,297.19) are less than those for the empty model with random intercept (AIC = 20,469.93 and BIC = 20,470.72). This indicates that fixed slope model with random intercept is a better fit as compared to the empty model with random intercept model.

4.3.3. Determinants of women unemployment in Ethiopia: The Random Coefficient Model

The variance components model which we have just specified and estimated above assumes that the only variation between regions is in their intercepts. We should allow for the possibility that the regions have different slopes. This implies that the coefficients of explanatory variables are random at level two. All variables included in the random intercept model are included in random coefficient model. Estimates of this model show that the random slope variances of all included variables except that of Mass Media is zero. This indicates that the effect of these variables is the same for each region. Therefore, those random slopes are excluded from the model. Results of random coefficient model are presented in Table 4.10.

TABLE 4.10: Result of Parameter Estimate of Random Coefficient Multilevel Model using Proc GLIMMIX.

Solutions for Fixed Effects							Odds Ratio Estimates		
Effect	**level**	**Estimate**	**S.E**	**DF**	**t - Value**	**Pr > |t|**	Estimate	95% Confidence Limits	
Intercept	-	-0.7251	0.2555	10	-2.84	0.0176*	-	-	-
EDUCATION	No education	0.2848	0.06833	20	4.17	0.0005*	1.330	1.153	1.533
EDUCATION	Primary	0.2369	0.05715	20	4.14	0.0005*	1.267	1.125	1.428
EDUCATION	Secondary+	.	.	.	.	.	-	-	-
MASSMEDIA	Not at all	0.4804	0.1158	20	4.15	0.0005*	1.617	1.270	2.058
MASSMEDIA	less a week	0.1518	0.1140	20	1.33	0.1979	1.164	0.918	1.477
MASSMEDIA	At least once a week	.	.	.	.	.	-	-	-
ECOSTATUS	Poor	0.2144	0.05278	20	4.06	0.0006*	1.239	1.110	1.383
ECOSTATUS	Middle	-0.0307	0.06101	20	-0.50	0.6200	0.970	0.854	1.101
ECOSTATUS	Rich	.	.	.	.	.	-	-	-
HOCCU	Not working	0.4551	0.07107	20	6.40	<.0001*	1.577	1.360	1.829
HOCCU	Agric employee	0.3157	0.05843	20	5.40	<.0001*	1.372	1.214	1.549
HOCCU	Non-Agric employee(ref.)	.	.	.	.	.	-	-	-
AGE	< 25	0.7644	0.05621	20	13.60	<.0001*	2.148	1.910	2.415
AGE	Between 25-34	0.01846	0.04875	20	0.38	0.7090	1.019	0.920	1.128
AGE	34^{+} (ref.)	.	.	.	.	.	-	-	-
PLRESIDE	Rural	0.3504	0.06101	9	5.74	0.0003*	1.420	1.237	1.630

PLRESIDE	Urban(ref.)	.	.	.	.	.	-	-	-
ADDICTION	No	0.2015	0.1059	10	1.90	0.0862	1.223	0.966	1.549
ADDICTION	Yes (ref.)	.	.	.	.	.	-	-	-
HHSIZE	≤ 5	0.03263	0.09899	20	0.33	0.7451	1.033	0.841	1.270
HHSIZE	Between 6-10	0.1108	0.09810	20	1.13	0.2721	1.117	0.911	1.371
HHSIZE	10^{+} (ref.)	.	.	.	.	.			
PRCHILD	No	-0.1432	0.04589	10	-3.12	0.0109*	0.867	0.782	0.960
PRCHILD	Yes (ref.)	.	.	.	.	.	-	-	-
MIGRATION	Visitor	0.00195	0.09735	10	0.02	0.9844	1.002	0.807	1.245
MIGRATION	Usual(ref.)	.	.	.	.	.	-	-	-
SEXHHEAD	Female	-0.2965	0.04374	10	-6.78	<.0001*	0.743	0.674	0.820
SEXHHEAD	Male(ref.)	.	.	.	.	.	-	-	-
PREGNANY	No	-0.2445	0.07181	10	-3.41	0.0067*	0.783	0.667	0.919
PREGNANY	Yes (ref.)	.	.	.	.	.	-	-	-
MARITAL	Married	0.4192	0.05242	10	8.00	<.0001*	1.521	1.353	1.709
MARITAL	Others(ref)	.	.	.	.	.	-	-	-
HEDU	No education	0.05631	0.07377	20	0.76	0.4541	1.058	0.907	1.233
HEDU	Primary	-0.0008	0.06534	20	-0.01	0.9899	0.999	0.872	1.145
HEDU	$Secondary^{+}$ (ref.)	.	.	.	.	.	-	-	-
DECISSION	No	0.03477	0.03498	10	0.99	0.3436	1.035	0.958	1.119
DECISSION	Yes (ref.)	.	.	.	.	.	-	-	-

Random effect	**B**	**S.E**	**Z-value**	**P- Value**
Var(U_{oj})= σ_0^2	0.3053	0.1416	2.16	0.0155*
Var(U_{2j})= σ_{2j}^2	0.0555	0.0236	2.35	0.0094*
Cov (U_{oj}, U_{2j})	-0.0749	0.0259	-2.89	0.0039*
Deviance = **19206.58**, AIC= **19256.58**, BIC = **19266.53**, Deviance based chi-square=**33.06**				

Ref=Reference Category, *=significant at 5% level of significance

In Table 4.10, the value of Var (U_{0j}) and Var (U_{2j}) are the estimated variance of intercept and slope of Mass Media respectively. These estimated variances are significant suggesting that intercept and slope of mass media vary significantly. So, there is a significant variation in the effect of mass media across regions in Ethiopia.

The effect of the intercept on region j is estimated to be -0.7251(0.2555) +U_{0j} and their variance 0.3056 (Standard error 0.1419). The intercept variance of 0.3056 (Standard error 0.1419) is interpreted as the between-region variance when all other variables are held constant (i.e. equal to zero). Their mean is -0.7251 (standard error 0.2555) and their variance is 0.3053 (standard error 0.1416). The between-region variance of slope of mass media is estimated to be 0.0555(standard error 0.0236). The individual region slopes of mass media vary about with a variance 0.0555(standard error 0.02362). The negative covariance estimate of -0.0749 (standard error 0.0259) between intercept and slopes of mass media, suggest that regions with a high intercept (above-average) tends to have a flatter-than-average slope.

