



Addis Ababa University

College of Natural and Computational Sciences

Graduate Studies program

Department of Statistics

**FACTORS ASSOCIATED WITH CHILD LABOUR IN ETHIOPIA: A MULTILEVEL
MODEL APPROACH**

BY

BEKALU MEHARI

ADVISOR: MEKONNEN TADESSE (Associate Prof.)

A Thesis submitted to the department of statistics of Addis Ababa University in partial fulfillment of the requirements for the Degree of Master of Science in Statistics (Biostatistics)

June, 2017

Addis Ababa, Ethiopia

Addis Ababa University
School of Graduate Studies

This is to certify that the thesis prepared by Bekalu Mehari, entitled: *Factors Associated with Child Labour 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

DECLARATION

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

Name: Bekalu Mehari

Signature: _____

Place of submission: Department of Statistics, Collage of Natural and Computational Sciences, Addis Ababa University

Date of submission: June, 2017

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

Mekonnen Tadesse (Associate Prof.) _____

Advisor's Name

signature

Abstract

FACTORS ASSOCIATED WITH CHILD LABOUR IN ETHIOPIA: A MULTILEVEL MODEL APPROACH

BEKALU MEHARI

ADDIS ABABA UNIVERSITY, 2017

Despite the world's promises to care for every child, the scourge of child labour still leaves countless children deprived of their most basic rights. The causes of child labour remain debatable. Child labour is a complex problem whose roots are deeply embodied in cultural, social, political, and economic structures and traditions of societies. Child labour is a serious problem in the world today. The problem is very severe in Sub-Saharan African countries. Among the Sub-Saharan African countries, Ethiopia is one of the countries with a high incidence of child labour. The major objective of this study is to identify and examine the key factors associated with child labour in Ethiopia. The study focused on children aged between 5 and 14 years using the 2013-ENLFS cross-sectional dataset. Descriptive statistics, binary logistic regression and multilevel logistic regression were used for analysis. The observed sample data revealed that 32% of 5-14 years old children were engaged in child labour. The binary logistic regression analysis results revealed that region of residence, sex of the child, age of the child, place of residence, child's relationship with the household head, literacy status of the household head, employment status of the household head and the household size are significant factors associated with child labour in Ethiopia. Conversely, school attendance of the child, marital status of the household head and sex of the household head were not associated with child labour. The results of the multilevel logistic regression indicated that the effect of place of residence varied across regions of Ethiopia.

Keywords: Child Labour, Ethiopia, Multilevel Logistic Regression Models, Odds Ratio

Acknowledgement

Firstly and foremost, thanks to the almighty God for granting me his limitless care, love and blessing all along the way.

Next I would like to thank with deep appreciation to my advisor Mekonnen Tadesse (Associate Prof.) for his valuable advice, suggestion and constructive comments which made the work more meaningful.

Special thanks go to my friends in CSA; Teshome Adno, Abay Getachew, Dawit Berhanu, Mulugeta Okubay, Seid Jemal, Sharew Ararsa and Zemecha Abdela for making the data available and for giving me enough information about the data.

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

TABLE OF CONTENTS

Abstract	i
Acknowledgment	ii
Acronyms	vi
List of Tables	vii
CHAPTER ONE	1
1. INTRODUCTION	1
1.1. Background of the study	1
1.2. Statement of the problem	5
1.3. Objective of the study	6
1.4. Significance of the study	7
1.5. Limitations of the study	7
1.6. Organization of the study	7
CHAPTER TWO	9
2. REVIEW OF LITERATURE	9
2.1. Introduction	9
2.2. Theoretical Viewpoints of child labour	9
2.3. Empirical Viewpoint	10
CHAPTER THREE	15
3. DATA AND METHODOLOGY	15
3.1. Data source	15
3.2. Variables to be included in the study	15
3.3. Method of data analysis	17
CHAPTER FOUR	36
4. STATISTICAL DATA ANALYSIS AND RESULTS	36
4.1. Descriptive Analysis	36
4.2. Bivariate Statistical Analysis	39
4.3. Determinants of Child Labour: Results of Logistic Regression Analysis	39
4.4. Goodness of Fit and Model Diagnostics for logistic regression	43
4.5. Multilevel logistic regression analyses	45
4.6. Goodness of Fit Test	54

CHAPTER FIVE	55
5. DISCUSSION, CONCLUSION AND RECOMMENDATIONS	55
5.1. Discussion	55
5.2. Conclusion	56
5.3. Recommendations	56
5.4. Areas of further research	57
References:	58
Appendices	62

Acronyms

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CSA	Central Statistical Agency
ECLS	Ethiopia Child Labour Survey
EDHS	Ethiopia Health and Demographic Survey
ENLFS	Ethiopia National Labour Force Survey
HHH	Household Head
ICC	Intra Class Correlation Coefficient
ILO	International Labour Organization
IPEC	International Programme on the Elimination of Child Labour
LR	Likelihood Ratio
ML	Maximum Likelihood
MLQ	Marginal Quasi Likelihood
OLS	Ordinary Least Square
PQL	Penalized Quasi Likelihood
S.E	Standard Error
MOLSA	Ministry of Labour and Social Affairs
NAPE	National Action Plan of Ethiopia
SC	Save the Children International
SNNP	Southern Nations Nationalities and Peoples
SPSS	Statistical Package for the Social Sciences
UN	United Nations
UN CRC	United Nations Convention on the Rights of the Child

List of Tables

Table 3.1. Description of Independent variables and coding.....	17
Table 4.1. Distribution of child-labour status by Socio-Demographic factors.....	37
Table 4.2. Test of Significance of Independent Variables Using Wald Test.....	41
Table 4.3. Hosmer-Lemeshow Test.....	44
Table 4.4. Area under ROC Curve.....	45
Table 4.5. Random intercept only model fit results.....	47
Table 4.6. Estimate of random effect for each region.....	47
Table 4.7. Parameter Estimates of the Random Intercept and Fixed Slope Model.....	48
Table 4.8. Results of Model Selection Criteria and Log Likelihood Ratio Test.....	49
Table 4.9. Parameter Estimates of the Random Intercept and Random Slope Model.....	51
Table 4.10. Results of Model Selection Criteria and Log Likelihood Ratio Test.....	52

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the study

Child labour doesn't have a universally accepted definition. In most cases it is difficult to differentiate between the concept and definition of child labour. It has been argued by some authors that it is impossible to come up with a definition of child labour that captures all its features due to the complexity of child labour phenomenon.

Often two major conventions, the ILO Convention 138 and the UN Convention on the Right of the Child are used as benchmarks for providing a working definition for child labour (Assefa Admassie, 2000). According to Article 32.1 of the UN Convention on the Rights of the Child, Child Labour is work performed by a child that is likely to interfere with his or her education, or to be harmful to their health or physical, mental, spiritual, moral or social development, the child may be engaged in wage or non-wage work; whether he or she may be working for their family or for others; whether employed in hazardous or non-hazardous occupations; whether employed on a daily or on a contract basis.

International Labour Organization (ILO) defines child labour as work that deprives children of their childhood, their potential and their dignity. It refers to work that is mentally, physically, socially or morally dangerous and harmful to children; and interferes with their schooling.

In addition ILO sets standards to identify child labourers from those children who are engaged in economic activities based on the following criteria (1) children between 5-11 years of age who were engaged in economic activity for at least an hour within the reference week, (2) children between 12-14 years of age who were involved in an economic activity for 14 or more hours in the reference week, (3) children between 5-14 years of age who were engaged in hazardous work for at least an hour during the reference week, and (4) children working in any economic activity during night for at least one hour during the period of the survey reference week. In line with these parameters, this study is based on information on children aged 5-14 years. This definition also satisfies the provisions of the Minimum Age Convention (No.138) which prohibits children under 15 years of age to enter into employment. The minimum age of a child worker is 5 years and this is because children under 5 years are assumed to be too young to work, and thus data for

younger children are not available in the 2013-ENLFS. The Central Statistical Agency (CSA) of Ethiopia, in its 2015 ECLS, also used the above standards to identify child labourers from the whole working children. The report of this recent survey is not yet disseminated.

In 2011, the Ministry of Labour and Social Affairs (MOLSA) listed down the following activities as hazardous occupations for children.

- Work in the transport for passengers, and goods by road, rail, air, and internal water way;
- Work in warehouses involving heavy weight-lifting, pulling or pushing or any other related type of labour;
- Work connected with electric power generation plants, transformers or transmission lines;
- Underground work such as in mines, quarries, and similar works;
- Working in construction;
- Work on streets cleaning, in sewers, digging tunnels, sorting and transporting liquid and solid wastes;
- Alcohol and cigarette production;
- Working in hotels, motels, night clubs and similar services;
- Working in very cold and hot environments and where there is radiation and waves;
- Metal welding, grinding, cutting and shaping;
- Cutting wood, splitting, shape making, etc., by using electric powered machines;
- Forest clearance;
- Work in chemicals and with mixtures of minerals that have recognized health hazards;
- Other occupations that have detrimental effects on the young workers' moral and physical development.

The idea of child labour does not oppose children's working at all. Involvement of children in a variety of jobs or economic activities which do not obstruct their education, their health and their holistic growth are often seen as positive. According to the definition of ILO, light work, which do not affect the health and education of children, are allowed from the age of 12 years. Thus, the central issue in child labour is mainly the type of work that the children are doing.

There have always been children involved in various types of activities and duties, ranging from very light and occasional or leisurely work, often serving as an introduction to adult life and work, to the most wretched activity in terms of its type or conditions of work, including the work place environment. Obviously, therefore, while some types of activities or occupations have very little or no negative effects on the working children and could, in fact, be beneficial as a learning and maturing process, other types are exploitative and/or hazardous with detrimental consequences to the schooling, health, spiritual, moral and normal development of the children (Kebebew Ashagrie, 2000). These latter activities violate the provisions of international conventions on the protection of children (ILO, 1999).

Ethiopia has ratified the above key international conventions concerning child labour. In addition, the Ethiopian Government has also established relevant laws and regulations related to child labour, including its worst forms.

Furthermore, the National Action Plan was prepared in 2013 in order to mitigate the incidence of child labour and to follow up the implementation of laws and regulations related to child labour. This National Action Plan is a tripartite (i.e. MOLSA, ILO and Employer) agreement document that gives the national definition of child labour according to the national context.

The child labour definition in the National Action Plan of Ethiopia (NAPE, 2013) takes into account school attendance of the child and is slightly different from the child labour definition of the ILO. This study, like many other studies on child labour used the ILO's definition so that the results/findings will be comparable with the results of other similar studies.

Child labour is a complex problem whose roots are deeply embodied in cultural, social, political, and economic structures and traditions of societies (Assefa Admassie, 2000). Child labour remains to be a serious problem in the world today. According to the International Labour Organization (ILO-IPEC, 2013), out of 264 million working children worldwide that are between the ages of 5-17, 168 million children were engaged in child labour. Of this 168 million, about half (85 million) were engaged in hazardous work which directly endangers their health, safety and moral development. The great majority of these are in developing countries particularly; in sub-Saharan Africa (21.4 % of the child population of the region), Asia and Pacific (9.3 %) and Latin America and the Caribbean (8.8 %). Many of these children are exposed to sexual,

emotional and physical abuses and thereby become vulnerable to various childhood emotional and behavioral disorders. In general letting children work at an early age may have negative outcome on the personal development of the children and the economic and social development of their country.

Ethiopia, as a developing country, might also suffer from this phenomenon given its current large proportion of young age population. The population of Ethiopia was 80.4 million in 2013, of which around 29.0% were between 5 and 14 years old. Of this age group of children, 51.4 % were males and 48.6 % were females. Further, 18.3 % of the child population lived in urban areas and the rest 81.7 % lived in rural parts of the country (ENLF, 2013). According to the country regulations, those under 15 are banned to be employed in Ethiopia and the workers between 15 to 18 years old are entitled to special protection (Ethiopian Labour Proclamation, 2004). Child labour has also been one of the main reasons for low school enrolment in rural Ethiopia (Assefa Admassie, 2000).

Child labourers are deprived of freedom, childhood, education, fun and play, and natural development. In principle, childhood should be a period of school learning, of recreation, of physical, mental and social development. Thus, understanding child labour is important not only because young children are made to undertake work obligations that may be beyond their physical capability but also because of its long term effect on human capital formation and on the future welfare of the children (Ravallion and Wodon, 1999).

To my knowledge, little research has been done on determinants of child labour in Ethiopia. All studies conducted so far to investigate determinants of child labour in Ethiopia did not capture the effect of relationship of a child with the head of the household, literacy status of the household head and gender of the household head which may have important implications to child labour. In addition, regional variability wasn't accounted for in most studies of child labour in Ethiopia. Hence, this study has attempted to investigate and identify factors associated with child labour in Ethiopia by incorporating these variables. Moreover, the study will attempt to identify factors that may explain the variation in child labour among regions of Ethiopia and examine the extent of the variation in child labour within and between regions of the country.

