DEBRE BERHAN UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

POST GRAGUTE STUDIES

DEPARTMENT OF STATISTICS

DETERMINANTS OF FOOD SECURITY IN RURAL HOUSEHOLDS OF THE AMHARA REGION

By:

SHEWAYIREF GINTO

A THESIS SUBMITTED TO SCHOOL OF GRADUATE STUDIES COLLEGE OF NATURAL & COMPUTATIONAL SCIENCE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE MASTER OF SCIENCE DEGREE IN BIOSTATISTICS

AUG 2021

DEBRE BERHAN,

ETHIOPIA

DEBRE BERHAN UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

SCHOLL OF POST GRAGUTE STUDIES,

DEPARTMENT OF STATISTICS

DETERMINANTS OF FOOD SECURITY IN RURAL HOUSEHOLDS OF THE AMHARA REGION

BY

SHEWAYREF GINTO

ADVISOR: GEBREMEDEHIN D.(ASS.PRO)

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES COLLEGE OF NATURAL & COMPUTATIONAL SCIENCE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE MASTER OF SCIENCE DEGREE IN BIOSTATISTIC

AUG, 2021

DEBRE BIRHAN,

ETHIOPIA

DECLARATION

I declare this thesis is my original work, has not been presented for degrees in any other University and all sources of materials used for the thesis have been duly acknowledged. The assistance received during the course these investigations has been duly acknowledged. Therefore, we recommended it to be accepted as fulfilling the thesis requirements.

Name: SHEWAYREF GINTO

________________ ___________________

Signature *Date*

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

GEBREMEDHIN D.(ASS.PRO)

________________ _____________

Signature *Date*

APROVAL SHEET

We, the undersigned, members of the board of examiners of the final open defense by Shewayref Ginto have read and evaluated his thesis entitled "**Determinants of food security in rural households of the Amhara region**; application of **ordinal logistic Regression Model**" and examined the candidate. This is therefore to certifying that the thesis has been accepted in partial fulfillment of the requirement for the degree of Master of Sciences in Statistics with specialization of Biostatistics.

Name of Chairman	Signature	Date
Name of Advisor	Signature	Date
Name of Internal Examiner	Signature	Date
Name of External Examiner	Signature	Date
Associate Dean, College Res/CS& Postgraduate	Signature	Date
Dean collage of Graduate Studies	Signature	Date

Final approval and acceptance of the thesis is contingent up on the submission of the thesis of the final copy of the thesis to the college of graduate studies (CGS) through the department of graduate committee (DGC) of the candidate's

Stamp of CGS

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

ACKNOWLEDGMENTS

I would like to thank my God and for helping me in my life and getting chance to pursue my graduate education at University of Debre Birhan.

I express my sincere gratitude to my advisor Gebremedin D. (Ass.pro) for his time, continuous, support, patience, motivation and immune's knowledge towards the writing of this thesis. His matured experience and knowledge guided me in all the time of research and writing of this thesis.

Furthermore, I would like to thank my friends and classmates that contributed in diverse ways to my success in this program; I say a big thank you. I would also gratefully and sincerely thank the Central Statistical Agency, for giving me HCE data of 2015/2016 and the data management staff of technical assistance.

Finally, my thanks go to my family, both home and aboard for their continued support, prayer, contributions and bearing with me throughout this program of study.

Contents

List of Abbreviations

CDR	Child Dependency Ratio
CSA	Central Statistical Agency
EA	Enumeration area
EEA	Ethiopian Economic Association
FAO	Food and Agricultural Organization
GAO	Government Accountability Office (USA)
HFS	Household Food Security
HH	Household
HCES	Household, Consumption and Expenditure Survey
IFPRI	International Food Policy Research Institute
Kcal	Kilo calorie
LDA	Linear Discriminate Analysis
MEDaC	Ministry of Economic Development and Cooperation
MoFED	Ministry of Finance and Economic Development
POM	proportional odds model
PPOM	Partial proportional odds model
SPSS	Statistical Package for Social Science
VIF	Variance Inflation Factor
WFS	World Food Summit