Generally, interpretation of significant covariance terms can be easily made in terms of the correlation coefficients between random intercept and random slopes. Positive covariance/correlation between intercept and slopes implies that regions with higher intercepts tend to have on average higher slopes on the corresponding predictors.

The intercept-slope correlation, for example intercept and slope of Mass Media, is estimated as:

$$\rho_{02} = \frac{-0.0749}{\sqrt{0.3053 * 0.0555}} = -0.575$$

The negative sign for the correlation between intercepts and slopes implies that regions with higher intercepts tend to have on average lower slopes on the corresponding predictors. This value indicates that women who live in those regions with high exposure of mass media are less likely to be unemployed than women who live in regions without any exposure of mass media.

The quantities AIC and BIC can be used to make an overall comparison of this more complicated model with the random intercept model with fixed slope model. We see that the value of fit statistics for random coefficient model (AIC = 19,256.58 and BIC = 19,266.53) is less than random intercept model **(**AIC=19,287.64, and BIC=19,297.19). This indicates that the random coefficient model is a better fit as compared to the random intercept and fixed effect model. The random coefficient model involves two extra parameters, the variance of the slope residuals (i.e. mass media), U_{2j} and their covariance with the intercept residuals U_{0j} and the change (which is also the change in deviance) can be regarded as a χ^2 value with 2 degrees of freedom under the null hypothesis that the extra parameters have population values of zero. The value of deviance based chi-square is given by (19,239.64 – 19,206.58 = 33.06, P-Value = 0.0001) which shows that the addition of this fixed effects and one random coefficient has significantly improved the fit of the more elaborate model to the data.

The parameters of observed variables can be interpreted much the same way as those from the standard logit model. Thus, everything else being equal except slight difference on random effect in the model, women who have no education were 33% more likely to be unemployed (OR=1.33) compared to women with secondary or higher education controlling for other variables in the model and random effect at level two while women with primary education were 26.7% (OR=1.254) more likely to be unemployed compared to women with secondary or higher education controlling for other variables in the model and random effect at level two.

Women who live in poor households were about 23.9% more likely to be unemployed than that of women who live in rich households controlling for other variables in the model and random effect at level two.

The odds of unemployment of women who have unemployed husband/partners were 57.7% more likely to be unemployed than women who have non-agric employee husband/partners controlling for other variables in the model and random effect at level two.

The odds of unemployment of women who have age less than 25 years were 2.148 (OR=2.148) times higher than the odds of unemployment of women who older than 34 years controlling other variables in the model and random effects at level two.

Women who live in rural households are 42% more likely to be unemployed (OR=1.42) than women who reside in urban households controlling for other variables in the model and random effects at level two. Conversely, the odds of unemployment of women is lower by a factor of 0.743 when the women is head of household compared to women who is not the head of the household controlling for other variables in the model and random effects at level two.

The odds of unemployment status for women who are not pregnant is lower by a factor of 0.783 compared to pregnant women controlling for other variables in the model and random effects at level two. Conversely, married women were about 35.3% more likely to be unemployed than other women (divorced, widowed) controlling for other variables in the model and random effects at level two.

Women who live in a household without a child aged less or equal to five years old were about 14.1% (OR = 0.859) less likely to be unemployed than women who live in households with child aged five years old and less controlling for other variables in the model.

4.3.4. Goodness of Fit Test

An overall evaluation of the multilevel logistic model was assessed using the deviance. The test is done by comparing the deviance (-2 log likelihood) of two models by subtracting the smaller deviance (model with more parameters) from the larger deviance (model with larger deviance). The difference is a chi-square test with the number of degrees of freedom equal to the number of different parameters in the two models. The significance of this chi square test indicates that the model is a good fit or is a better model than the smaller model. Similarly, the overall model evaluation was assessed using AIC and BIC. Based on the result we obtained in Table 4.10 (random coefficient model), the deviance chi-square is significant and the value of AIC and BIC is less than the model we obtained in the random intercept with fixed slope model. So, we conclude that the model is a good fit.

4.3.5. Diagnostic Checking

The diagnostic test results for detection of outliers and influential values are presented in Appendix C. As can be seen in Appendix C all types of residuals are less than 3 in absolute value for both level one and level two. Similarly, the value of leverage and influential observation is less than one. Thus, from the goodness of fit test and diagnostic test results presented in the Appendix C, we can say that the fitted model is adequate.

Chapter 5

5. Discussion, Conclusions and Recommendations

5.1. Discussion and Conclusions

This study found evidence that some of the demographic and socioeconomic variables considered have significant influence on women unemployment. Place of residence, age, marital status, exposure to any mass media, husband/partners occupation, sex of household head, economic status of the household, educational level, and presence of a child aged less or equal to 5 years in the household were found to be important determinants of unemployment among reproductive age women (15-49 years). Based on the findings in the preceding chapter, this study arrives at the following conclusions.

Women who had no education were more likely to be unemployed (OR=1.33) compared to women with secondary or higher education controlling for other variables in the model. Women with primary education were more likely (OR=1.254) to be unemployed compared to women with secondary or higher education. The reason might be that education may enable women to make independent decisions, to be accepted by other household members and community, and paves the way to have greater access to job opportunity. This is in agreement with the findings in other studies. Foley (1997); Yang, (1992), and Bhorat, (2007) showed that the higher the level of education, the more likely to participate in the market and the more likely to be employed

Women younger than 25 years of age are the most likely to be unemployed than older women. This is also consistent with the findings in other studies. Michael (1985) and

Foley (1997) showed that that women in the youngest age group surveyed are most affected by unemployment.

Similarly, place of residence is a significant factor contributing to being unemployed. Rural women are less likely to participate in market activities (more unemployment rate) than women in urban areas (Yang, 1992). The results of this study also show that there is a difference in the unemployment status of women residing in rural and urban areas. Women who reside in rural areas are more likely to be unemployed than women who reside in urban areas. This may be due to fact that women residing in rural areas have a high shadow value of home production activities.