1.2. Statement of the problem

The consequence of child labour greatly contributes to the poverty rate among the community as it increases the dropouts of children from schools; it also decreases the school enrollment rate. In contrast, education prepares one with life skills which enables him/her to move from poverty to prosperity, it is a part of any solution of eliminating /reducing the child labour (Abas, 2014).

Similar to other African countries, child labour is a problem in Ethiopia. Different researchers have indicated that the situation is very severe in the rural parts of the country (Assefa Admassie, 2000; and Bisrat Abebe, 2014). The 2013-ENLFS also confirmed that the situation of child labour is very high (49.2%) in rural parts of the country and relatively low (7.7%) in the urban parts of the country. In Ethiopia, child labour is the major problem that children face even if it is not given adequate emphasis as the issue requires (Save the Children, 2010). According to Guarcello and Rosati (2007), Ethiopia had one of the highest rates of child labour in the world.

The Ethiopia Demographic and Health Survey (EDHS-2011) reported that about 27% of children under 15 (with 17% for children aged 5-11 and 55% for children aged 12-14) were engaged in child labour. Most of these children were involved in household chores, followed by family businesses. However, paid work was very limited. Less than 1% of children aged 5-11 and 2% of children aged 12-14 were engaged in paid work, and 14% of 5-11 year olds and 30 % of 12-14 year olds worked for a family business. Furthermore, 18% of 5-11 year olds and 40% of 12-14 year olds were engaged in household chores for 20 hours or more per week. In gender terms, child labour was higher among boys (31%) than girls (24%), and much higher in rural areas (30%) than urban areas (13%). Child labour was found to be much less prevalent among children with educated mothers (6%) and decreased with increasing wealth quintiles.

The 2013-ENLFS reported that 62.3% of the total children aged 10-14 years were involved in economic activities. In terms of sex differences, 66.3% of boys and 59.8% of girls were involved in economic activities. The proportion of urban children aged 10-14 years engaged in economic activities was higher for boys (20.4%) than for girls (18.0%). Similarly, the proportion of rural children aged 10-14 years engaged in economic activities was higher for boys (73.5%) than for girls (68.5%).

Thus high number of children were engaged in economic activities and many of them were considered as child labourers because they are either under the minimum age for work, or above that age but engaged in work that poses a threat to their health, safety or morals, or are subject to conditions of forced labour (ILO, 2010). This high incidence of child labour occurrence might be related to several factors like poverty, inequality, socio-economic vulnerability, inadequate and inappropriate education opportunities and cultural norms and values.

What explains the high child labour rate in Ethiopia is not clear as very few studies have been conducted about the problem of child labour. Any program aimed at improving the welfare of children cannot be successful unless the key determinants of child labour are identified and investigated, there is a need to have a clear understanding of the factors and nature of child labour in Ethiopia. Without this knowledge, it would be difficult to formulate policies and interventions that will help to eliminate/reduce the phenomenon of child labour across the country. Therefore, this study will attempt to identify the major determinants of child labour in Ethiopia and answer the following major research questions depending on the availability of indicators in the 2013-ENLFS.

1. What are the major determinants of child labour in Ethiopia?
2. Does the age of a child determine whether he/she is engaged in child labour?
3. Is it more likely for males to be engaged in child labour than females?
4. Is it more likely for children residing in rural areas to be engaged in child labour than children residing in urban areas?
5. Are children currently attending school less likely to be engaged in child labour than children not currently attending school?
6. Is there variation in child labour among regions of Ethiopia?

1.3. Objective of the study

General objective:

The general objective of this study is to investigate and identify the socio-demographic determinants of child labour in Ethiopia by employing single and multilevel logistic regression models.

Specific objectives:

1. To identify socio-demographic determinants associated with child labour.
2. To examine the extent of the variation in child labour status within and between regions of Ethiopia.
3. To compare results from traditional (single-level) model with results from multilevel model in identifying the determinants of child labour.

1.4. Significance of the study

This study will have the purpose of identifying the major contributing risk factors that force children to be engaged in work in Ethiopia. Previous literature suggest that determinants of child labour are largely country specific, indicating that any policies aimed at reducing child labour must look carefully at the determinants of child labour in context. The results of this study will be helpful for the formulation of policies and strategies that will facilitate the reduction of child labour in the country. The findings may also provide helpful information for those interested in conducting further studies on the issue.

1.5. Limitations of the study

- It is evident that household income is a very important determinant of whether children are engaged in child labour or not. But, unfortunately in the ENLFS-2013 dataset, the available data on the monthly income of employed family members excludes the self-employed (which represents above 70 % of the total sampled households). So in this study the explanatory variable ‘monthly income of the household’ couldn’t be incorporated as a determinant of child labour.
- The other limitation of this study is that only children who are living with their families or households were targeted in the survey. Children who were living in institutions, factories, in the street and any other places outside households were not included in the 2013-ENLFS. As the survey is a household-based survey this study couldn’t get information from children who were not living with households.

1.6. Organization of the study

The research was categorized under five chapters. The first chapter deals with the background of the study, statement of the problem, purpose of the study/objectives, significance of the study,

limitations of the study, and organization of the study. The second chapter deals with a review of related literature. The third chapter describes the source of data, variables of the study and the methodology used for analysis. The fourth chapter provides the results of the study and their interpretations. Finally, chapter five presents conclusion and recommendations.

CHAPTER TWO

2. REVIEW OF LITERATURE

2.1. Introduction

In undertaking empirical studies, researchers need to read related materials to the problem or topic to find out what has already been written about the topic of the studies. The review process involves a systematic identification, location and analysis of documents that contain information which are related to the research problem. The idea is to avoid unplanned replication, identify useful methodology techniques, develop research hypotheses, and identify contradictory findings and to facilitate interpretation of results of the study. This chapter reviews available literature on child labour.

2.2. Theoretical Viewpoints of child labour

The labour market perspective is mostly concerned about the possible impact of child labour on adult labour markets. It argues that the relationship between child labour and adult employment is a negative relationship. This labour market perspective points out that adult unemployment and child servitude worsen working class poverty due to the reason that child labour would succeed adult labour. Since child labour increases the supply of work force in the labour market, it will reduce salary rates and increase adult joblessness. The perspective advocates policies that discourage economic participation of children in order to protect adult employment and wages. This labour market perspective argues that the state has the highest responsibility in eradicating child labour by using compulsory education and minimum prohibitions on work (Bachman, 2000).

Another strand of theoretical explanation on why children do work is the poverty hypothesis. Families that allocate their time optimally between various forms of work and school presumably compare the current value of the child's labour against the future value of increased productivity of an educated worker. There is no particularly compelling reason why the productivity gains from educating a child from a poor family should be any larger or smaller than the gains for a child from a high-income family. Nevertheless, poverty could have a direct effect on schooling decisions. Families who are barely surviving are likely to discount the future heavily, thereby giving less weight to future income earned by their educated children (Drusilla, 2001).

2.3. Empirical Viewpoint

Tharmpornphilas (2009) investigated the various factors that influence a household's decision of sending a child to work. The study presented a detailed empirical analysis of the determinants of child labour in Thailand. Econometric analysis was carried out using data from the Thailand labour force survey (National Statistics Office Thailand, 2003). A multiple regression analysis was done using number of hours children who worked in the last 7 days before the survey (child time of work) as the response variable. The explanatory variables were classified as: 'children' characteristics that are age, sex and wage; 'household' characteristics: household's monthly income, region of residence, number of children, gender of household head, age of household head, parental education, and occupation of household head. However, the 'school' and 'community' characteristics were not incorporated due to the limitation of the dataset. The results showed that wage impacted significantly on the time that children allocate to work. Age had a significant effect on boys but insignificant on girls. This implied that the older a boy became, the more time he would allocate to work. Boys and girls in urban areas were found to work fewer hours compared to their counterparts in the rural areas. The effect was higher on boys than in girls. The size of the household affected the working time of children positively. This implies that the larger the households the more their children worked. Girls tend to benefit from household's head age. This implies that older boys worked more compared to girls. Educated parents were found to allocate fewer working hours for their children. The occupation of the household head was also found to affect the working time of children. Children from households in which occupation of the household head is related to agriculture were found to be involved in some form of work.

Ahmed et al. (2012) studied determinants of child labour in Khyber province in Pakistan. The research was based on cross-sectional data collected from one hundred sampled households through field survey out of which 50 were taken from schools and 50 from market. The binary choice model was used to analyze the determinants of child labour. The results of the econometric model show that education of the head of the household and household's average income are significantly and negatively correlated with child labour. Age of the child and family size are insignificantly and positively correlated with child labour. The results of the study concluded that parental education is necessary for better future of children.

Laurent and Sebastian (2010) investigated characteristics and determinants of child labour in Cameroon, using data from the Cameroonian survey on employment and informal sector. In this study, Binary Probit and Tobit models were employed using child time of work as the dependent variable. Independent variables used were grouped into: individual's characteristics which are: the sex, the age; the relationship with the household (biological child of the household head). Household environment: the household income, the residence, the size and the composition of household. Finally, household head's characteristics: the level of education, the type of employment, the age and sex. They found in the estimated models, an increase in income together with increase in household size resulted to a reduction in the time that a child spent on work. Also their study found that children who resided in urban areas together with an increase in the level of education of household head could result to a reduction in the time spent by children working. All the variables in the study were found to be statistically significant except the level of education of the household head which was statistically insignificant.

Coster and Adekoya (2014) examined determinants of child labour participation and schooling choice of children aged 5-14 years. Using data from 212 rural households randomly selected through multistage random sampling technique from ten rural communities in Ogun State, Nigeria. Descriptive and inferential statistics were used to study the data. The results show that about 50% of mothers had no formal education. About 52.8% of the sample households are low income earners. The study showed that child labour participation is gender sensitive and that child labour and schooling increases with age. Male children of different age categories attended school more than their female counterparts. Even though children work to supplement family needs, they are likely to be at receiving end because of early entry into labour force which may affect their health. There is tradeoff between child labour and schooling. The results revealed that parent, child, household and community characteristics determined child labour and schooling choices of children. The findings from this study affirmed the relationship between child labour and poverty.

Moyi (2011) examined the causes and magnitude of child labour in Kenya. The data used for this study was drawn from the second round (2000) of the Multiple Indicator Cluster Survey (MICS). Multinomial logistic regression was used as analytical tool. The variables used in this study were classified as children, household, and community characteristics that influence child labour and

school participation. Age, gender, and child's relationship to the head of the household and the children's number of siblings, gender of the head of household, and education of the head of household, have impact on school and/or work participation. The study found children's participation in work and/or school to be affected by their age and gender, how they are related to the household head, education level of the household head, wealth of the household and the children present in the household. Although the study found that children of the household head were less likely to be working exclusively and attending neither school nor work; however, they had a higher probability of combining both school and work. Urban children were found to combine school and work four times less likely compared to their rural counterparts. The study recommended that policy makers formulate policies which would factor education inequality dimension between children who combine work and school and those who do not combine if the effect of working is going to hinder children from attaining education.

Mudzongo and Whitsel (2013) attempted to identify determinants of child labour in Malawi and Tanzania. The study also aimed to investigate the effect of household and community characteristics in determining the likelihood of children participating in child labour. Data from the Tanzania National Panel Survey (TNPS) (2008- 2009) (National Bureau of Statistics Republic of Tanzania, 2009) and the Integrated Household Survey of Malawi (2010-2011) (National Statistical Office of Malawi (NSO, 2011) were used. The target group of the study was children aged 5–14 years. The study used logistic regression modelling to identify the effect of factors at individual, household and the community levels. The findings showed that older children in Tanzania are more likely to be engaged in child labour than younger children while females are less likely to be engaged in child labour in Malawi than boys. However, there was no significant difference between genders in Tanzania. Child's school enrollment factor is negatively associated with child labour in Malawi and Tanzania. Household factors have no effect on child labour for both countries. When analyzed the community factors only, the study found that children who reside in rural areas are more likely to be child labourers than children who reside in urban areas. School enrolled children were less likely to be child labourers. The study concluded that school enrollment is the most significant negative predictor of child labour in both Malawi and Tanzania. The second general finding stated that the availability of a market place in the community leads to greater likelihood that children from both the Malawi and Tanzania are engaged in child labour. In Malawi, the father's education factor is negatively

associated with child labour. When the father has been to school, the children are less likely to be child labourers

The result of the Ethiopian National Child Labour Survey (2001) shows that the total children population of the country in the age group of 5-17 was 18,197,783. Out of this, 9,483,611 children were engaged in productive activities during the reference period, consisting 52.1% of the child population in the given age group. The participation rate in productive activities was 62% for boys and 41.9% for girls. For domestic activities, this figure was 22% for boys and 44% for girls. In rural areas, children were more frequently engaged in productive activities than in domestic activities, whereas in urban areas the opposite was true. Desegregated data by education status of children aged 5-17 years indicate that 43.9% have attended school while 56.1% have never gone to school. The school status of children aged 5-17 years indicated that the majority, about 61.7%, were not at school or have dropped out of school for various reasons.