LIST OF TABLES

LIST OF FIGURES

Abstract

Food security is a difficult concept to measure since it deals in very broad terms with the production, distribution and consumption of food. Food insecurity on the other hand lends itself more readily to measurement and analysis. Food security refers to the availability of food whereas famine and hunger are the consequence of the non-availability of food, in other words the results of food insecurity. This study investigated the determinants of food security and identified the major factors that jointly discriminate the rural households of Amhara region into food insecure, marginally food secure and food secure households. The study is made based on the 2015/2016 Household Consumption and Expenditure Survey (HCES) which were conducted by Central Statistical Agency (CSA). Calorie method was employed to determine food security. To achieve the objective of this study descriptive statistics, chi-square test of association and partial proportional odds model and related tests were used for data analysis using socio-economic and demographic related variables as explanatory variables and household food security as the response variables. The descriptive results revealed that about 41.14% of the households are food insecure, 13.60% marginally food secure and 45.26% were food secure. The result of the partial proportional odds model revealed that the variables marital status, education, household size, religion, income and employment status were found to be significant determinants of household food security. As rural part of Amhara region is constantly facing food insecurity and famines, there is a need for integrating famine relief and prevention strategies at the regional level with the overall development strategy.

Keywords: Determinants of food security, ordinal logistic regression, chi-square test, partial proportional odds model

1. INTRODUCTION

1.1. Background

We are living in a world where more than one billion people are poor, 800 million are food insecure, and where about 170 million children are malnourished. While food insecurity occurs in most countries to varying degrees, 75 % of the food insecure lives in rural areas of developing countries (Bedeke 2012).Food is both a need and human right, but food insecurity is prevalent in today's world in general, and in sub-Saharan Africa in particular. Since early 2007, food-related riots have occurred in 15 countries, including 7 in sub-Saharan Africa (Benson, Minot et al. 2008)

Africa, which reversed from being a key exporter of agricultural commodities into being a net importer, has the highest percentage of undernourished and has shown the least progress on reducing the prevalence of undernourishment in the last 30 years(Clover 2003). Chronic food insecurity now affects about 200 million people who are suffering from malnutrition. Acute food insecurity in 2003 affected 38 million people in Africa who are facing the outright risk of famine, with 24,000 dying from hunger daily. Famines are the most visible and extreme manifestation of acute food insecurity. Out of the 39 countries worldwide that faced food emergencies at the beginning of 2003, 25 are found in Africa including Ethiopia (Endris and Nura 2018, Oluoko-Odingo, Odingo et al. 2018).

As part of Africa, Ethiopia faces daunting poverty and food insecurity challenges that are worsening over time. About half of Africa's food insecure population lives in Ethiopia, Chad, Zaire, Uganda, Zambia and Somalia (Adenew 2004). In the 1990s, an estimated 30 million Ethiopians were food insecure, and food crises were persistent. Among this food insecure people the majority reside in the rural areas of the country. About 52% of the rural population and 36% of the urban population consume under the minimum recommended daily intake of 2100 calorie per person per day (Devereux and Sussex 2000).

The world development report indicators for the year 2013/14 reveal the prevalence of child malnutrition (children under age 5) is 48% during the period 2008-2015 (ALEMU and Ayalew 2016).

Ethiopia has reasonably good resource potential for development –agriculture, biodiversity, water resource, minerals, etc. Yet, Ethiopia is faced with complex poverty, which is broad, deep and structural. The proportion of the population below the poverty line is 44% in 2007/8 (Bogale and Korf 2009).

The presence of hunger in Ethiopian households due to insufficient resources to obtain food has been a long-standing challenge to Ethiopian government, donors, and other local and international organizations like food and agriculture organization (FAO) (Mohamed 2017). Despite significant amounts of food aid assistance over recent years, there has been little measurable impact in reducing food insecurity. The reason behind is that food insecurity is a complex, multidimensional phenomenon which varies through continuum of successive stages as the condition becomes more severe. Each stage consists of characteristic conditions and experiences of food insufficiency to fully meet the basic needs of household members, and of the behavioral responses of household members to these conditions (Leroy, Ruel et al. 2015).