Household economic status is one of the most important determinants of unemployment status among women in Ethiopia. According to our findings, as compared with women residing in higher economic status households, the risk of being unemployed for women in poor households was highly significant. This finding is consistent with other studies (Bhorat, 2007) showing that women of very poor or poor (low economic status) households have the highest rates of unemployment.

Unmarried women and heads of household are more likely to work than are married women (Scott, 1992; Yang, 1992). The present study also showed that women who live in female headed households are more likely to be employed than women who live in male headed households. Married women are more likely to be unemployed than unmarried women. This may be due to high involvement of women in home production activities.

Some studies have revealed that labor supply of women increases if the husband/partners are self-employed (Franz, 1985). This is in agreement with our finding that women who have employed husband/partners are more likely to be employed than women who have unemployed husband/partners. Presence of young children has also been found to be a major determinant of unemployment in previous studies. If there are many teenagers and babies in the household, the women labor supply decreases (Franz, 1985; Bhorat, 2007) consistent with our finding that women who live in households without children aged 5 years or less in the household are less likely to be unemployed than women who live in households with children aged 5 years or less in the household. This may be due to the presence of helpers/servants in the household to care for those children and help home production activities.

Women who were pregnant in a given year had a lower probability of participating in the labor market than women who had not been pregnant (Scott, 1992). This study also showed that non pregnant women were less likely to be unemployed than pregnant women.

From the above discussion, it can be seen that the individual-level predictors of unemployment identified in the present study have all been well-established in the literature. The other finding in this study is the identification of variable at the regional level that explains the variation in unemployment between the regions of Ethiopia. There are no studies involving multilevel modeling of unemployment in Ethiopia that included variables at higher levels. The present study also identified socio-economic indicators of the region as predictors of unemployment. This is the exposure of mass media in different regions of Ethiopia. According to the final model, this level-two variable explains all of the regional-level variation in unemployment found in the data.

5.2. Recommendations

The results obtained from this study are of great concern to policy makers because of the negative effects of unemployment on the loss of output, on the society and on the psychological well being of the unemployed and immediate family members. In order to formulate policies to control the rising problem of unemployment in Ethiopia, it is important not only to understand the effect of reforms on the incidence of unemployment among the women, but also on the duration of unemployment and, on the probability of exiting unemployment and how it differs with demographic and economic characteristics. In response to this challenge, this paper suggests the following possible solution to tackle women unemployment problem in Ethiopia.

- Policies that promote education and create more job opportunities should be implemented. For example, re-schooling or training of the less educated women, increasing vocational training and labor market information. This would encourage more women to go to work, and thus generate the income required that would enable more families in the regions to be able to increase their living standards.

- The government should give more support and emphasis on those regions with high rates of unemployment. Additionally, further research on socio-cultural practices, distribution of education, women‟s workload, and other related factors should be emphasized. In order to decreases unemployment levels in regions with lower levels, the socio-economic status of the regions has to be raised. As a consequence, differences in the level of unemployment between regions would be reduced, and job opportunity would be more uniform across all regions.

- The government should take a measure of action to support the very poor, and to bring about rapid economic growth at the national level. To this effect, it is important to develop community-based interventions giving priority to very poor households to participate in the labor market, education, health facility and areas of job access.
- The government or concerned bodies should make the coverage of mass media uniform across all regions of Ethiopia.
- Efforts should be made to improve place of residence of women who live in rural areas of Ethiopia.

References

Adebayo, A. (1999): "Youth unemployment and National Directorate of Employment and self-employment programmes". Nigerian Journal of Economics and Social Studies 41 (1), 81-102.

Agresti, A. (2002), Categorical Data Analysis. Wiley Interscience, New York.

Agresti, A. (2007), an Introduction to Categorical Data Analysis. John Wiley And Sons, Inc, New York

Bakare A.S (2011). The Determinants of Urban Unemployment Crisis in Nigeria: An Econometric Analysis Department of Economics, Adekunle Ajasin University, Nigeria.

Ledezma, L. R., Orlando, M. B., and Zuniga, G. (2003). "Determinants of the Labor Force Participation of Women in Venezuela, 1980-2000." Report prepared for the XXIV Latin American Studies Association. International Congress, Dallas, Texas, USA.

Belsley D. A, E. Kuh, and R. E. Welsch. (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

Benefo K.D and Pillai V.K (2003). The determinants of female non-family work in Africa

Berhanu Denu, Abraham Tekeste and Hannah Van Der Deij (2005), Characteristics and Determinants of Youth Unemployment, Underemployment, Underemployment and Inadequate Employment in Ethiopia, Employment Policies Unit Employment Strategy Department.

Bhorat, H. (2007). Unemployment in South Africa: Descriptors and Determinants. Paper Presented To The Commission On Growth And Development, World Bank, Washington Dc.

Breslow, N. E. And D. G. Clayton (1993) Approximate Inference in Generalized Linear Mixed Models. J. Am. Statist. Assoc., 88:9–25.

Bryk, A.S., and Raudenbush, S.W (1992).Hierarchal Linear Models, Newbury Park, Sage

C.S.A. (1999). Report on Labor Force Survey 1999 (Statistical Bulletin), Addis Ababa, Ethiopia.

C.S.A. (2005). Report on Labor Force Survey 1999 (Statistical Bulletin), Addis Ababa, Ethiopia.

Cook, R. D. (1998). Detection of influential observations in linear regression. Technimetrics, 19, 15-18.

Copas, J.B. (1988). Binary Regression Models for Contaminated Data. Journal of Royal Statistical Association, B 50 (2): 225-265.

Cox, D., and Psacharopoulos, G. (1992) "Female Participation and Earnings in Venezuela, 1987." In Case Studies on Women"s Employment and Pay in Latin America. World Bank Report, No. 12175, Washington, D.C., Pp. 451-461.

Dantwala, M. (1971): The Definition and Measurement of Unemployment in Developing Countries. pp 29-39

Englama, A. (2001) "Unemployment: Concepts and Issues." Bullion Publication of the Central Bank of Nigeria 25 (4), 1-6.

Eita, J.H. (2010). Determinants of Unemployment in Namibia. South African Journal of Economic and Management Sciences.