The 2000-EDHS was used by Dawit Senbet (2010) to look into determinants of child labour versus schooling in rural Ethiopia. The findings suggested that incidence of child labour versus schooling depended on age of the child, education attained, proximity to water, age/gender of the household head, presence of infants (for girls), household size, ownership of cropland (for boys) and ownership of cash crops. A child was less likely to work if he/she was younger, and had more schooling, was a direct offspring, was in a larger household (for boys) and the household head was younger (for girls). The probability for girls to work increased especially with the existence of infants in the household, and boys in wealthier households seemed to engage in working at the cost of schooling, while the opposite was true for girls. Ownership of cropland strongly increased the likelihood of boys engaging in work while the reverse was true for girls. In contrast ownership of cash crops favored boys' going to school instead of working.

Save the children Denmark carried out a study on "Child Labour in Ethiopia with special focus on Child Prostitution" in 2003. The study result stated that poverty, death of one or both parents, the need to become economically independent, and violence within the home environment are major causes which push children to engage in child labour in Ethiopia. Child labour exposes the child into various health problems. However, most of the study subjects believe that children from poor families should work, but under conditions which do not affect their health conditions. In order to reduce and prevent child labour, alleviation measures such as: poverty reduction,

awareness creation, family support, fulfilling the basic needs of children, and family planning were recommended as essential and critical.

Solomon Sorsa and Alemu Abera (2006) conducted a study in three towns in the Southern Nations, Nationalities and Peoples Region (SNNPR): Hawassa, Wolayta Soddo and Arba Minch, that considered 323 working children who were not at school. The reasons for child labour included: poverty (61%), loss of parents (17%), disagreement with parents (8%), parental separation (6%), shortage of food (5%) and displacement due to war (1%). Almost all of the respondents' parents had a low level rank occupation with 64% having a monthly income of less than 50 birr, and 79% of the respondents reported that they were from poor families. Among the respondents, 51% were domestic child labourers, 23% were street child labourers and 18 percent were working in private organizations. Two-thirds of the child labourers were working for more than 10 hours a day and 82% of them had a daily income of less than five birr. About half of them stayed in the job for more than two years and most of them did not visit their parents or relatives for long periods of time. The vast majority, 84%, reported having encountered one or more health problems. Malaria-like illnesses and diarrheal diseases were the major health problems reported. About 19% were sexually active, yet 23% of them have never heard about HIV/AIDS. About three-quarters of them did not attend any kind of health education program. However, it should be noted that the study focused on child labourers and did not compare their condition with children at home and in school.

CHAPTER THREE

3. DATA AND METHODOLOGY

3.1. Data source

The dataset used in this study was obtained from the Ethiopia National Labour Force Survey (ENLFS) conducted by the Central Statistical Agency (CSA) in 2013. The 2013-ENLFS is a nationally representative cross-sectional survey of individuals aged 5 or above. This study used the dataset which provided information about individual, household and community characteristics among children in the 5 to 14 years age group. Data about work status were collected on children as young as five years old since children in rural and urban areas of Ethiopia engage in collection of fire wood, looking after cattle, shoe shining, street vendor, petty trading and the like at their early age (ENLS, 2013).

The survey utilized a multistage cluster sample based on the 2007 Population and Housing Census sampling frame and was designed to obtain and provide information on the basic indicators of the demographic, economic and social characteristics 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).

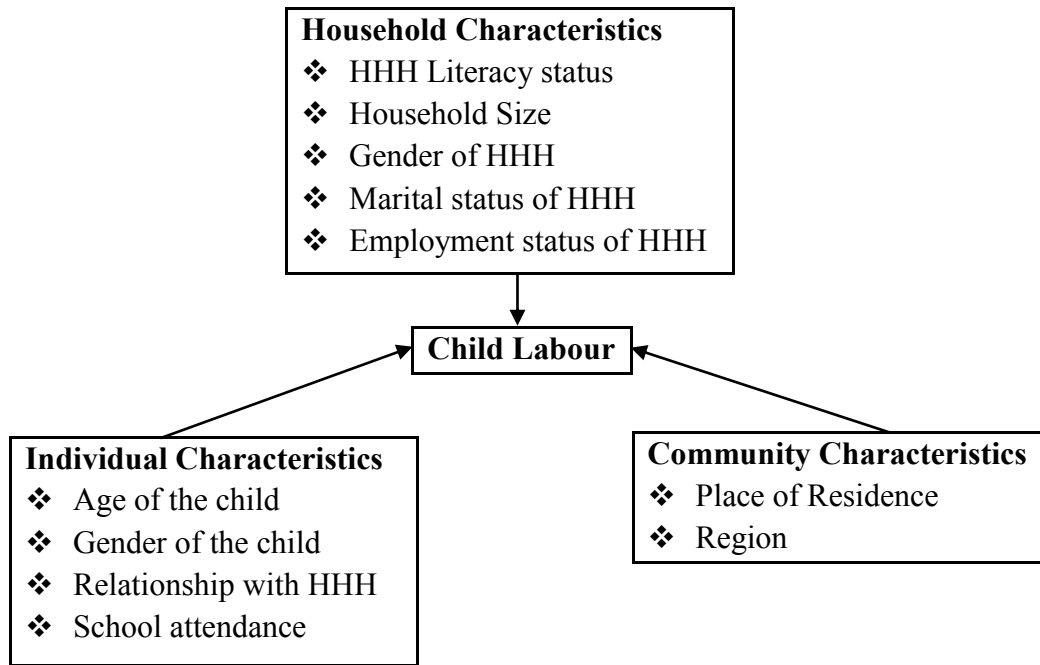
This multistage 2013-ENLFS dataset has a two-level hierarchical structure with children at level 1, nested within 9 regions and 2 city administrations at level 2. The respective relevant dataset for this research was extracted from the 2013-ENLFS.

3.2. Variables to be included in the study

The variables considered in this study include which might be the determinants of child labour in Ethiopia. The potential independent variables that were considered to affect child labour status were selected based on findings of available similar studies and the available data on the subject.

Figure 1 below shows that child labour is a result of a variety of factors from multiple levels, individual, household, and community. The research question is: what are the factors that increase the likelihood of child labour in Ethiopia? Specific hypotheses are formulated on the factors within the three levels of determinants - individual, household and community factors.

Figure 1: Determinants of child labour used in this analysis



3.2.1. The Response variable

The response variable is a binary indicator of whether a child is engaged in any economic activities and considered as a child labourer or not. Therefore, the outcome for the i^{th} child is represented by a random variable Y_i with two possible values coded as 1 and 0 as follows.

$$Y_i = \begin{cases} 1, & \text{if the } i^{th} \text{ child is working and considered as a child labourer} \\ 0, & \text{otherwise} \end{cases}$$

To be considered as a child labourer, a child must be: (1) between 5-11 years of age who is engaged in economic activity for at least an hours within the reference week; or (2) between 12-14 years of age involved in an economic activity for 14 or more hours in the reference week, or (3) between 5-14 years of age who is engaged in hazardous work for at least an hour during the reference week, or (4) between 5-14 years of age involved in any economic activity during night for at least one hour during the survey reference week.

3.2.2. Predictor variables

Independent variables/factors that are expected to influence child labour in Ethiopia are presented in Table 3.1

Variable	Notation	Description	Values/categories
Idreg	x ₁	Region	1=Tigray, 2=Afar, 3=Amhara, 4=Oromiya, 5=Somali, 6=Ben-Gumuz, 7=SNNP, 8=Gambela, 9=Harari, 10=Addis Ababa, 11=Dire Dawa
sexc	x ₂	Sex of the child	0=Female, 1=Male
agec	x ₃	Age of the child	0=5-11, 1=12-14 years
plres	x ₄	Place of residence of the child	0=Rural, 1=Urban
schatt	x ₅	School attendance	0=Not attending, 1=Attending
rship	x ₆	Relationship of children with the HHH	0= Not son/daughter of the HHH, 1=Son/daughter of the HHH
sexhh	x ₇	Sex of the household head	0=Female, 1=Male
lithh	x ₈	literacy status of the HHH	0=Illiterate, 1=Literate
emphh	x ₉	Occupation of the HHH	0=Not employed, 1=Employed
hhmarital	x ₁₀	Marriage status of the HHH	0=Not Married, 1=Married
hhsizec	x ₁₁	Household size	1=Less than or equal to 5, 2= 6-10, 3=Greater than or equal to 11

3.3. Method of data analysis

Descriptive statistics, ordinary binary logistic regression and multilevel logistic regression analyses will be used to study/identify factors associated with child labour. In the multilevel analysis, three multilevel models (an empty model, model controlling for the individual-level variables, and a model controlling for community-level variables) have been constructed.

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, categorical, or a mix of any of these. Hosmer and Lemeshow (2000) have described logistic regression focusing on its theoretical and applied aspects.

Often the outcome variable in the social data is in general not continuous instead it 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 (Agresti, 2002). There are two primary reasons for choosing the logistic regression. First from mathematical point of view it is an extremely flexible and easily used function, and second it lends itself to a clinically meaningful interpretation (Hosmer and Lemeshow, 2000).

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, the dependent variable must be categorical, the categories (groups) must be mutually exclusive and exhaustive; a case can only be in one group and every case must be a member of one of the groups, and

larger samples are needed than linear regression because maximum likelihood coefficients are large sample estimates. A minimum of 50 cases per predictor is recommended.

Logistic regression determines the impact of multiple independent variables presented simultaneously to predict membership of one or other of the two dependent variable categories. There are two main uses of logistic regression. The first is the prediction of group membership. Since logistic regression calculates the probability of success over the probability of failure, the results of the analysis are in the form of odds ratio. Logistic regression also provides knowledge of the relationships and strengths among the variables.

3.3.1.1. The Binary Logistic Regression Model

Consider a collection of k independent variables which will be denoted by the vector (Hosmer and Lemeshow, 2000) $\mathbf{x}' = (x_1, x_2, \dots, x_k)$. Let the conditional probability that the outcome is present be denoted by $P(Y=1 | \mathbf{x}) = p(\mathbf{x})$.

Then, the logit of the multiple logistic regression is given by the equation

$$\text{logit}[p(x)] = \log\left[\frac{p(x)}{1-p(x)}\right] = \log[e^{g(x)}] = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k \dots\dots\dots(3.1)$$

where, $g(x) = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k$

The probability of success can be expressed as

$$\pi_i = P(Y_i / X_{1i}, \dots, X_{ki}) = \frac{\exp(X'\beta)}{1 + \exp(X'\beta)} \dots\dots\dots(3.2)$$

With further rearrangement we obtain the odds of success.

$$\text{odds}(Y_i = 1) = \frac{\pi_i}{1 - \pi_i} = \exp(X'\beta) \dots\dots\dots(3.3)$$

The above three equations give suitable representations of log-odds, the success probability, and odds, respectively. Indeed, these representations facilitate interpretations of parameter estimates. The parameter refers to the effect of x_i on the log odds that $Y = 1$, controlling the other x 's. For

example, the multiplicative effect on the odds of a one-unit increase in x_i at fixed levels of the other variables in the model.

3.3.1.2. Parameter Estimation

The maximum likelihood and non-iterative weighted least squares are the two most competing estimation methods used in fitting logistic regression model (Hosmer- Lemeshow, 1989). When the assumption of normality of the predictors does not hold, the non-iterative weighted least squares method is less efficient (Maddala, 1997). In contrast, the maximum likelihood estimation method is appropriate for estimating the logistic (logit) model parameters due to the less restrictive nature of the underlying assumptions (Hosmer and Lemeshow, 2000). Hence, in this study the maximum likelihood estimation technique will be applied to estimate unknown parameters of the model.

Consider the logistic model, $p_i = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}}$

where, $x_i = (x_{1i}, x_{2i}, \dots, x_{ki})'$ is the vector of independent variables, and $\beta = (\beta_0, \beta_1, \beta_2 \dots \beta_k)'$ is the vector of coefficients.

Since observed values of Y say, y_i 's ($i= 1, 2 \dots n$) are independently distributed as Bernoulli with parameter p_i , $y_i \sim \text{Bernoulli}(1, p_i)$, the likelihood function of Y is given by:

$$l(\beta, y) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} = \prod_{i=1}^n \left[\frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}} \right]^{y_i} \left[\frac{1}{1 + e^{x_i'\beta}} \right]^{1-y_i} \dots \dots \dots (3.4)$$

For convenience in multiple logistic regression models, the likelihood equations can be written in matrix notation as:

$$\frac{\partial L(\beta)}{\partial(\beta)} = X'(Y - \pi) \dots \dots \dots (3.5)$$

Now, theoretically putting

$$\frac{\partial L(\beta)}{\partial(\beta)} = 0$$

produces, $\hat{Y} = \hat{\pi}$ satisfying $X'(Y - \hat{Y}) = 0$. 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 eq (3.4). 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 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 $H_1: \beta_i \neq 0$).