1.2. Statement of the problem

Ethiopia had been faced with many droughts and many people had died of famine than other problems particularly in the epidemic periods of 1957-58, 1964-65, 1983-84, 1998-99, and 2003 (EEA, 2005).Since the last major famine of 1984/5 (when excess mortality may have reached one million), Ethiopia has been affected by recurrent droughts (Devereux, 2000). Some droughts were exacerbated by civil conflict, which undermined food production and inhibited government and donor responses to the harvest failure(Devereux and Sussex 2000).

In Ethiopia, the seriousness of the famine and food shortage varies from one area to another on the state of natural resources and the extent of development of these resources(Mitiku, Fufa et al. 2012)

Most famines and food crises have been geographically concentrated along two broad belts of the country. The first belt consists of the mixed farming production system area of the central and northern highlands, stretching from northern Shewa through Wollo and Tigray. The land resources, mainly the soils and vegetation of this part of the country have been highly degraded because of the interplay between some environmental and human factors such as relief, climate, population pressure and the resultant over-cultivation of the land, deforestation of vegetation and overgrazing (Endris and Nura 2018).

The second belt is made up of the low-lying agro-pastoral lands ranging from Wollo in the North, through Hararge and Bale to Sidamo and Gamo Gofa in the South (Endris and Nura 2018).

The study area, Amhara region, is one of the food deficient regions of the country, which falls in the first and second drought prone belt. As a result of the food deficient situation in the region, where even in a good year farm households can only meet 60% of their total food needs and the remaining is filled by food aid -both free and Food-For-Work (Endris and Nura 2018).

Although the seriousness of food shortage varied from year to year, farm households faced seasonal food shortage almost every year. Food secure and food insecure farm households reside as neighbors and could share common climate and weather situation and mainly similar socioeconomic, cultural and land topography. Yet, one faces seasonal food crises and become dependent on food aid, while the other remains food secure, requiring no food aid. This clearly shows poverty and transitory food insecurity are deep-rooted in the study area. Although drought plays a paramount role in triggering food crises, the difference in consumption status of farm households between good year and bad year is not so significant to claim that drought is the central cause of famine or transitory food insecurity. This implies poverty and seasonal food insecurity in the region are mainly determined by structural, socio-economic, cultural, demographic and other factors. Researchers studied the determinate of food security using some set of variables and statistical methods such as binary logistic regression model (Dagne 2016) and (Bogale and Shimelis 2009) and discriminant analysis (Asmelash 2014). In this study partial proportional odds models is more appropriate for determinant of food security. Hence, the central question of this study is:

- ✓ How can households‟ food security status be measured?
- ✓ What is the extent of food insecurity as disaster risk among rural households of Amhara region?
- ✓ What are the determinants of food insecurity among rural households‟ of Amhara region?

1.3. Objective of the study

1.3.1. Main objective

The main objective of this study is to assess the status of food security and its major determinants in the rural households of the Amhara region.

1.3.2. Specific Objectives

The **specific objectives** of this study include:

1. To examine the effects of some variables that may influence food insecurity of rural households;
2. To describe the relationship between food insecurity and its determinants; and
3. To apply logistic regression analysis in classifying rural households of Amhara region based on their status of food security.

1.4. Significance of the study

This study was carried out for academic purposes and it is confined only Amhara region in Ethiopia. However, the findings of the study are thought to be very helpful to have a deeper comprehension about the food security status of the Amhara region rural community in general and the surrounding area in particular. It contributes a lot to figure out the food insecurity problems of the rural households that are practically challenging them at present. Therefore, the result of this study is thought to be a crucial input to the current government's endeavor to alleviate the prevailing food insecurity problems in rural areas and bring about sustainable development.

2. LITERATURE REVIEW

2.1. Conceptual Framework

2.1.1. Concepts and Definitions of Food Security

Since the World Food Conference in 1974 due to food crises and major famines in the world, the term Food Security was introduced, evolved, developed and diversified by different researchers. (Maxwell and Smith 1992) listed 194 different studies on the concept and definition of Food Security and 172 studies on indicators. A review that updates this literature (Clay, 1997) provides an additional 72 references. In the work by Maxwell and Frankenberger, a distinction is made between process indicators (those that describe food supply and food access) and outcome indicators (those describe food consumption) (Maxwell, Ahiadeke et al. 1999).