Falae, S.O. (1971) "Unemployment in Nigeria". Nigerian Journal of Economics and Social Studies 13 (1) 65

Foley, M.C. (1997). Determinants of Unemployment in Russia, Center Discussion Paper No. 779, Yale University.

Franz, Wolfgang, 1985. "An Economic Analysis of Female Work Participation, Education, and Fertility: Theory and Empirical Evidence for the Federal Republic of Germany," Journal of Labor Economics, University of Chicago Press, vol. 3(1), pages S218-34, January.

Goldstein, H. (2003) Multilevel Statistical Models. 3rd Edition London: Arnold; New York: Oxford University Press Inc.

Hilbe, J. M.(2009), Logistic Regression Models, Chapman & Hall, London

Hongyu Yang, 1992. "Female Labor Force Participation and Wages: A Case Study of Panama." In Case Studies on Women"s Employment and Pay in Latin America.

Hosmer, W.D. and S. Lemeshow (2000), Applied Logistic Regression. 2nd Ed., John Wiley and Sons, New York.

Hox J.J.,(2002), Multilevel Analysis: Techniques And Application, Lawrence Erlbaum Associates, Inc., London

ILO (2009a), Global Employment Trends. ILO. Geneva

ILO (2009b). Global Employment Trends updates. ILO. Geneva.

ILO (1983). International Conference of Labour Statistical report of the Conference, Geneva.

ILO (2010). Growth-employment-poverty reduction linkages: a framework for recovery and accelerated progress towards the Millennium Development Goals, Economic Report on Africa 2010.

Katherine Scott (1992). The female labor force participation in Bolivia

Kingdon, G., and Knight, J. (2001). Race and the Incidence of Unemployment in South Africa., CSAE working papers/2001-18.

Lippman, S., and McCall, J. (1976). The Economics of Job Search: A Survey. Economic Inquiry, 14, 155-189.

Longford N.T. (1993).Random Coefficient Models. Oxford, Clarendom Press.

Mahmood, Z., N. Akhtar, M. Amin and M. Idrees. (2011). Causes of unemployment among the education segments in Peshawar division, Pakistan: a statistical study. Sarhad J. Agric. 27(1): 139-142 Models Project, London.

Mortensen, D. (1970) Job Search, the Duration of Unemployment and the Phillips Curve. American Economic Review, 30, 847-862.

Newsom (2011). Significance Testing in Multilevel Regression

Onah, F.O. (2010) "Urban Unemployment Situation in Nigeria". Some Implication for the Third Millennium Global Journal of Social Sciences 2 (1) 21-26.

Psacharopoulos, G., and Tzannatos, Z. (1992). "Female Labor Force Participation: An International Perspective." The World Bank Research Observer, 4.

Robert T. Michael, 1985. "An Economic Analysis of Female Work Participation, Education, and Fertility: Theory and Empirical Evidence for the Federal Republic of Germany." Journal Of Labor Economics, 3:1, Part 2: Trends In Women‟s

Ryan, T. (1997). Modern Regression Methods. Wiley, Inc., New York.

Seife Dendir (2004): Unemployment Duration in Poor Developing Economies: Evidence From Urban Ethiopia, Radford University, USA.

Serneels P. (2004), the Nature of Unemployment in Urban Ethiopia, CSAE Working Paper, University Of Oxford.

Serneels P. (2007), Unemployment in Urban Ethiopia, Blackwell Publishing Ltd, University Of Oxford.

Snijders T. A.B. and Roel J. Bosker (1999). "An Introduction to Basic and Advanced Multilevel Modeling". Department Of Statistics, University Of Poone.

Elster, S., and Kamlet, M.S. (1990). "Reference Groups, Occupational Class and Married Women‟s Labor Force Participation." Research in Human Capital and Development, 6, pp. 89-123.

Tansel A. And Tasci, H.M (2004), Determinants of Unemployment Duration For Men And Women In Turkey, Metu, Ankara-Turkey

The Encyclopedia Americana International 1995. 27, 316

UNDP (1995), Fourth World Conference on Women in Beijing.

UNECA (2005). Economic Report on Africa 2005: Meeting the Challenge of u unemployment and Poverty in Africa Addis Ababa. Ethiopia.

United nation (1996) International Standard Industrial Classification of All Economic Activities.3rd Revision, Statistical Papers, Series M, No.4. Rev.3. New York

Winter, C. (1992), "Female Earnings, Labor Force Participation and Discrimination in Venezuela, 1989." In Case Studies on Women‟s Employment and Pay in Latin America. World Bank Report, No. 12175, Washington, D.C., Pp. 463-475.

Appendixes

Model Summary

Table 4.11: Test of Significance of Individual Predictors for Logistic Regression Using Score Test

Summary of Stepwise Selection			
Effect	**DF**	**Score Chi-Square**	**Pr > Chisq**
REGION	10	854.1272	<.0001
PRESIDE	1	445.0730	<.0001
AGE	2	257.9592	<.0001
MARITAL	1	172.6103	<.0001
MASSMEDIA	2	134.3116	<.0001
HOCCU	2	79.7034	<.0001
SEXHHEAD	1	52.6974	<.0001
ECOSTATUS	2	31.8354	<.0001
EDUCATION	2	23.0730	<.0001
PREGNANCY	1	11.2117	0.0008
PRCHILD	1	11.2316	0.0008

Table 4.12: Test of Overall Significance of Individual Predictors in Random Coefficient Model

The GLIMMIX Procedure Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F- Value**	**Pr > F**
EDUCATION	2	20	10.08	0.0009*
MASSMEDIA	2	20	8.89	0.0017*
ECOSTATUS	2	20	12.67	0.0003*
HOCCU	2	20	25.54	<.0001*
AGE	2	20	129.10	<.0001*
PLRESIDE	1	9	32.97	0.0003*
ADDICTION	1	10	3.63	0.0860
HHSIZE	2	20	2.38	0.1186
PRCHILD	1	10	9.73	0.0109*
MIGRATION	1	10	0.00	0.9844
SEXHHEAD	1	10	45.94	<.0001*
PREGNANCY	1	10	11.59	0.0067*
MARITAL	1	10	63.96	<.0001*
HEDU	2	20	0.68	0.5169
DECISSION	1	10	0.99	0.3436