The Wald test is based on the behavior of the log-likelihood function at the ML estimate, 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:

$$W = \left[\frac{\hat{\beta}}{SE(\hat{\beta})} \right]^2 \dots\dots\dots(3.6)$$

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

3.3.1.4. The 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. Goodness-of-fit statistics 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 $g \times 2$ contingency table of observed and estimated expected frequencies. A formula defining the calculation of \hat{C} is as follows:

$$\hat{C} = \frac{\sum_{k=1}^g (O_k - E_k)^2}{V_k} \dots\dots\dots(3.7)$$

where $E_k = nP_k$, $V_k = nP_k(1-P_k)$, 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. If the observed number of events differs significantly from what is expected by the model, the statistic \hat{C} will be large and there will be evidence against the null hypothesis that the model is adequate to fit the data. 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 not significant, the model estimates are adequate to fit the data at an acceptable level.

3.3.1.5. The likelihood ratio test

The likelihood ratio 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:

$$G = [(-2\ln L_0) - (-2\ln L_1)]$$

where, L_0 is the likelihood of the null model and L_1 is the likelihood of the saturated model. The statistic G plays the same role in logistic regression as the numerator of the partial F-test does in linear regression.

Under the global null hypothesis, $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$, the statistic G follows a chi-square distribution with k degrees of freedom and measures how well the independent variables affect the response variable (Hosmer and Lemeshow, 2000).

3.3.1.6. The ROC Curve

One way of finding out how well a model can make predictions is by creating a receiver operating characteristic (ROC) curve. Let

$$\hat{y}_i = \begin{cases} 1 & \text{if } \hat{\pi}(x_i) > \pi_0 \\ 0 & \text{otherwise} \end{cases}$$

where \hat{y}_i is the predicted value of y_i (the outcome) and π_0 is a cutoff limit. (Agresti, 2007) makes the following definitions:

Sensitivity = $P(\hat{y} = 1|y = 1)$ and Specificity = $P(\hat{y} = 0|y = 0)$:

A receiving operating curve, ROC, is a curve that plots Sensitivity against Specificity for all possible values of π_0 to be more precise, the ROC curve plots Sensitivity against 1-Specificity. This creates a concave curve that connects the points (0, 0) and (1, 1) in the XY-plane. The area under the curve reflects the predictive ability the model in question has, where a larger area

reflects a better predictive ability. Obtaining the area 0.5 means predictions of the model were no better than random guessing (Agresti, 2007).

Hosmer and Lemeshow (2000 page.162) provide guidelines on how to interpret different areas under the ROC-curve (AUC):

$$\text{If} = \begin{cases} 0.5 < AUC < 0.7 & \text{Poor} \\ 0.7 < AUC < 0.8 & \text{Acceptable} \\ 0.8 < AUC < 0.9 & \text{Excellent} \\ 0.9 < AUC & \text{Outstanding} \end{cases}$$

3.3.1.7. Model Diagnostics

To identify an observation is outliers or influential, the following rules of thumbs are employed in this study.

- **Leverage Values (Hat Diag)** is a 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 leverage values larger than one are considered to be potentially highly influential (Belsley et al., 1980). The formula is:-

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}$$

where:- h_i = Leverage value, n = Number of observations and S_{xx} is Standard error

- **DFBETAS** measure how much an observation has affected the estimate of a regression coefficient (there is one DFBETA for each regression coefficient, including the intercept). If DFBETAS is less than unity, this implies no specific impact of an observation on the coefficient of a particular predictor variable, while DFBETA of a case greater than 1.0, and implies the observation is an outlier (Cook and Weisberg, 1982).

$$\text{DFBETAS}_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{SE_{\hat{\beta}_{j(i)}}}$$

- **Cook's D** measure the 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 and Lemeshow, 2000). It can be calculated by:-

$$D_i = \frac{r_i^2 \text{Var}(\hat{y}_i)}{P \text{var}(e_i)} = \frac{r_i^2 h_{ii}}{P(1-h_{ii})}$$

where:- D_i = Cook's distance, r_i^2 = Standardized residual, h_{ii} = Leverage and P = Number of predictors.

3.3.2. Multilevel Analysis

The 2013-ENLFS dataset used for this study is based on multistage stratified cluster sampling. The structure of data in the population is hierarchical, and a sample from such a population can be viewed as a multistage sample. For multistage clustered samples, the dependence among observations often comes from several levels of the hierarchy. In order to draw appropriate inferences and conclusions from multistage stratified clustered survey data we may require tricky and complicated modeling techniques like multilevel modeling.

Multilevel analysis is a methodology for the analysis of data with complex patterns of variability, with a focus on nested sources of variability. These are used when the data structure is hierarchical with elementary units at level 1 nested in clusters at level 2, which in turn may be nested in (super) clusters at level 3, and so on. The latent variables, or random effects, are interpreted as unobserved heterogeneity at the different levels which induce dependence among all lower-level units belonging to a higher-level unit. Random intercepts represent heterogeneity between clusters in the overall response and random coefficients represent heterogeneity in the relationship between the response and explanatory variables.

Logistic regression requires the assumptions of independence of the observations conditional on the explanatory variables and uncorrelated residual errors. These assumptions are not always met when analyzing nested data. But the multilevel logistic regression analysis considers variations due to hierarchy structure in the data. It allows the simultaneous examination of the effects of group level (cluster and division) and individual level variables on individual level outcomes

while accounting for the non-independence of observations within groups. The analysis also allows the examination of both between group and within group variability as well as how group level and individual level variables are related to variability at both levels.

3.3.2.1. Two-Level Model

In this study, the clustering of the data points within geographical regions offers a natural 2-level hierarchical structure of the data, i.e. children are nested within regions. Let y_{ij} be the binary outcome variable for individual i in region j , coded '0' or '1', associated with level-one unit i nested within level two unit j . Also let p_{ij} be the probability that the response variable for individual i in region j equals 1, and $p_{ij} = \Pr(y_{ij} = 1)$. Here, y_{ij} follows a Bernoulli distribution. Like the logistic regression the p_{ij} is modeled using the link function, logit. The two-level logistic regression model can be written as,

$$\log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_0 + \beta_1 x_{ij} + u_{0j} \quad (3.8)$$

where u_{0j} is the random effect at level 2.

Therefore, conditional on u_{0j} , the y_{ij} 's can be assumed to be independently distributed. Here, u_{0j} is a random quantity and follows $N(0, \sigma_u^2)$. The basic data structure of two-level logistic regression is a collection of N groups (units at level-two (regions)) and within group j ($j=1, 2, \dots, N$) a random sample of n_j level-one units. The outcome variable is dichotomous and denoted by y_{ij} ($i = 1, 2, \dots, n_j, j = 1, 2, \dots, N$) for level-one unit i in group j .

If we let the success probability in group j be denoted by p_j , the dichotomous outcome variable for the individual i in group j , y_{ij} ; which is either 0 or 1 can be expressed as the sum of the probability in group j , p_j (the average proportion of j levels in group j , $E(y_{ij}) = p_j$) and some individual-dependent residual ε_{ij} , that is, $y_{ij} = p_j + \varepsilon_{ij}$. The residual term is assumed to have mean zero but has a peculiar property that it can assume only the values $-p_j$ and $1-p_j$. The variance of the residual is, $Var(\varepsilon_{ij}) = p_j(1-p_j)$.

3.3.2.2. Testing heterogeneous Proportionality

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:

$$\chi^2 = \sum_{j=1}^N n_j \frac{(\hat{P}_j - \hat{P}.)^2}{\hat{P}.(1-\hat{P}.)} \text{-----} (3.9)$$

where $\hat{P}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$ is the proportion of successes in group j, $\hat{P} = \frac{1}{M} \sum_{j=1}^N \sum_{i=1}^{n_j} Y_{ij}$ is the overall proportion of successes, N is the number of groups, n_j is the number of samples in the jth group.

This statistic follows approximately chi-square distribution with N-1 degrees of freedom. The approximation is valid if the expected numbers of success and of failures in each group, $n_j \hat{P}_j$ and $n_j(1-\hat{P}_j)$, respectively, are at least 1 while 80 percent of them are at least 5 (Agresti, 2002).

Estimation of between and within group variance. The true variance between the group dependent probabilities, i.e., the population value of $\text{Var}(P_j)$, can be estimated by:

$$\tau^2 = S^2_{\text{between}} - \frac{S^2_{\text{within}}}{\tilde{n}} \text{-----} (3.10)$$

where \tilde{n} is given by:

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

For dichotomous outcome variables, the observed between-groups variance is closely related to the chi-squared test statistic (3.9). They are connected by the formula

$$S^2_{between} = \frac{\hat{P} \cdot (1 - \hat{P})}{\bar{n}(N - 1)} \chi^2 \text{-----} (3.11)$$

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

$$S^2_{within} = \frac{1}{M - N} \sum_{j=1}^N n_j \hat{P}_j (1 - \hat{P}_j) \text{-----} (3.12)$$

where \hat{P}_j is the proportion of successes in group j , \hat{P} is the overall proportion of successes, n_j is the sampled observation in group j , \bar{n} is the average of the sampled observation in groups, M is the total sampled observations, and N is the number of groups (in our case regions).

3.3.2.3. *The Empty Logistic Regression Model*

This is the simplest case of a hierarchical two level model for a dichotomous outcome variable in which there are no explanatory variables at all. This model only contains random groups and random variation within groups.

We focus on the model that specifies the transformed probabilities $f(p_j)$ to have a normal distribution. This is expressed, for a general link function $f(p)$, by the formula

$$f(p_j) = \beta_0 + u_{0j} \text{-----} (3.13)$$

where β_0 is the population average of the transformed probabilities and u_{0j} the random deviation from this average for group j . If $f(p)$ is the logit function, then $f(p_j)$ is just the log-odds for group j . Thus, for the logit link function, the log-odds have a normal distribution in the population of groups, which is expressed by

$$\text{logit}(p_j) = \beta_0 + u_{0j} \text{-----} (3.14)$$

For the deviations u_{0j} , it is assumed that they are independent random variables with a normal distribution with mean zero and variance σ_0^2 .

This model does not include a separate parameter for the level-one variance. This is because the level-one residual variance of the dichotomous outcome variable follows directly from the success probability, as indicated by the following equation.

$$Var(\varepsilon_{ij}) = p_j(1 - p_j).$$

The probability corresponding to the average value β_0 , denoted by π_0 is defined by

$$f(\pi_0) = \beta_0.$$

For the logit function, the so-called logistic transformation of β_0 , is defined by

$$\pi_0 = \text{logit}(\beta_0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \dots\dots\dots (3.15)$$

Due to the non-linear nature of the logit link function, there is no simple relation between the variance of the deviations u_{oj} . However, there is an approximate formula which is valid when the variances are small and is given by

$$Var(p_j) = (\pi_0(1 - \pi_0))^2 \sigma_0^2 \dots\dots\dots (3.16)$$

3.3.2.4. The Random Intercept Logistic Regression Model

Random intercept models are models where only the intercept of the level-1 dependent variable is modeled as an effect of the level-2 grouping variable and possibly other level-1 or level -2 (or higher). Random intercept regression models are also called “means as outcome regression models”. In this model the intercept vary between groups. This reflects that some groups tend to have, on average, higher responses and others tend to have lower. Random intercepts are used to model unobserved heterogeneity in the overall response. This allows knowing how the variability of the overall probability of the occurrence of child labour seems across regions. 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.

We assume that there are variables which are potential explanation for observed success or failure. These variables are denoted by $X_h = \{x_{hij}, i = 1, 2, \dots, n_j, h = 1, 2, \dots, k, j = 1, 2, \dots, N, n_j$ is the number of observation in region j , k is the number of predictor variables, N is the number of groups (regions)}. Since some (or all) of these variables could be level-1 variables, the success probability is not necessarily the same for all individuals in a given group. Therefore, the success probability p_{ij} depends on the individual as well as the group.

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 variable and a random group-dependent deviation u_{0j} . That is,

$$\text{logit}(p_{ij}) = \log \left[\frac{p_{ij}}{1 - p_{ij}} \right] = \beta_{0j} + \sum_{h=1}^k \beta_h x_{hij} = \beta_0 + \sum_{h=1}^k \beta_h x_{hij} + u_{0j} \dots\dots\dots (3.17)$$

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 $\beta_{0j} = \beta_0 + u_{0j}$. The first part of the right hand side of (3.17) 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_{0j} is called the random part of the model. It is assumed that the residual, u_{0j} are mutually independent and normally distributed with mean zero and variance σ_0^2 (Snijders and Bosker, 1999). From (3.17) we have

$$\text{logit}(p_{ij}) = \log \left[\frac{p_{ij}}{1 - p_{ij}} \right] = \beta_{0j} + \sum_{h=1}^k \beta_h x_{hij} \dots\dots\dots (3.18)$$

Solving for p_{ij} we have

$$p_{ij} = \frac{e^{\beta_0 + \sum_{h=1}^k \beta_h x_{hij} + u_{0j}}}{1 + e^{\beta_0 + \sum_{h=1}^k \beta_h x_{hij} + u_{0j}}} \dots\dots\dots (3.19)$$

Thus, a unit difference between the x_h values of two individuals in the same group is associated with a difference of β_h in their log-odds, or equivalently, a ratio of $\exp(\beta_h)$ in their odds. Level-2

residual, u_{0j} is the effect of being in group j on the log-odds that $y = 1$ with level-2 (residual) variance σ_0^2 , or the between-group variance in the log-odds that $y = 1$ after accounting for the predictor(s).