Food security was understood as adequacy of food supply at global and national levels until the mid1970's. This view favored merely food production oriented variables and overlooked the multiple forces which in many ways affect food access. Evidences show that during the last two decades, food production has been increasing in the world. However, large amount of food at global level does not guarantee food security at national level. Moreover, availability of enough food at national level does not necessarily ensure household food security. For instance, in 1990, the calorie supply at global level was more than 110 percent compared to the total requirement. However, during the same period, more than 100 million people were affected by famine and more than a quarter of the world's population was short of enough food (Ranganathan, Vennard et al. 2016). Although food production has been increasing from time to time, food insecurity, malnutrition and hunger and much more serious problems would remain the main agenda in the globe today (Barrett, Lentz et al. 2013, Ranganathan, Vennard et al. 2016).

As the occurrence of hunger, famine, and malnutrition are increasing from time to time in developing countries, the conceptual framework of food security has also progressively developed and expanded. The idea of food security attained wider attention since the 1980s after the debate on 'access' to food and the focus of the unit shifted from global and national levels to household and individual levels (GEMEDA 2020). This paradigm came with new concept and definition of food security and it led to two additional major shifts in thinking; from a first food approach to a livelihood perspective and from objective indicators to subjective perceptions (Maxwell 1994).

The most commonly accepted definition of Food security is “access by all people at all times to enough food for an active and healthy life” (Hindle 1990). Food insecurity is a situation in which individuals have neither physical nor economical access to the nourishment they need. A household is said to be food insecure when its consumption falls to less than 80% of the daily minimum recommended allowance of caloric intake for an individual to be active and healthy. In particular, food insecurity includes low food intake, variable access to food, and vulnerability- a livelihood strategy that generates adequate food in good times but is not resilient against shocks. These outcomes correspond broadly to chronic, cyclical, and transitory food insecurity, and all are endemic in Ethiopia (Desalegn and Ali 2018).

During the debates that preceded the World Food Summit (WFS) held in Rome in 1996, it was established that "There is food security when all people at all times have sufficient physical and economic access to safe and nutritious food to meet their dietary needs including food preferences, in order to live a healthy and active life"(Boon 2007). When an individual or population lacks, or is potentially vulnerable due to the absence of, one or more factors outlined above, then it suffers from, or is at risk of, food insecurity. Based on the WFS (1996), the definition focuses on three distinct but interrelated elements, all three of which are essential to achieving food security:

- **Food availability:** having sufficient quantities of food from household production, other domestic output, commercial imports or food assistance,
- **Food access**: having adequate resource to obtain appropriate foods for a nutritious diet, which depends on available income, distribution of income in the household and food prices,
- **Food utilization:** proper biological use of food, requiring a diet with sufficient energy and essential nutrients, potable water and adequate sanitation, as well as knowledge of food storage, processing, basic nutrition and child care and illness management.

The concept of food security also has spatial and temporal dimensions. The spatial dimension refers to the degree of aggregation at which food security is being considered. It is possible to analyze food security at the global, continental, national, sub-national, village, household, or individual level (Hoddinott 1999).

The temporal dimension refers to the time frame over which food security is being considered. In much of the food security literature, temporal dimension is almost universally classified in to two states-chronic or transitory (Kalkuhl, Kornher et al. 2013). Chronic food insecurity is a long-term or persistent inability to meet minimum food consumption requirements; while transitory food insecurity is a short-term or temporary food deficiency. An intermediate category is cyclical food insecurity, such as seasonality. Transitory is often used to imply acute, with the corollary assumption that chronic equates to mild or moderate food insecurity (Devereux, Mthinda et al. 2007).

The worst form of transitory food insecurity is famine (Devereux, Mthinda et al. 2007). Hence, transitory food insecurity faced by farm households will be understood in this study as a seasonal food shortage of any magnitude ranging from mild to severe. It can also be noted that the concepts of transitory food insecurity and seasonal food shortages are synonymous and will be used interchangeably in this study. As the Ethiopian farming system is mainly dependent on rain-fed agriculture, seasonality adversely affects the food security situation of the country.