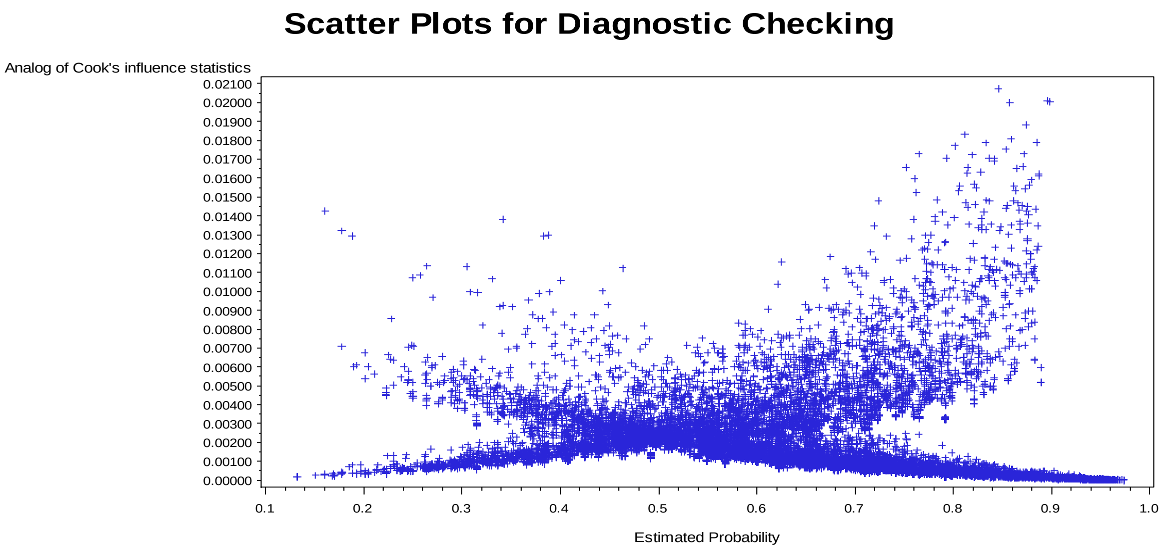
***= Significant at 5% Level**

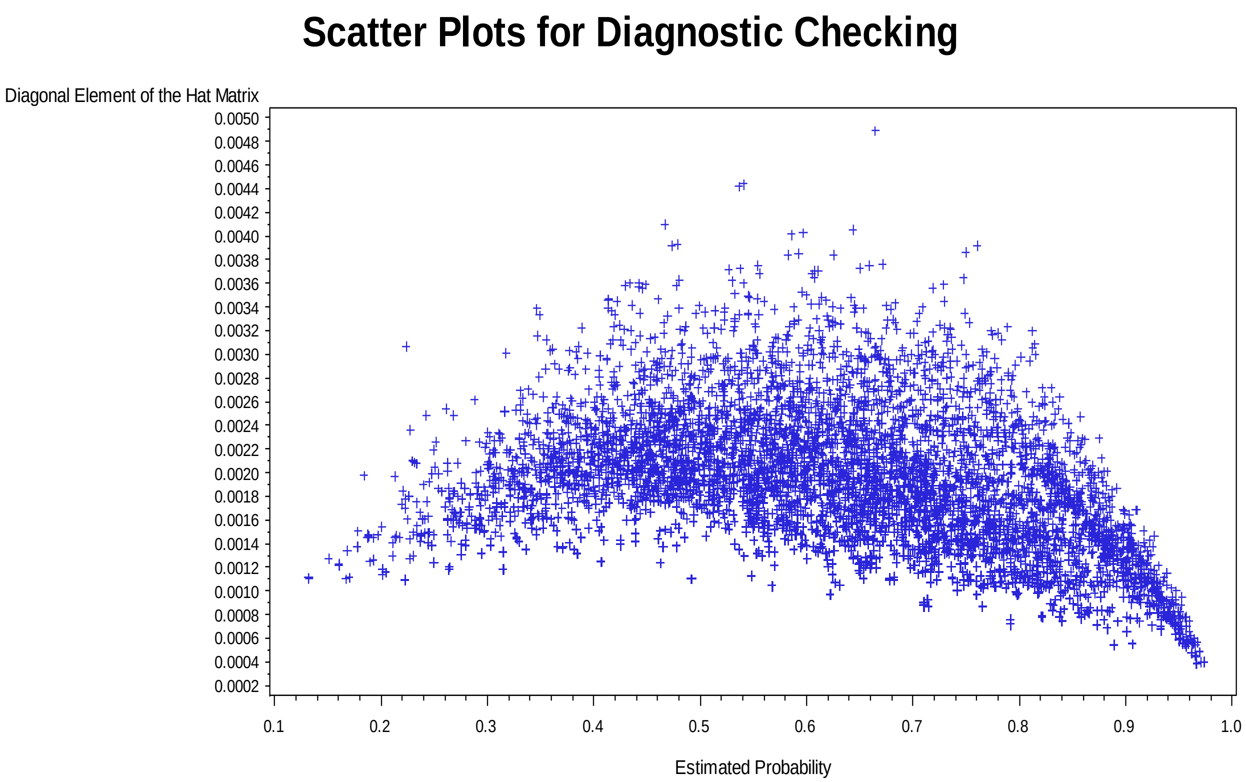
Appendix A: Result of Diagnostic Tests for Outliers and Influential Value for Standard logistic regression