Random intercept models have many applications, for instance estimating the regional effects on child labor controlling for individual (child's) level factors, and within the model evaluate and compare the situation of the regions child labour status. This can be done by obtaining the random intercept for each region. This regional effect is a measure of the situation of child labour due to the region relative to the average of all regions. If the odds of child labour for regional effects is sufficiently larger than one, the region is considered to have performed worse than the average; if it is significantly smaller than one, the region is considered to have better performance than the expected.

3.3.2.5. The Random Coefficient Logistic Regression Model

Random coefficients models are ones where the coefficient(s) of lower-level predictor(s) is/are modeled as well. Random coefficients model explain unobserved heterogeneity in the effects of explanatory variables on the response variable. In logistic regression analysis, linear models are constructed for the log-odds. The multilevel analogue, random coefficient logistic regression, is based on linear models for the log-odds that include random effects for the groups or other higher level units.

Consider explanatory variables which are potential explanations for the observed outcomes. Denote these variables by $X_h = \{x_{hij}, i = 1, 2, \dots, n_j, h = 1, 2, \dots, k, j = 1, 2, \dots, N\}$. Since some or all of these variables could be level-one variable, the success probability is not necessarily the same for all individuals in a given group. Therefore, the success now consider a model with group-specific regressions of logit of the success probability, $\text{logit}(p_{ij})$ on a single level one explanatory variable x_1 is

$$\text{logit}(p_{ij}) = \log \left[\frac{p_{ij}}{1 - p_{ij}} \right] = \beta_{0j} + \beta_{1j}x_{1ij} \dots\dots\dots(3.20)$$

The intercepts β_{0j} as well as the regression coefficients, or slopes, β_{1j} are group-dependent. These group-dependent coefficients can be split into an average coefficient and the group-dependent deviation:

$$\beta_{0j} = \beta_0 + u_{0j}, \quad \text{and} \quad \beta_{1j} = \beta_1 + u_{1j} \dots \dots \dots (3.21)$$

Substituting (3.21) in (3.20) we have

$$\logit(p_{ij}) = \log \left[\frac{p_{ij}}{1 - p_{ij}} \right] = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x_{1ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j} x_{1ij} \dots \dots \dots (3.22)$$

From (3.21) there are two random group effects, the random intercept u_{0j} and the random slope u_{1j} . It is assumed that the level-two residuals u_{0j} and u_{1j} have means zero given the value of the explanatory variable X. Thus β_1 is the average regression coefficient like β_0 is the average intercept. The part, $\beta_0 + \beta_1 x_{1ij}$ of equation (3.22) is called the fixed part of the model and the other part, $u_{0j} + u_{1j} x_{1ij}$ is called the random part. The term $u_{1j} x_{1ij}$ can be regarded as a random interaction between group and explanatory variable X. This model implies that the groups are characterized by two random effects: their intercept and their slope. These two group effects, u_{0j} and u_{1j} will not be independent, but correlated. The random intercept variance $\text{var}(u_{0j}) = \sigma_0^2$, the random slope variance $\text{var}(u_{1j}) = \sigma_1^2$, and the covariance between the two random effects $\text{cov}(u_{0j}, u_{1j}) = \sigma_{01}$ are called variance components (Snijders and Bosker, 1999).

The model for a single explanatory variable discussed above can be extended by including more variables that have random effects. Suppose that there are k level-one explanatory variables x_1, x_2, \dots, x_k , and considering the model where all x-variables have varying slopes and random intercept. That is,

$$\logit(p_{ij}) = \log \left[\frac{p_{ij}}{1 - p_{ij}} \right] = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + \dots + \beta_{kj} x_{kij} \dots \dots \dots (3.23)$$

Letting

$$\beta_{0j} = \beta_0 + u_{0j} \text{ and}$$

$$\beta_{hj} = \beta_h + u_{hj}, \text{ for } h = 1, 2, \dots, k$$

$$\logit(p_{ij}) = \log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_0 + \sum_{h=1}^k \beta_h x_{hij} + u_{0j} + \sum_{h=1}^k u_{hj} x_{hij} \dots\dots\dots(3.24)$$

The first part of this model, $\beta_0 + \sum_{h=1}^k \beta_h x_{hij}$ is the fixed part and the second part $u_{0j} + \sum_{h=1}^k u_{hj} x_{hij}$ is the random part of the model (Snijders and Boskers, 1999).

3.3.2.6. Intra-class Correlation Coefficient (ICC)

The fundamental reason of using multilevel analysis is the existence of more similarities between children in the same region than in different regions. This leads to the existence of intra-class (intra-Regional) correlation of children’s to be engaged in child labour. ICC is the degree of resemblance between level one units belonging to the same group. It is an indication of the proportion of variance at the second level (region) and it can also be interpreted as the expected (population) correlation between two randomly chosen individuals within the same group (Hox, 2010).

In two- level model, the ICC is calculated in the intercept only model. This model can be derived from “Eq. (3.24)” by excluding all explanatory variables, which results in the following equation: $\logit(p_j) = \beta_0 + u_{0j}$.

The ICC is then calculated as follows:

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} \dots\dots\dots(3.25)$$

where, σ_e^2 is variance of individual (lower) level units.

Since the logistic distribution for the level one residual variance implies a variance of $\frac{\pi^2}{3} \approx 3.29$

(Snijders and Bosker, 1999) and this formula can be reformulated as:

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + 3.29} \dots\dots\dots (3.26)$$

3.3.2.7. Estimation and Testing Technique for Multilevel logistic model

Parameter estimation for multilevel logistic model is not straightforward like the methods for ordinary single level logistic regression. The most common methods for estimating multilevel logistic models are based on likelihood. Among the methods, Marginal Quasi Likelihood or MQML [Goldstein (1991), Goldstein and Rasbash (1996)] and Penalized Quasi Likelihood or PQL [Laird (1978) and Breslow and Clayton (1993)] are the two prevailing approximation procedures. Both MQML and PQL are based on Taylor series expansion to achieve the approximation. Based on the first and second term of Taylor expansion, MQML and PQL are often known as first order MQML and second-order MQML, first-order PQL and second-order PQL respectively. After applying these quasi likelihood methods, the model is then estimated using iterative generalized least squares (IGLS) or reweighted IGLS (RIGLS) [Goldstein (2003)]. Besides, there are other estimation methods: Maximum Likelihood Method (several simulation based; McCulloch (1997)), Bayesian methods using Markov Chain Monte Carlo (MCMC), adaptive Gaussian quadrature (AGQ) and the Iterative Bootstrap method. Using MCMC simulation technique has come to the forefront of statistical research over the last one and half decade [Gelfand et al. (1990)] and also it is being used with greater extent in multilevel modeling recently. An important part of modeling involves testing parameters and models to see which parts of the multilevel model are statistically important. For fixed coefficients of multilevel logistic regression tests about parameters are done using the Wald test. The random part of multilevel logistic regression parameters is estimated based on t-test or Z-test. Parameter estimation in hierarchical generalized linear models is more complicated than the hierarchical linear models. The most frequently used kind of approximation method used is based on a first-order or second-order Taylor series expansion of the link function. There are different methods of parameter estimations which are implemented by various software packages such as, MLwiN,

STATA and SAS. In this study, the multilevel data have been analyzed with the help of STATA software package (Snijders and Bosker, 1999).

3.3.2.8. Multilevel model selection criteria: AIC and BIC

Akaike's (1974) information criterion is a measure of fit defined as

$$AIC = -2\ln L + 2k$$

where $\ln L$ is the maximized log-likelihood of the model and k is the number of parameters estimated. Schwarz's (1978) Bayesian information criterion is another measure of fit defined as

$$BIC = -2 \ln L + k \ln N$$

where N is the sample size.

AIC and BIC can be viewed as measures that combine fit and complexity. Fit is measured negatively by $-2 \times \ln(\text{likelihood})$; the larger the value, the worse the fit. Complexity is measured positively, either by $2 \times k$ (AIC) or $\ln(N) \times k$ (BIC). Given two models fitted on the same data, the one with the smaller AIC or BIC fits the data better than the one with the larger AIC or BIC (Akaike, 1974 and Schwarz, 1978).

CHAPTER FOUR

4. STATISTICAL DATA ANALYSIS AND RESULTS

This chapter presents the results of the study. The first section presents descriptive statistics of the variables included in the study. The second section presents the single level multiple logistic regression and multilevel logistic regression results, and the last part is discussion of the results.

4.1. Descriptive Analysis

In order to have an overall picture of the distribution of child labour in Ethiopia, we start with a descriptive analysis of the variables included in the study.

The total number of children aged 5 to 14 years covered in the present study was 12,238. Among these, 3,922 children (about 32%) were engaged in child labor during the last 7 days before the survey (Table 4.1).

This is a sign of how severe the situation of child labor in Ethiopia is. In this section the description of the sample data used in the study are presented. This study has focused on the effects of individual, household and community level characteristics on child labour. The target group consists of boys and girls aged 5-14 years. Out of the total number of sampled children (12,238), 6,094 (49.8%) were girls.

Table 4.1: Distribution of child-labor status by Socio-Demographic factors

Variables	Child Labour	Total no. of children	Percentage of Child Labor	d.f	Chi-square	p-value
Region						
Tigray (Ref)	323	935	34.5	10	658.428	<.0001
Afar	144	457	31.5			
Amhara	912	2114	43.1			
Oromya	1145	2943	38.9			
Somali	146	743	19.7			
Ben-Gumuz	139	518	26.8			
SNNPR	790	2412	32.8			
Gambela	94	550	17.1			
Harari	83	350	23.7			
Addis Ababa	15	823	1.8			
Dire Dawa	131	393	33.3			
Total	3922	12238	32.0			
Sex of the child						
Male	2114	6144	34.4	1	31.552	<.0001
Female (Ref)	1808	6094	29.7			
Age categories of children						
5 – 11 years (Ref)	2663	8585	31.0	1	13.970	<.0001
12 – 14 years	1259	3653	34.5			
Place of residence of the child						
Urban	391	5080	7.7	1	2339.05	<.0001
Rural (Ref)	3531	7182	49.2			
School attendance of the child						
Attending	2107	7756	27.2	1	237.700	<.0001
Not attending(Ref)	1815	4482	40.5			
Child's relationship with the household head						
Son/daughter	3422	10463	32.7	1	14.343	<.0001
Not son/daughter (Ref)	500	1775	28.2			
Sex of the household head						
Male	3231	9423	34.3	1	94.449	<.0001
Female (Ref)	691	2820	24.5			
Literacy status of the household head						
Literate	1521	6287	24.2	1	366.3	<.0001
Illiterate (Ref)	2401	5951	40.3			
Employment status of the household head						
Employed	3737	10937	34.2	1	212.500	<.0001
Unemployed (Ref)	185	1301	14.2			
Current marital status of the household head						
Married	3335	10017	33.3	1	39.329	<.0001
Not married (Ref)	587	2221	26.4			
Family size						
5 and below (Ref)	1464	5185	28.2	2	90.695	<.0001
6-10	2395	6745	35.5			
11 and above	63	308	20.5			

Source: Own calculation based on the 2013-ENLFS report

The proportion of children engaged in child labour varied from one region to the other in Ethiopia. Table 4.1 shows that region of residence is significantly associated with incidence of child labour ($p < 0.001$). The highest (43.1%) percentage of child labour was observed in Amhara region followed by Oromiya region (38.9%). On the other hand, Addis Ababa city administration and Gambela region had the lowest percentage (1.8%, 17.1%) respectively, for incidence of child labour among children in the age group of 5-14 years. Hence, there seems to be some variation in the proportion of child labourers amongst children in different regions.

The proportion of child labourers also differs by sex. While higher proportion (34.4%) of boys aged 5-14 years were engaged in child labour, a relatively lower proportion (29.7%) of girls aged 5-14 were engaged in child labour. Similarly, the proportion of child labourers differs with their age groups. About 31.0% of the children in the age group of 5-11 years were child labourers while 34.5% of children in the age group of 12-14 years were child labourers.

Likewise, the proportion of child labour, observed indicates high variation in the proportion of children engaged in child labour by place of residence (urban and rural). Accordingly, higher proportion (49.2%) of children residing in rural areas and a relatively small proportion (7.8%) of the children in urban areas were engaged in child labour.

A high percentage of child labour was observed among children who have not been attending school at the time of the survey (40.5%) as opposed to the lower percentage of child labour which was recorded for children who have been attending school (27.2%). Likewise, the proportion of children engaged in child labor varied with the relationship of children to the head of the household. The higher percentage of child labour was observed for children who are the son/daughter of the household head (32.7%) as opposed to the lower percentage of child labour which was recorded for children who are not the son/daughter of the household head (28.2%). With regard to gender of the household head, higher percentage of child labourers (34.3%) resided in male headed households and relatively small percentage of child labourers (24.5%) resided in female headed households.