2.1.2. Theoretical Approaches to Food Security

The general approach has pointed out a number of environmental and socio-economic attributes assumed to explain famine and food security. The principal ones include: rapid population growth, war and civil strife, drought, ecological degradation, government mismanagement, unequal access to resources and unequal exchange, socio-economic and political dislocation (Tolossa 2001). The argument of this approach is that one or a combination of theses can disrupt food production. However, production failure may or may not result in famine or food insecurity. Due to this fact, the attributes (factors) are not precise explanations of the causation of the process of famine. It is in response to this major problem weakness that the specific approaches (models) of famine emerged (Fisseha 2017).

2.2. Empirical Review of Causes and Determinants of Food Insecurity

The empirical review for this study is organized under three sections. The first section presents some causes of food insecurity documented in Ethiopia and other developing countries of the world particularly in Africa. The second part presents determinants of food security in Ethiopia. The last part presents and generalizes the findings of certain previous studies concerning the determinants of food insecurity.

2.2.1. Causes of Food Insecurity

2.2.1.1. Causes of Food Insecurity in Other Developing Countries

Achieving food security in its totality continues to be a challenge not only for the developing nations, but also for the developed world. The difference lies in the magnitude of the problem in terms of its severity and proportion of the population affected.

(Mwangi and Kariuki 2015)mentioned the main causes of food insecurity in developing countries. Some of them include: unstable social and political environments that preclude sustainable economic growth , war and civil strive, macro-economic imbalances in trade, natural resource constraints, poor human resource base , gender inequality, inadequate education, poor health, natural disasters, such as floods and locust infestation, and the absence of good governance. All these factors contribute to either insufficient national food availability or insufficient access to food by households and individuals.

A study by (Boussard, Daviron et al. 2005, 2009)found that 99% of the food in Sub-Saharan Africa is grown under rain fed agriculture. Hence, food production is vulnerable to adverse weather conditions. The reason behind is that there was an over decline in farm input investment including fertilizers, seeds, and technology adoptions.

Other causes include rapid population growth, limited access to agriculture-related technical assistance, underdeveloped agricultural sector and lack of knowledge about profitable soil fertility management practices leading to expansion in to less-favorable lands. Barriers to market are also causes of food insecurity in Africa (Mwaniki 2006). As he mentioned some barriers of market access were poor infrastructure, market standards, limited information, and requirements for large initial capital investments, limited product differentiation, and handicapping policies.

Diseases and infection are also identified as causes of food insecurity. (Karuhanga 2008) found that diseases such as malaria, tuberculosis and mainly HIV/AIDS not only reduce the man hours available to agriculture and household food acquisition, but also increase the burden of household in acquiring food.

Migration of male labor is also recognized as a cause of food insecurity. A study conducted in Lesotho village found that women and children suffered from lack of food and hygiene because women were too exhausted to cook and clean at times of peak agricultural work (Momsen 2008)). (Kenneth 2008)observed that growing cash crops at the expense of subsistence crops has largely contributed to seasonal food deficiency among the Gernieri in Gambia. He also observed that illness of adults at critical times in the production process adversely affects labor efficiency and productivity, which in turn contributes to seasonal food shortage.

Deterioration in the ecological conditions of production has also been seen as a cause of seasonal hunger in several African countries. Associated with this, (Ogbu 1973)noted insufficient farm land, low yields on farms and high storage losses of staples as the principal causes of food shortage in Nigeria.

A similar research conducted by (Toulmin and Gustavsen 1996)noted that the people of Bambara of Kala in Mali face seasonal food shortages that are mainly induced by two principal factors: one of the factors is climatic, specifically low and highly variable rainfall making the people very vulnerable to crop failure. The second class of risk is demographic, consisting of high level of mortality, varying level of fertility and vulnerability of all producers to sickness and disability.

2.2.2. Determinants of household food security

Those discussed in the above section determinants of food security at National, Regional or community levels. However, a study by Keshav (2006) shows that commonly used indicators of food security at the regional and national level or community level is often poor predictors of household food security. The study also made comparison among households based on depth and severity of food insecurity and found that socio-economic factors are the main determinants of food insecurity. The study concluded that both depth and severity of food insecurity are higher in occupational castes, small farms and less livestock holders, laborers, and households having minimum expense.