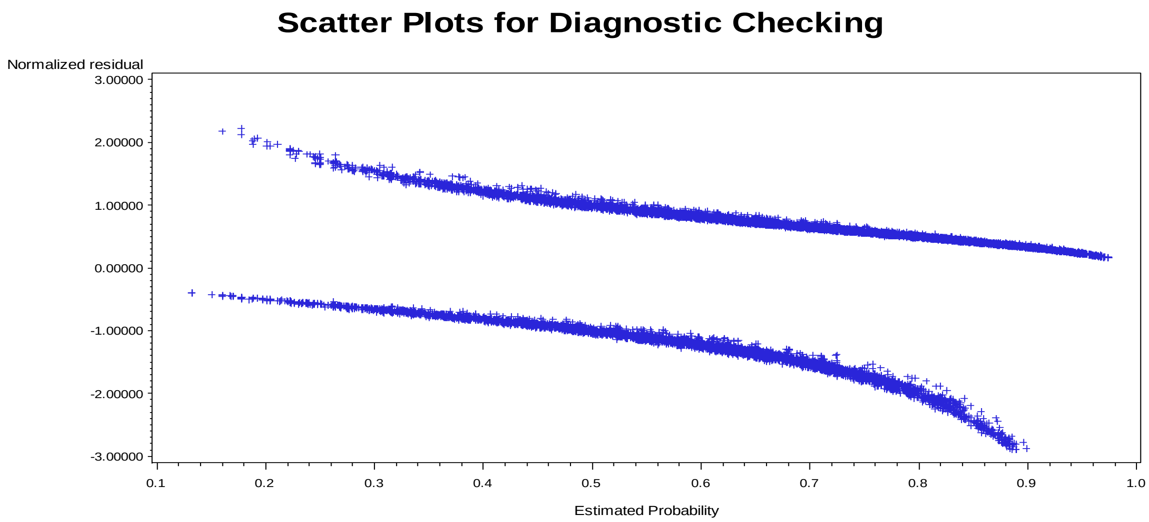
The MEANS Procedure				
Variable	Label	N	Minimum	Maximum
COO_1	Analog of Cook's influence statistics	16386	9.7203296E-6	0.0207443
phat	Estimated Probability	16386	0.1322376	0.9738963
chires	Pearson Residual	16386	-2.9780334	2.2859585
devres	Deviance Residual	16386	-2.1397974	1.9124177
hat	Diagonal Element of the Hat Matrix	16386	0.000384184	0.0048891
SRE_1	Standard residual	16386	-2.1173626	1.8908816
ZRE_1	Normalized residual	16386	-2.8974963	2.2242790
DFB0_1	DFBETA for constant	16386	-0.0121090	0.0146914
DFB1_1	DFBETA for Region(1)	16386	-0.0032886	0.0048227
DFB2_1	DFBETA for Region(2)	16386	-0.0087658	0.0047903
DFB3_1	DFBETA for Region(3)	16386	-0.0028893	0.0050216
DFB4_1	DFBETA for Region(4)	16386	-0.0030105	0.0050859
DFB5_1	DFBETA for Region(5)	16386	-0.0081103	0.0065432
DFB6_1	DFBETA for Region(6)	16386	-0.0029197	0.0050798
DFB7_1	DFBETA for Region(7)	16386	-0.0031359	0.0049859
DFB8_1	DFBETA for Region(8)	16386	-0.0040296	0.0050314
DFB9_1	DFBETA for Region(9)	16386	-0.0037938	0.0040297
DFB10_1	DFBETA for Region(10)	16386	-0.0031365	0.0036911
DFB11_1	DFBETA for Education(1)	16386	-0.0024934	0.0044022
DFB12_1	DFBETA for Education(2)	16386	-0.0019363	0.0028097
DFB13_1	DFBETA for MassMedia(1)	16386	-0.0017862	0.0016479
DFB14_1	DFBETA for MassMedia(2)	16386	-0.0015841	0.0015750
DFB15_1	DFBETA for EcoStatus(1)	16386	-0.0022018	0.0019460
DFB16_1	DFBETA for EcoStatus(2)	16386	-0.0029888	0.0021919

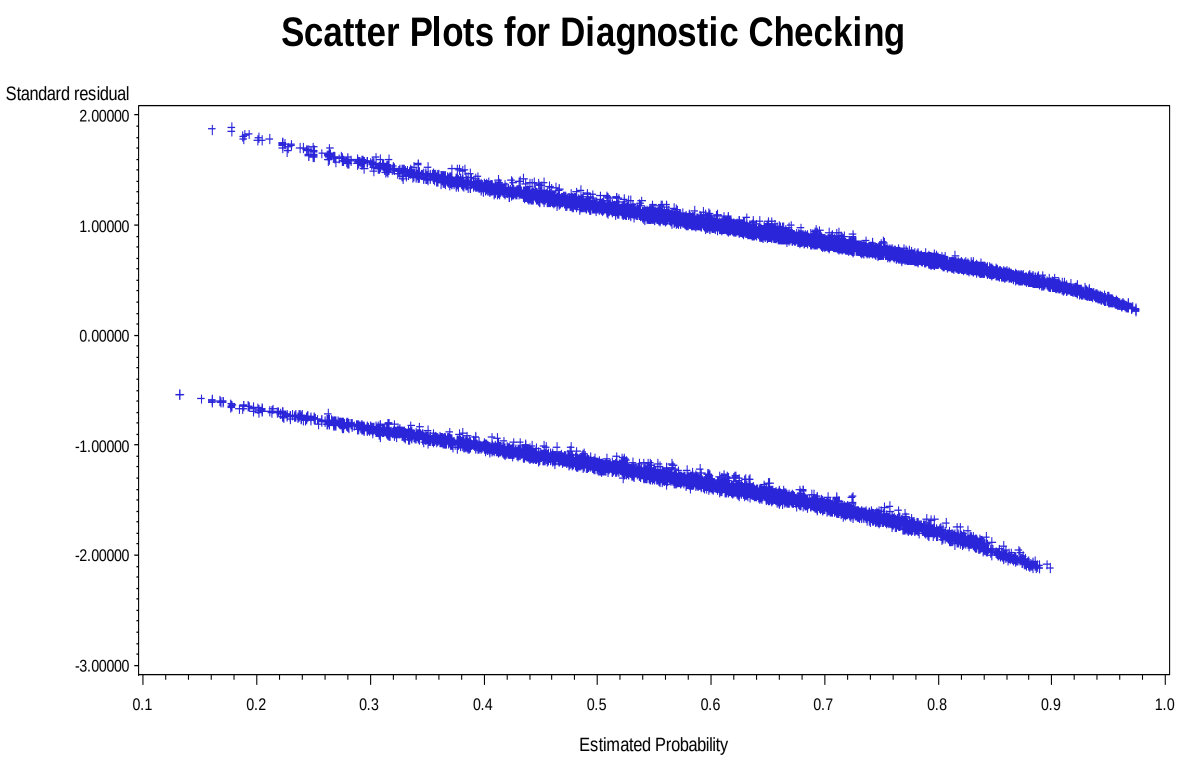
DFB17_1	DFBETA for Hoccu(1)	16386	-0.0045564	0.0035263
DFB18_1	DFBETA for Hoccu(2)	16386	-0.0022518	0.0025190
DFB19_1	DFBETA for Age(1)	16386	-0.0024210	0.0024555
DFB20_1	DFBETA for Age(2)	16386	-0.0017437	0.0016259
DFB21_1	DFBETA for Plreside(1)	16386	-0.0025783	0.0030394
DFB22_1	DFBETA for Addiction(1)	16386	-0.0079826	0.0084927
DFB23_1	DFBETA for HHsize(1)	16386	-0.0071740	0.0078400
DFB24_1	DFBETA for HHsize(2)	16386	-0.0069782	0.0077497
DFB25_1	DFBETA for Prchild(1)	16386	-0.0019222	0.0019315
DFB26_1	DFBETA for Migration(1)	16386	-0.0079896	0.0073281
DFB27_1	DFBETA for SexHHead(1)	16386	-0.0016038	0.0013423
DFB28_1	DFBETA for pregnancy(1)	16386	-0.0034527	0.0042484
DFB29_1	DFBETA for Marital(1)	16386	-0.0022837	0.0022574
DFB30_1	DFBETA for Hedu(1)	16386	-0.0026512	0.0040628
DFB31_1	DFBETA for Hedu(2)	16386	-0.0029373	0.0039033
DFB32_1	DFBETA for Decision(1)	16386	-0.000612483	0.00059150