The incidence of child labour also differs according to literacy and current marital status of the household head. The higher percentage (40.3%) of child labour was observed among those children having illiterate household heads as opposed to the lower percentage (24.2%) of child

labour which was recorded among children having literate household heads. When we see child labour by marital status of the household head, higher percentage (33.3%) of child labourers was observed among married household heads and a relatively lower percentage (26.4%) of child labourers was observed among non-married household heads.

Employment status of household head and the household size affected children engaged in child labour at different magnitudes. While 34.2% of the children having employed household heads were engaged in child labour, 14.2% of the children from non-employed household heads were engaged in child labour. With regard to the household size, the highest percentage (35.5%) of child labour was observed in households of size six to ten and the lowest percentage (20.5%) of child labour was observed in households of size eleven or more.

4.2. Bivariate Statistical Analysis

The bivariate statistical analysis addresses the marginal effect of a predictor variable on the response variable without taking into account other predictors. The analysis was based on a cross tabulation of the response variable, child labour and the predictor variables independently. The bivariate chi-square test results in Table 4.1 indicated that all 11 variables are statistically significant at 5% ($p < 0.05$) significance level. The chi-square test results show that the incidence of child labour is associated with region, sex of the child, age categories of the child, place of residence of the child, school attendance of the child, relationship of children to the household head, sex of household head, literacy status of the household head, employment status of the household head and family size.

4.3. Determinants of Child Labour: Results of Logistic Regression Analysis

Multiple logistic regression was used to analyze the effect of the independent variables on child labour while controlling for the other independent variables. Enter method of variable selection procedure in SPSS was employed to select the important determinants of child labour among children. The statistical significance of the individual regression coefficients have been tested using the Wald chi-square statistic. Accordingly, region, sex of a child, age of a child, place of residence, relationship of a child with the household head, literacy and employment status of the household head and household size were found to be significant predictors for child labour. Conversely, school attendance of the child, sex of the household head and marital status of the household head were not significant (see Table 4.2).

The multiple binary logistic regression model fit results in Table 4.2 revealed that children aged 5-14 years in Amhara region were 46.7% more likely to be engaged in child labour than children in Tigray region. Similarly, children in Oromiya were 26.5% more likely to be engaged in child labour than children in Tigray region (aOR=1.265; 95% CI: 1.067, 1.499). However, children in Somali region were 48.2% less likely to be engaged in child labour compared to children in Tigray region. Similarly, children in Beneshangul-Gumuz were 35.7% less likely to be engaged in child labour compared to children in Tigray region (OR=0.643; 95% CI: 0.498, 0.831) and children in Addis Ababa were 79.2% less likely to be engaged in child labour when compared with children in Tigray region. On the other hand, there was no statistically significant difference in child labour between children living in Tigray region and children living in Afar, SNNP and Dire Dawa.

The results also revealed that boys were 21.6% more likely to be engaged in child labour when compared with girls (aOR=1.216; 95% CI: 1.116, 1.326). Likewise, older children in the age group 12-14 were 38.9% more likely to be engaged in child labour compared to younger children in the age group 5-11 controlling for other variables in the model.

Place of residence and relationship of children to the household head were found to be significantly associated with child labour. Children who resided in the urban areas were 89.5% less likely to be engaged in child labour compared with those from the rural areas (aOR=0.105; 95% CI: 0.092, 0.120). Children of the household head were 35.0% less likely to be engaged in child labour compared with those who are not children of the household head.

Literacy status of the household head is associated with child labour. Children from literate household heads were 16.3% less likely to be engaged in child labour than children having illiterate household heads. Conversely, the employment status of the household head has positive effect to child labour. Children from households having employed household heads were 85.9% more likely to be engaged in child labour than children having non employed household heads.

There was no significant difference in child labour between children from households of size six to ten family and children from households of size five or less. But, children from households of size eleven or more were 33.1% less likely to be engaged in child labour than children from households of size five or less.

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

Variables	Estimate	S.E.	Wald	D.f	Sig.	aOR	95.0% C.I. for OR	
							Lower	Upper
Region			245.552	10	.000*			
Tigray(Ref)								
Afar	-.083	.135	.379	1	.538	.920	.706	1.199
Amhara	.383	.091	17.889	1	.000*	1.467	1.228	1.752
Oromiya	.235	.087	7.311	1	.007*	1.265	1.067	1.499
Somali	-.657	.126	27.230	1	.000*	.518	.405	.663
Beneshangul-Gumuz	-.441	.131	11.377	1	.001*	.643	.498	.831
SNNP	-.105	.089	1.373	1	.241	.901	.756	1.073
Gambela	-.802	.143	31.256	1	.000*	.449	.339	.594
Harari	-.366	.157	5.400	1	.020*	.694	.510	.944
Addis Ababa	-1.568	.277	32.120	1	.000*	.208	.121	.359
Dire Dawa	.276	.145	3.644	1	.056	1.318	.993	1.749
Sex of the child			19.744	1	.000*			
Male	.196	.044	19.744	1	.000*	1.216	1.116	1.326
Female(Ref)								
Age of the child			43.735	1	.000*			
5-11 years(Ref)								
12-14 years	.328	.050	43.735	1	.000*	1.389	1.260	1.530
Place of residence			1.177E3	1	.000*			
Urban	-2.253	.066	1.177E3	1	.000*	.105	.092	.120
Rural(Ref)								
School attendance			.005	1	.945			
Attending	-.003	.047	.005	1	.945	.997	.909	1.093
Not attending(Ref)								
Child's relationship to the household head			34.639	1	.000*			
Son/daughter	-.430	.073	34.639	1	.000*	.650	.564	.751
Not son/daughter(Ref)								
Sex of the household head			1.766	1	.184			
Male	.107	.081	1.766	1	.184	1.113	.950	1.304
Female(Ref)								
Literacy status of the household head			13.429	1	.000*			
Literate	-.178	.049	13.429	1	.000*	.837	.760	.920

Illiterate(Ref)								
Employment status of the household head	40.460 1 .000*							
Employed	.620	.098	40.460	1	.000*	1.859	1.536	2.251
Not Employed(Ref)								
Marital status of the household head	1.355 1 .244							
Married	-.102	.087	1.355	1	.244	.903	.761	1.072
Not married(Ref)								
Household size	9.189 2 .010*							
5 and below(Ref)								
6-10	.059	.049	1.430	1	.232	1.060	.963	1.167
11 and above	-.402	.162	6.125	1	.013*	.669	.487	.920
Constant	-.412	.133	9.574	1	.002*	.662		

* Significant (p<0.05)

The negative sign in the column labeled ‘Estimate’ indicates an inverse relationship of an explanatory variable with the log odds of the dependent variable. In contrast the positive coefficient columns labeled ‘Estimate’ indicate a direct relationship to the log odds of the dependent variable.

The statistical significance of individual regression coefficients is tested using the Wald chi-square statistic. Region, sex of the child, age category of the child, place of residence, school attendance of the child, relationship of the child with the household head, literacy status of the household head, employment status of the household head and household size were found to be significant predictors of child labour at 5% level of significance. The test of significance of the intercept also indicated that it is significant implying that it should be included in the model. Thus, the estimated model is given by:

$$\log it(\hat{p}) = \beta_0 + \sum_{i=1}^{11} \beta_{1i} X_{1i} + \sum_{j=1}^1 \beta_{2j} X_{2j} + \sum_{k=0}^1 \beta_{3k} X_{3k} + \sum_{l=0}^1 \beta_{4l} X_{4l} + \sum_{m=0}^1 \beta_{5m} X_{5m} + \sum_{n=0}^1 \beta_{6n} X_{6n} \\ + \sum_{o=0}^1 \beta_{7o} X_{7o} + \sum_{p=0}^1 \beta_{8p} X_{8p} + \sum_{q=0}^1 \beta_{9q} X_{9q} + \sum_{r=0}^1 \beta_{10r} X_{10r} + \sum_{s=1}^3 \beta_{11s} X_{11s}$$

where:

\hat{p} = predicted probability of child labour, β_0 = constant, X_{1i} = children's region of level i, X_{2j} =sex of children at level j, X_{3k} =age of children at level k, X_{4l} =place of residence of children at level l, X_{5m} =school attendance of children at level m, X_{6n} = relationship of children with the household head at level n, X_{7o} =sex of the household head at level o, X_{8p} = literacy status of the household head at level p, X_{9q} = employment status of the household head at level q, X_{10r} = marital status of the household head at level r and X_{11s} = household size at level s.

For all explanatory variables, the first category was taken as the reference category. The fitted logistic regression model with only significant variables is:

$$\begin{aligned} \logit(\hat{p}) = & -0.412-0.083X_{12}+0.383X_{13}+0.235X_{14}-0.657X_{15}-0.441X_{16}-0.105X_{17}-0.802X_{18}- \\ & 0.366X_{19}-1.568X_{110}+0.276X_{111}+0.196X_{21}+0.328X_{31}-2.253X_{41}-0.003X_{51}-0.430X_{61}+ \\ & 0.107X_{71}-0.178X_{81}+ 0.620X_{91} -0.102X_{101} + 0.059X_{112} - 0.402X_{113} \end{aligned}$$

4.4. Goodness of Fit and Model Diagnostics for logistic regression

4.4.1. Goodness of fit of the logistic regression model

After a logistic regression model has been fitted, it is necessary to see the appropriateness, adequacy and usefulness of the fitted model. Thus, a global test of the goodness of fit of the resulting model should be performed. The most commonly used techniques are Pearson's Chi-square, Hosmer-Lemeshow test and the Wald goodness of fit tests.

4.4.1.1. Likelihood Ratio Test

The most common assessment of overall model fit in logistic regression is the likelihood ratio test, which is simply the chi-square difference between the null model (constant only) and the model containing a set of predictors. Under model summary in the Appendix A we see that the -2 Log Likelihood statistics is 12252.814. This statistic shows how much improvement is needed before the predictors provide the best possible prediction of the dependent variable, the smaller the value of the statistic the better the model. The statistic for the model that had only the intercept is $-2LL0 = 3099.234 + 12252.814 = 15352.048$. The inclusion of the parameters reduced the -2 Log Likelihood statistic by $15352.048 - 12252.814 = 3099.234$, which is reflected in the model chi-square for the omnibus test. The omnibus test of model coefficients tests whether the

model with the predictors is significantly different from the model with only the intercept. The omnibus test may be interpreted as a test of the capability of all predictors in the model jointly to predict the response (dependent) variable.

The omnibus test of model coefficients is used to assess the overall fit of the logistic regression model. The results ($\chi^2=3099.234$, d.f=21, p-value<0.001), show that the fit is adequate, implying that at least one of the predictors is significantly related to the response variable. That is, the null hypothesis that there is no difference between the model with only a constant and the model with independent variables was rejected (see Appendix A).

4.4.1.2. The Hosmer-Lemeshow Goodness-of-Fit test

The Hosmer-Lemeshow goodness of fit test divides subjects into g classes(often deciles) based on predicted probabilities and then computes a chi-square from observed and expected frequencies (usually in a 10x2 contingency table). A non-significant chi-square implies that there is no significant difference between the observed and the model predicted values and hence the estimated model adequately fit the data. Since the p-value is 0.264 (greater than 0.05), we do not reject the null hypothesis that there is no difference between observed and model-predicted values, implying that the estimated model fits the data at an acceptable level (see Table 4.3).

Table 4.3: Hosmer-Lemeshow Test

Hosmer-Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
10.015	8	0.264

4.4.1.3. The Roc Curve

Another way of testing the goodness of fit of the model is by using ROC curve. The area under the ROC curve, which ranges from zero to one, provides a measure of the ability of the model to discriminate between children engaged in child labour and children not engaged in child labour. The area under the ROC curve, 0.784 indicates acceptable discrimination according to the Hosmer-Lemeshow’s rule. The discrimination is statistically significant with p-value=0.000 (Table 4.4). Consequently, the fitted model acceptably discriminates children engaged in child labour from children not engaged in child labour. In addition to the output for the areas under

the ROC curve presented in table 4.4, the graph of the ROC Curve is presented in the appendices.

Table 4.4: Area under ROC Curve

Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.784	0.004	0.000	0.776	0.793

4.4.2. Model diagnostics

So far, we have discussed some summary statistics and examined the goodness of fit of our model. Before concluding that the model is adequate, it is a common practice to examine other measures used to detect the presence of outliers and influential observations among others. An observation is influential if it is individually or together with several other observations, has a demonstrably larger impact on the calculated values of various estimates than is the case for most of the other observations.

The results reveal that each one of model diagnostic measures are less than unity. DFBETAs less than unity imply no specific impact of an observation on the coefficient of a particular predictor variable, while Cook's distance less than unity showed that an observation had no overall impact on the estimated vector of regression coefficients β . A value of the leverage statistic less than one show that no subject has a substantial impact on the predicted values of the model. Based on the above goodness of fit tests and diagnostic checking results, we can say that our model is adequate.

4.5. Multilevel logistic regression analyses

A two-level structure (with children as first level unit and region as second level unit) has been used. Children selected randomly from the same region are assumed to be more similar than children selected randomly from different regions. In this section, three multilevel logistic regression models (empty model, random intercept with fixed effects model and random coefficient with random intercept model) have been considered and model comparison, testing goodness of fit and interpretation of fixed effects in terms of odds ratio are presented.