APPENDIX B: Scatter Plots for Diagnostic Checking for Standard Logistic Model

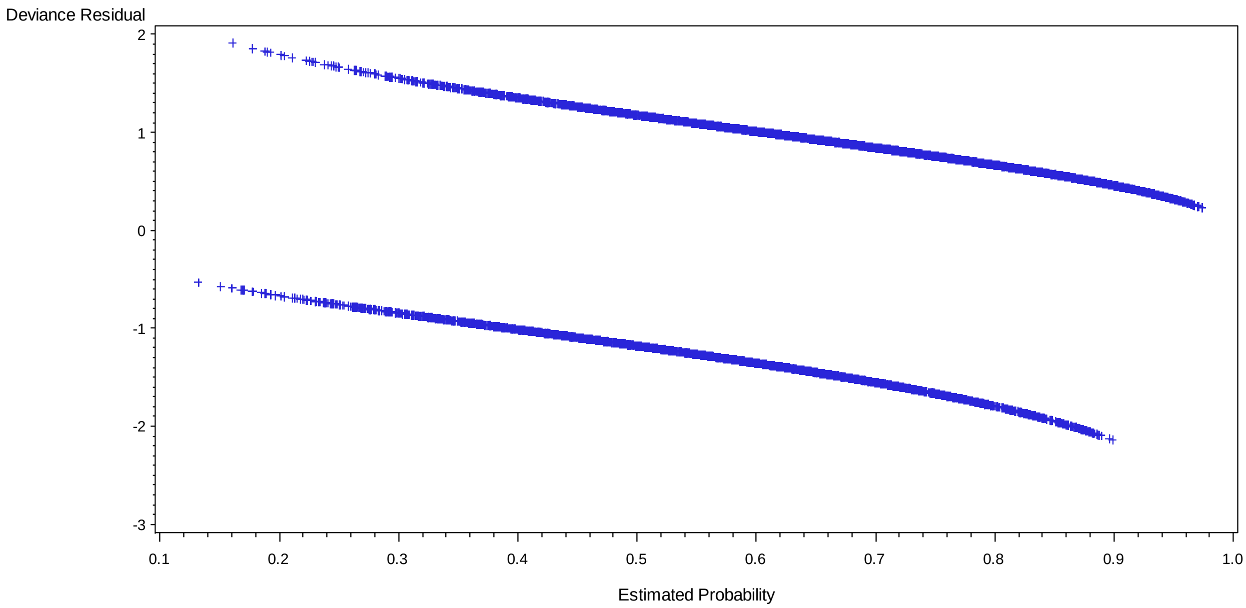




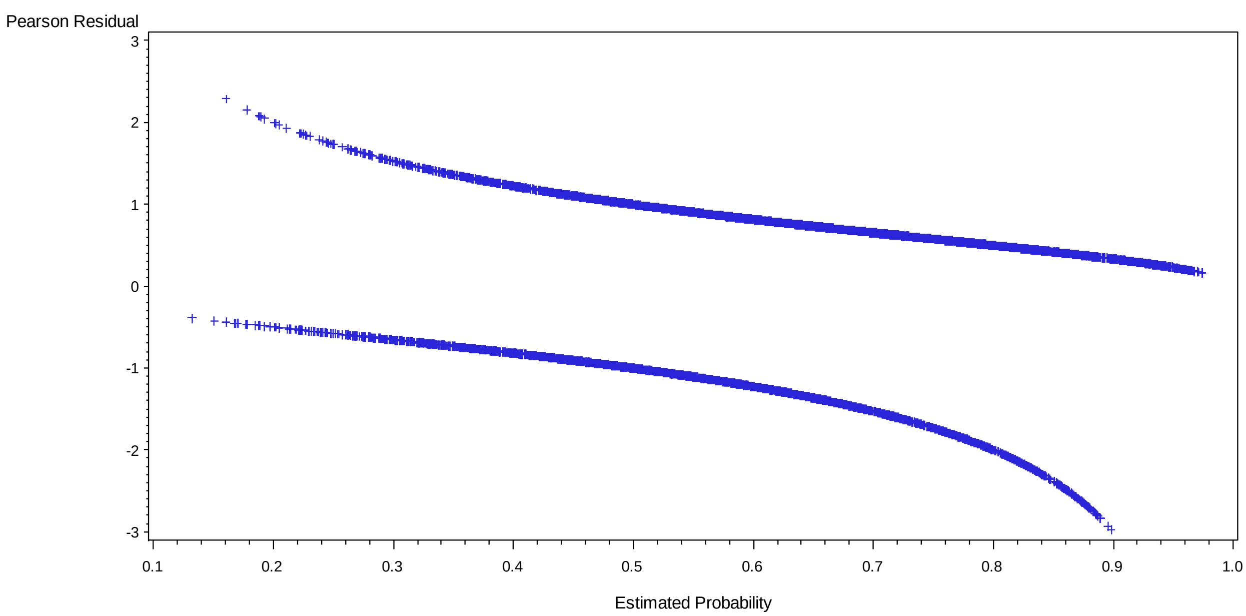

Scatter Plots for Diagnostic Checking
Normalized residual
3.00000
2.00000
1.00000
0.00000
-1.00000
-2.00000
-3.00000
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Estimated Probability


Scatter Plots for Diagnostic Checking
Standard residual
2.00000
1.00000
0.00000
-1.00000
-2.00000
-3.00000
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Estimated Probability

Scatter Plots for Diagnostic Checking

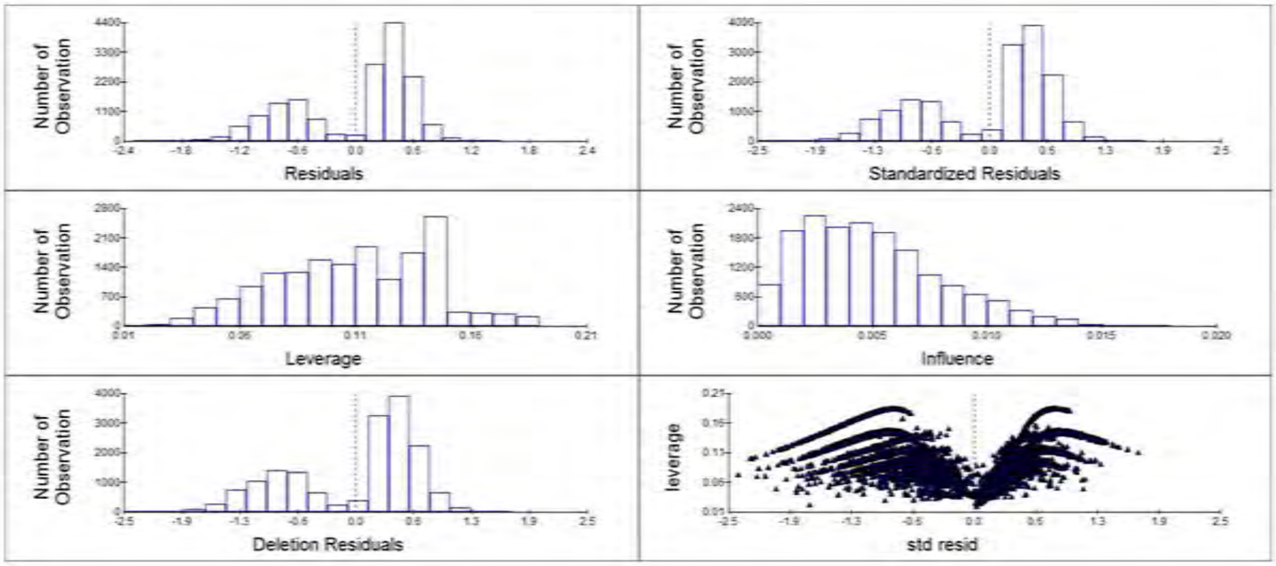


Scatter Plots for Diagnostic Checking

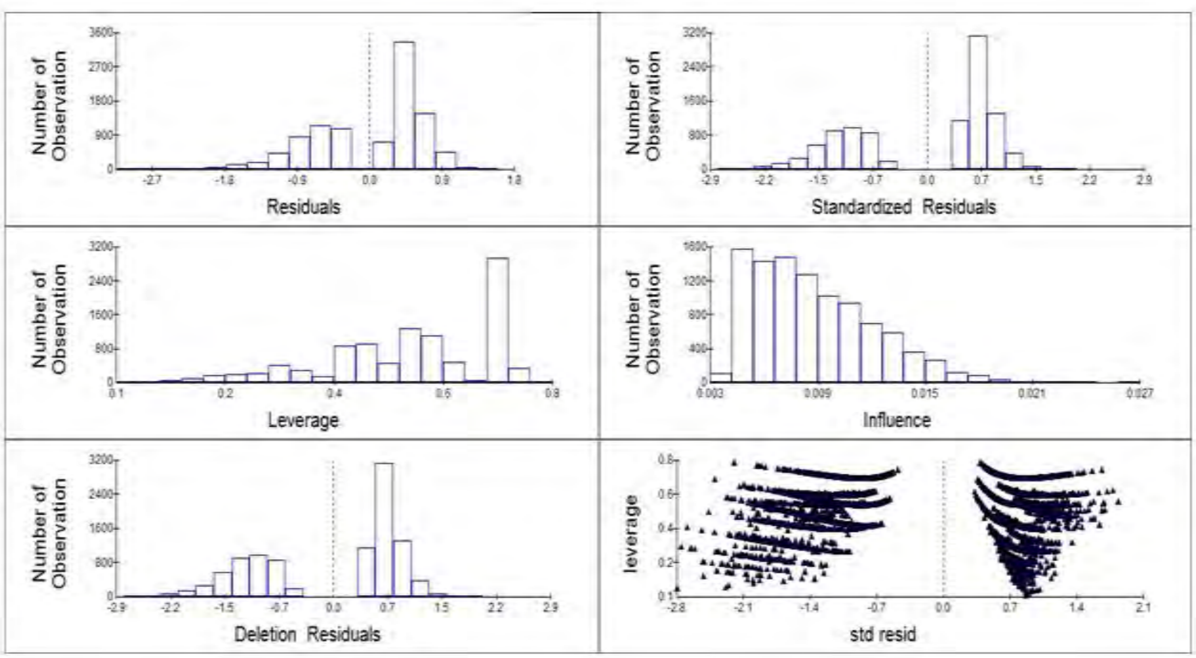


Appendix C: Diagnostic Checking for the Multilevel Model

1. Residual Plots for Level-One



2. Residual Plots for Level-Two



Declaration

I, the undersigned, declare that the thesis is my original work, has not been presented for degree in any other university and that all sources of material used for the thesis have been duly acknowledged.

Declare by

Name---------------------------------------

Signature------------------------------------

Date--

This thesis has been submitted for examination with my approval as a University advisor.

Name-------------------------------------

Signature---------------------------------

Date--------------------------------------