For the proper application of multilevel analysis in general and multilevel logistic analysis in particular, the first logical step is to test for heterogeneity of proportions between groups (in our

case between Regions). The chi-square test was applied to assess heterogeneity between regions means. The test results are $\chi^2 = 658.428$ with d.f = 10 (p=0.001). Thus, there is evidence of heterogeneity with respect to the incidence of child labour among the regions of Ethiopia.

4.5.1. Empty model with random intercept/intercept only model

The empty two-level model also called the null two-level model for a dichotomous outcome variable refers to a population of groups (level-two unit, i.e. regions) and specifies the probability distribution for group-dependent probabilities, π_j . It is the model that incorporates only the grand mean and random intercept (regional effect) without covariate.

It is given by: $\text{logit}(\pi_j) = \beta_0 + u_{0j}$, where $u_{0j} \sim \text{IID}(0, \sigma_0^2)$.

The intercept β_0 also known as the grand mean is shared by all regions while the random effect u_{0j} , also known as level two residual is specific to region j. It shows how the mean in a particular region deviates from the grand mean. σ_0^2 is the between regions variance. The random effect is not directly estimated but is summarized in terms of their estimated variances.

Presented in Table 4.5 are the estimates of fixed effects and random effects. The estimate of the fixed part of the model is -1.142 with a p- value of 0.000 implying that the average log odds of child labour is significantly different from zero. The fixed part of the model is interpreted as the grand mean of log odds of child labour with odds $\exp(-1.142) = 0.319$. The average probability of the occurrence of child labour is $\frac{\exp(-1.142)}{1+\exp(-1.142)} = 0.242$ which means that the chance for the incidence of child labour is 0.242 on average. The table also contains the variance estimate of the random effects at regional level, $\sigma_0^2 = 0.8694$ which implies that the between-region variance of child labour is 0.8694. At the bottom of the table, the value of the test statistics and the corresponding p-value for testing the hypothesis $H_0: \sigma_0^2 = 0$ that there is no cross-regional variation in child labour are presented. Since the value of the test statistic is 760.32 with p= 0.0000, the null hypothesis is rejected, and we conclude that there is strong evidence of heterogeneity or cross-regional variation in child labour incidence.

Table 4.5: Random intercept only model fit results

Childlabor	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	-1.142049	.2831676	-4.03	0.000	-1.697047	-.5870503
Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]		
REGION: Identity						
	var(_cons)	.8693587	.390514	.3604467	2.096799	
LR test vs. logistic regression: chibar2(01) = 760.32 Prob>=chibar2 = 0.0000						
Model	Obs	LL value	df	AIC	BIC	
M0	12238	-7295.865	2	14595.73	14610.55	

We can now write the model for the j^{th} region as $\text{logit}(\pi_j) = -1.142 + u_{0j}$. From the model we can say that the average probability of children to be engaged in child labour in the absence of covariates in region j is less than the average when u_{0j} is negative while it is higher than the average when u_{0j} is positive.

Estimated values of u_{0j} are presented in Table 4.6. The results indicate that the probability of child labour is less than the average in Somali, Gambela, Harari and Addis Ababa while it is higher than the average in the remaining regions. The worst situation has been in Amhara followed by Oromiya.

Table 4.6: Estimate of random effect for each region

Region	u_{0j}
Tigray	.5002457
Afar	.3614399
Amhara	.8640296
Oromiya	.6896441
Somali	-.2636785
Beneshangul-Gumuz	.1374313
SNNP	.421753
Gambela	-.4308043
Harari	-.0258892
Addis Ababa	-2.655339
Dire Dawa	.4430601

Using the figures in Table 4.5, we can calculate the intra-class correlation coefficient, which is a measure of the correlation between two individuals who are in the same higher level unit (region). A low ICC indicates a relatively small between region variations. In other words regions tend to perform at comparable levels to reduce the incidence of child labour. As ICC increases, regions tend to perform with ever increasing variations to reduce the incidence of child labour. We have a between regions variance of 0.8695 and a level one variance of $\pi^2/3 = 3.29$. From Equation (3.26), the intra-class correlation coefficient is 0.209. The ICC for this model implies that 20.9% of the variation in the incidence of child labour can be explained by grouping the children in regions. The remaining (100-20.9%=79.1%) of the variation in the incidence of child labour is explained within region-lower level units.

4.5.2. Random Intercept and Fixed Effect Model

In a random intercept and fixed coefficient multilevel logistic regression model, we allow the probability of child labour to vary across regions assuming that the effects of the explanatory variables are the same for each region. That is, the random intercept varies across regions, but children level explanatory variables are fixed across regions.

Table 4.7: Parameter Estimates of the Random Intercept and Fixed Slope Model

Variables	B	S.E.	Z	P> z	aOR	95.0% C.I. for OR	
						Lower	Upper
Sex of the child							
Male	.1962	.0441	4.45	.000*	1.2167	1.1160	1.3265
Female(Ref)							
Age of the child							
5-11 years(Ref)							
12-14 years	.3285	.0496	6.62	.000*	1.3888	1.2601	1.5307
Place of residence							
Urban	-2.2653	.0678	-34.4	.000*	.1038	.0912	.1181
Rural(Ref)							
School attendance							
Attending	-.0041	.0471	-0.09	.930	.9959	.9081	1.0921
Not attending(Ref)							
Child's relationship to the household head							
Son/daughter	-.4304	.0731	-5.89	.000*	.6502	.5635	.7504
Not son/daughter(Ref)							

Sex of the household head							
Male	.1086	.0808	1.35	.179	1.1147	.9516	1.3059
Female(Ref)							
Literacy status of the household head							
Literate	-.1794	.0487	-3.68	.000*	.8358	.7597	.9195
Illeterate(Ref)							
Employment status of the household head							
Employed	.6234	.0975	6.40	.000*	1.8653	1.5410	2.2579
Not employed(Ref)							
Marital status of the household head							
Married	-.1021	.0873	-1.17	.242	.9029	.7609	1.0714
Not married(Ref)							
Household size							
5 and below(Ref)							
6-10	.0586	.0490	1.20	.231	1.0604	.9634	1.1672
11 and above	-.4059	.1624	-2.50	.012*	.6664	.4847	.9160
Constant	-.6754	.1901	-3.55	.000*	.5090	.3507	.7387
Random Effect		Estimate	S.E	Z-value	P-value		
Var(U _{0j})= σ_0^2		.2484	.118	2.1	0.018		
LR test vs. logistic regression: chi2(1) = 225.75 Prob>=chi2 = 0.000							

* Significant (p<0.05)

Model	Obs	LL value	df	AIC	BIC
M1	12238	-6151.97	13	12329.94	12426.3

Table 4.8: Results of Model Selection Criteria and Log Likelihood Ratio Test

	Random Intercept Only Model	Random Intercept with Fixed Effects Model
Log likelihood(LL)	-7295.865	-6151.97
-2LL=Deviance	14591.73	12303.94
AIC value	14595.73	12329.94
BIC value	14610.55	12426.3
Wald test		Wald chi2(11)= 1573.34 Prob > chi2 = 0.000

Table 4.7 provides estimates of the fixed slopes and associated odds ratios. Values of the Wald test statistic used for testing the significance of individual predictors are given in the 4th column

with the corresponding p-values in the 5th column. The overall goodness of fit test results, Wald chi-square (11) = 1573.34 with $p = 0.000$ indicate that all explanatory variables jointly are significant. We see that the inclusion of level one covariates decreased regional variations from 0.8695 (level-two variance without covariates) to 0.2484, it indicates that there is a significant variation between regions in the incidence of child labour.

The deviance of the random intercept model, 14591.73 is reduced to 12303.94 when we included covariates for the same random intercept with fixed slope which implies that the random intercept with fixed slope model is better than the random intercept only model. The BIC and AIC values in Table 4.8 also ensure this as the smaller the values of AIC and BIC the better the model is.

Moreover, the values of $\chi^2(1) = 225.75$ and $p = 0.0000$ (see Table 4.7) lead to the rejection of the null hypothesis that the random effect is zero as in the assumption of ordinary logistic regression. From this we can conclude that the random effect at regional level is significantly different from zero.

4.5.3. The Random Coefficient Model

The variance components model which we have just specified and estimated in the preceding section 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 the explanatory variables are random at level two. All variables included in the random intercept model are included in the random coefficient model. Estimates of this model show that the random slope variances of all included variables except that of “Place of Residence” 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 the random coefficient model are presented in Table 4.9.

Table 4.9: Parameter Estimates of the Random Intercept and Random Slope Model

Variables	B	S.E.	Z	P> z	aOR	95.0% C.I. for OR	
						Lower	Upper
Sex of the child							
Male	0.1988	0.0442	4.5	0.000*	1.220	1.1188	1.3303
Female(Ref)							
Age of the child							
5-11 years(Ref)							
12-14 years	0.3239	0.0497	6.51	0.000*	1.3824	1.2541	1.5239
Place of residence							
Urban	-2.5923	0.2838	-9.13	0.000*	0.0748	0.0429	0.1305
Rural(Ref)							
School attendance							
Attending	0.0048	0.0471	0.10	0.920	1.0048	0.9162	1.1019
Not attending(Ref)							
Child's relationship to the household head							
Son/daughter	-0.4267	0.0732	-5.83	0.000*	0.6527	0.5654	0.7534
Not son/daughter(Ref)							
Sex of the household head							
Male	0.1131	0.0810	1.40	0.163	1.1198	0.9553	1.3125
Female(Ref)							
Literacy status of the household head							
Literate	-0.1745	0.0488	-3.58	0.000*	0.8399	0.7633	0.9241
Illiterate(Ref)							
Employment status of the household head							
Employed	0.6071	0.0981	6.19	0.000*	1.8351	1.5142	2.2240
Not employed(Ref)							
Marital status of the household head							
Married	-0.1114	0.0876	-1.27	0.203	0.8946	0.7535	1.0621
Not married(Ref)							
Household size							
5 and below(Ref)							
6-10	0.0472	0.0491	0.96	0.337	1.0483	0.9521	1.1543
11 and above	-0.4153	0.1629	-2.55	0.011*	0.6601	0.4798	0.9083
Constant	-0.5273	0.1830	-2.88	0.004*	.5902	0.4123	0.8448
Random effect		Estimate	S.E	Z-value	P-value		
Var(U_{0j})= σ_0^2		0.7359	0.414	1.78	0.0375		

$\text{Var}(U_{4j}) = \sigma_{4j}^2$	0.1924	0.093	2.08	0.0188
$\text{Cov}(U_{0j}, U_{4j}) = \sigma_{04}$	-0.1823	0.152	-1.20	0.2301
LR test vs. logistic regression: $\text{chi2}(3) = 266.47$ Prob > $\text{chi2} = 0.000$				

* Significant ($p < 0.05$)

Model	Obs	LL value	df	AIC	BIC
M2	12238	-6131.61	15	12293.22	12404.4

Table 4.10: Results of Model Selection Criteria and Log Likelihood Ratio Test

	Random Intercept and Fixed Slope Model	Random Coefficient Model
Log likelihood(LL)	-6151.97	-6131.61
-2LL=Deviance	12303.94	12263.22
AIC value	12329.94	12293.22
BIC value	12426.31	12404.4
Wald test		Wald $\text{chi2}(11) = 234.31$ Prob > $\text{chi2} = 0.000$

In Table 4.9, the value of $\text{Var}(U_{0j})$ and $\text{Var}(U_{4j})$ are the estimated variance of intercept and slope of place of residence respectively. These estimated variances are significant and suggest that the intercept and slope of place of residence vary significantly. So, there is a significant variation in the effect of place of residence across regions in Ethiopia.

The effect of intercept on region j is estimated to be $-0.5273 (0.1830) + U_{0j}$ with variance 0.7359 (standard error = 0.4142). The intercept variance of 0.7359 (standard error = 0.4142) is interpreted as the between-region variance when all other variables are held constant (i.e. equal to zero). The between-region variance of slope of place of residence is estimated to be 0.1924 (standard error = 0.0925).

The quantity AIC and BIC can be used to make an overall comparison of this more complicated model with the previous one. As shown in Table 4.10, values of AIC and BIC for the random coefficient model (AIC = 12293.22 and BIC = 12404.4) are less than the AIC and BIC values for the random intercept model (AIC = 12329.94, and BIC = 12426.3). This indicates that the random coefficient model is a better fit compared to the random intercept and fixed effect model. The random coefficient model involves two extra parameters, the variance of the slope residuals (for

place of residence), U_{4j} 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 the deviance based chi-square is given by $(12303.94-12263.22) = 40.72$, $p\text{-value} = 0.0001$) 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 in much the same way as those from the standard logit model. Thus, everything else being equal except a slight difference on random effect in the model, boys were 22.0% more likely to be engaged in child labour (aOR=1.22) compared to girls. Likewise, children in the age group 12-14 years were 38.2% (aOR=1.382) more likely to be engaged in child labour compared to children in the age group 5-11 years controlling for other variables in the model and random effects at level two.

Similarly, children who lived in urban households were 92.5% less likely to be engaged in child labour (aOR= 0.0748) than children who resided in rural households controlling for other variables in the model and random effects at level two.

The odds of children who are the sons/daughters of the household heads were 34.7% less likely to be engaged in child labour than children who are not the sons/daughters of the household heads controlling for other variables in the model and random effect at level two.

Likewise, children from households having literate household heads were 16.0% (aOR=0.840) less likely to be engaged in child labour than children from households having illiterate household heads controlling other variables in the model and random effects at level two.

Children who have employed household heads were 1.8351(aOR=1.8351) times more likely to be engaged in child labour than children who have unemployed household heads controlling for other variables in the model and random effects at level two.

Finally, children from households of size eleven and above were 34.0% (aOR=0.660) less likely to be engaged in child labour compared to children from households of size five or less controlling for other variables in the model and random effects at level two. However, the odds

of child labour for children from households of size 6 to 10 was not significantly different from the odds of child labour for children from households of size five or less.

4.6. Goodness of Fit Test

Based on the result we obtained in Table 4.10 (random coefficient model), the deviance chi-square is significant and the values of AIC and BIC are less than the AIC and BIC values we obtained for the random intercept with fixed slope model. So, we conclude that the random coefficient model is a better fit.

CHAPTER FIVE

5. DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1. Discussion

This study intended to identify determinants of child labour (whether a child is engaged in child labour or not) among 5-14 years old children in Ethiopia based on evidence from Ethiopian National Labour Force Survey (ENLFS, 2013) data. Consequently, descriptive analysis, single level multiple logistic regression and multilevel logistic regression analyses were used.

Male children were more likely to be engaged in child labour compared to female children controlling for other variables in the model. The reason might be that girls are more engaged in household chores (domestic works) than boys. This is in agreement with the findings in other studies (Assefa Admassie, 2000; Dawit Senbet, 2010; Moyi, 2011; Webbink et al., 2010; and Mudzongo and Whitsel, 2013). They concluded that girls are less likely to be engaged in child labour than boys because girls are more involved in housework. Older children were more likely to be engaged in child labour than younger children. This is also consistent with the findings in other studies. (Mudzongo and Whitsel, 2013; Cockburn and Dostie, 2007; and Laurent and Sebastian, 2010).

Children's place of residence was found to be significantly associated with the incidence of child labour. Children residing in rural areas were found to be more likely to be engaged in child labour than children in urban areas. This finding is in line with the finding that child labour is, for obvious reasons, particularly widespread in developing countries, and more so in rural than in urban areas (Nkamleu, 2009). The same result was also confirmed by Mudzongo and Whitsel (2013) and Laurent and Sebastian (2010). Assefa Admassie (2000) also found out that the subsistence nature of economic life, tradition, and cultural values force rural communities in Ethiopia to involve their children in economic activities thereby perpetuating child labour.

Children of household heads were less likely to be engaged in child labour compared to those children who are not. It is assumed that being a direct child of the household head is the basis for parental altruism and children who are not the son/daughter of the household heads may, therefore, be more involved in work (Dawit Senbet, 2010; and Ainsworth, 1996).

Children from households having literate household heads were less likely to be engaged in child labour than children from households having illiterate household heads. This finding is consistent with the finding by Abou (2014).

Household heads employment status affect child labour positively. This means that children from employed household heads were more likely to be engaged in child labour than children from non-employed household heads. This result was also confirmed by Tharmapornphilas (2009).

Household size was found to be negatively associated with the incidence of child labour. Children from households of size 11 or above were less likely to be engaged in child labour compared to children from households of size five or less. This finding is consistent with the finding by Laurent and Sebastian (2010).

5.2. Conclusion

The descriptive results show that about 32% of the sampled children were engaged in child labour in Ethiopia. The findings of this study identified region, sex of the child, age of the child, place of residence, child's relationship with the household head, literacy status of the household head, employment status of the household head and the household size as significant predictors of the incidence of child labour in Ethiopia.

Results from both single level multiple logistic regression and multilevel logistic regression analyses indicated that older children were more likely to be engaged in child labour than younger children. Also, children were less likely to be engaged in child labour as the size of households increase.

The multilevel logistic regression results revealed that only the effect of place of residence varied across regions whereas the effects of other covariates on the incidence of child labour remained fixed across regions.

5.3. Recommendations

Based on the findings of our study, we forward the following recommendations that will help to decrease the incidence of child labour in Ethiopia:

- Raise adult literacy levels by strengthening existing programs and creating new ones throughout Ethiopia.

- Improve the severe situation of child labour in rural parts of Ethiopia by establishing the school infrastructures and creating awareness within the society.
- Address regional disparities in child labour by allocating more resources to regions with high child labour, while at the same time taking into account gender effects.
- Enforce the implementation of the existing laws and regulations properly and protect children from child labour especially in rural parts of the country.
- Give more support and emphasis to regions with high rates of child labour.

5.4. Areas of further research

A survey specific to child labour with detailed questionnaire is required to provide more information and arrive at a more complete list of determinants of child labour in Ethiopia. Moreover, additional factors on socio-economic, demographic and community characteristics, such as household income (wealth), schooling access, school quality/type, infrastructure and socio-cultural practice should be included as potential predictors of child labour since they may contribute to the unexplained effects – so far, little effort has been applied in investigating them. Contrary to our expectation, children from employed household heads were more likely to be engaged in child labour than children from unemployed household heads. This calls for further research as some researchers have confirmed that the probability of children to be engaged in child labour decreases as the income of households increase, and most of the time the household heads are the breadwinners of the households.

References:

- Abas, A.T. (2014). Modelling the key Determinants of Child Labour in Somalia. University of Nairobi.
- Abou, E.P. (2014). A Re-examination of the Determinants of Child Labour in Côte d'Ivoire: AERC Research Paper 289 African Economic Research Consortium, Nairobi December 2014.
- 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.
- Ahmed, S.S., Haider, W. and Khan, D. (2012). Determinants of Child Labour in Khyber Pakhtunkhwa: An Econometric Analysis. Munich Personal RePEc Archive (MPRA) Paper No. 73526. Online at <https://mpra.ub.uni-muenchen.de/73526/>.
- Ainsworth, M. (1996). Economic aspects of child fostering in Cote d'Ivoire. Research in Population Economics, 8(1), 25-62.
- Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.
- Assefa Admassie (2000). The Incidence of Child Labour in Africa with Empirical Evidence from Rural Ethiopia: ZEF Discussion paper on development policy no. 32. Bonn.
- Bachman, S. (2000). A new Economics of Child Labour: Searching for Answers Behind the Headlines. Sweden: Pearson.
- Belsley D. A, E. Kuh, and R. E. Welsch. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.
- Bisrat Abebe (2014). What Determine Child Labour in Rural Ethiopia? (Panel Data Analysis). Graduating Thesis, Addis Ababa University, Ethiopia.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed Models. J. Am. Statist. Assoc., 88, 9–25.
- Central Statistical Agency (2014). Analytical Report on The 2013 National Labour Force Survey. Statistical Bulletin. Addis Ababa.
- Central Statistical Agency (CSA): Ethiopia Child Labour Survey (ECLS, 2001). Addis Ababa.

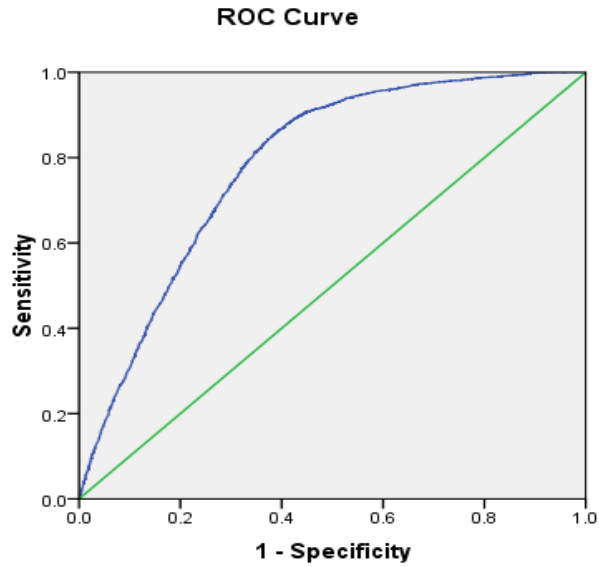
- Cockburn, J. and Dostie, B. (2007). Child work and schooling: The role of household asset profiles and poverty in rural Ethiopia. *Journal of African Economies*, 16(4), 519–563.
- Cook, R. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, New York.
- Copas, J.B. (1988). Binary Regression Models for Contaminated Data. *Journal of Royal Statistical Society. Series B*, vol. 50 No. 2, pp. 225-265.
- Coster, A.S and Adekoya, M.I (2014). Determinants of Child Labour and Schooling in Rural Farming Households in Ogun State, Nigeria: *Asian Journal of Agriculture and Food Science* (ISSN: 2321 – 1571) Volume 02 – Issue 04, August 2014.
- Dawit Senbet (2010). Determinants of Child Labour versus Schooling in Rural Ethiopia, *European Journal of Social Sciences*; Nov2010, Vol. 17 Issue 3, p414.
- Drusilla, K. (2001): Child labour in Latin America: Policy and Evidence. Working Paper, Department of Economics, Tufts University.
- EDHS (2011). *Ethiopian Demographic and health survey (EDHS)*, Addis Ababa, Ethiopia and Calverton, Maryland, USA.
- Ethiopian Labour Proclamation (2004). No. 377/2003. Addis Ababa, Ethiopia.
- Gelfand, A. E., Hills, S.E., Racine-Poon, A. and Smith A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85, 972–985.
- Goldstein, H. (1991). Nonlinear multilevel models with an application to discrete response data. *Biometrika*, 78, 45–51.
- Goldstein, H. (2003). *Multilevel Statistical Models*. 3rd edition (2003) London: Arnold.
- Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *J. Roy. Statist. Soc. A*, 159, 505–513.
- Guarcello, L. and Rosati, F. (2007). *Child Labour and Youth Employment: Ethiopia Country Study*. Social Protection Discussion Paper No. 0704, the World Bank.
- Hosmer, W.D. and Lemeshow, S. (2000). *Applied Logistic Regression*. 2nd Ed., John Wiley and Sons, New York.
- Hox, J.J. (2010). *Multilevel Analysis Techniques and Applications*. Second Edition, Utrecht University, The Netherlands.

- ILO-IPEC (2013). Marking progress against child labour - Global estimates and trends 2000-2012.
- Kebebew Ashagrie (2000). Current Progress in Implementing New Methods and Conducting Innovative Surveys for Measuring Exploitation of Children. Session I-PL 0. ILO, Geneva.
- Laird, N. M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika*, 65, 581–590.
- Laurent, N. and Sebastian, D. (2010). Characteristics and determinants of Child Labour in Cameroon. CSAE Conference: Economic Development in Africa, St. Catherine's College, Oxford. UK.
- Maddala, G.S. (1997). Introduction to Econometrics. 2nd Ed. Prentice-Hall, Inc., USA.
- McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of American Statistical Association*, 92, 162–170.
- Moyi, P. (2011). Child labour and school attendance in Kenya: Educational Research and Reviews Vol. 6(1), pp. 26-35, Academic Journals.
- Mudzongo, C.C. and Whitsel, C.M. (2013). Determinants of Child Labour in Malawi and Tanzania. *Journal of Community Positive Practices*, XIII(3) 2013, 3-24 ISSN Print: 1582-8344; Electronic: 2247-6571.
- National Action Plan of Ethiopia (2013). Addis Ababa, Ethiopia.
- National Bureau of Statistics Republic of Tanzania (2009). Basic Information Document National Panel Survey (NPS 2008-2009).
- National Statistical Office of Malawi (NSO) (2011). Malawi Second Integrated Household Survey (IHS2) 2010-2011.
- National Statistic Office Thailand (2003). Thailand Labour Force Survey.
- Nkamleu, G.B. (2009). Determinants of Child Labour and Schooling in the Native Cocoa Households of Côte d' Ivoire. AERC Research Paper No. 190.
- Ravallion and Wodon (1999). Does Child Labour Displace Schooling? Evidence on Behavioral Responses to an Enrollment Subsidy. *The Economic Journal*, 110 462, 158-175.
- Save the Children Denmark (2003). Child Labour in Ethiopia with special focus on Child Prostitution. Addis Ababa.

- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461– 464.
- Snijders, T. and Bosker, R. (1999). *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling*. London/ Thousand Oaks/ New Delhi: Sage Publications.
- Solomon Sorsa and Alemu Abera (2006). A study on child labour in three major towns of southern Ethiopia. *Ethiop.J.Health Dev.* 2006;20(3).
- Tharmmapornphilas, R. (2009). *Determinants of Child Labour in Thailand*, Columbia University. UMI Number: 3348405.
- Webbink, E., Smits, J. and de Jong, E. (2010). Hidden child labour: Determinants of housework and family business work of children in 16 developing countries. NiCE Working Paper 10-110 October 2010.

Appendices

Appendix A



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.784	.004	.000	.776	.793

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

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