A number of studies made use of various methodologies to identify determinants of food security in different parts of Ethiopia. According to studies conducted by (Asmelash 2014); livestock ownership, farmland size, family labour, farm implements, employment opportunities, market access, level of technology application, level of education, health status, weather conditions, crop disease, rainfall, oxen ownership and family size were identified as major determinants of farm households' food security in Ethiopia.

A study by (Haile, Alemu et al. 2005)conducted in Koredegaga Peasant Association, Oromia Zone, identified that farmland size, per capita aggregate production, fertilizer application, household size, ox ownership, and educational attainment of farm households heads had a significant influence on food security. The computed partial effects at sample means using results from the logistic regression model indicated that a unit change in farmers' access to fertilizer or educational level of household heads or farmer's access to land or access to family planning improve the probability of food security in the study area.

Another similar study by (Aragie and Genanu 2017)conducted in North Wollo revealed that per capita land holding, cereal production, livestock, educational level of household heads, fertilizer use and family size were the major determinants of food security. They constructed food balance sheet and food security causation was examined using a binary logistic regression model.

2.2.3. Generalizations of the Causes and Determinants of Food Insecurity

From the theoretical and empirical causes and determinants of food insecurity, it can be generalized that food insecurity is a function of environmental crises, rapid population growth, poor assets basis, socio-cultural related issues, and poor access to market and infrastructure. Hence, in this sub-topic it is attempted to review relevant literatures particularly conducted in Ethiopia.

2.2.3.1. Demographic Factors

The population of Ethiopia is rising from time to time. Currently the Ethiopian population is about 115 million which grows by 2.6 % (Ayelign and De Saeger 2020).According to (Ayelign and De Saeger 2020)the average household size is also large when compared with other Sub-Saharan countries. At the micro level, household size is one of the factors expected to have influence on food security status of households. The majority of farm households in Ethiopia are small scale semi-subsistence producers with limited participation in non-agricultural activities since land holding size and financial capital to purchase agricultural inputs is very limited. (Kidane, Alemu et al. 2005) in his work found that family size tends to exert more pressure on consumption than the labor it contributes to production.

Another demographic factor that strongly influences household food security is sex of the household head. Studies by (Degefa, Ababneh et al. 2006), and (Kidane, Alemu et al. 2005)independently conducted in different parts of rural Ethiopia came out with common conclusion that the livelihood of female headed households was disadvantaged when compared with their male counterparts. This is due to the fact that, the researchers justify, female household heads have limited access to livelihood assets like land, education, saving, labor force and oxen (drought power), livestock and credit services.

2.2.3.2. Environmental Crises

The combined effect of land based resources degradation like deforestation, soil erosion, flooding, and loss of agricultural and pasture land leads to production decline (Melese 2016). Rapid population growth and recurrent drought are causing serious resource degradation. (Adgo, Selassie et al. 2014) described that the seriousness of shortage of productive (fertile) land in the highland areas, coupled with population pressure, have forced the cultivation of the steep and moderate slopes which are highly degraded because of soil erosion.

Climate is one of the important elements of the natural environment that positively or negatively affects the food security status of rural households. Many studies indicated that inadequate and erratic rainfall is one of the environmental phenomena, causing food crises in many rain fed farming and drought prone areas across the world. In Ethiopia more than 95% of food grain production is from rain fed subsistence farm ((FIKIRE and Bekele 2014). A study conducted in Ethiopia by Devereux (2002) revealed that a 10% decline in rainfall below its long term average reduces national food production by 4.4%.

2.2.3.3. Socio-Cultural Factors

Education has a tremendous influence on the food security status of households. Educational attainment by the household head could lead to awareness of the possible advantages of modernizing agriculture by means of technological inputs; enable them to read instructions on fertilizer packs and diversification of household incomes which, in turn, would enhance household's food supply (Mannaf and Uddin 2012).

Socio-cultural events such as eating habit and food preference, cultural ceremonies and festivals also influence the food security status of the given communities and way of saving or expenditure, also directly or indirectly affects the food security situation of that particular community.

2.2.3.4. Access to Infrastructure

Access to infrastructure such as market center and roads promote livelihood diversification and agriculture intensification. Adequate infrastructure, especially main and feeder roads that improve access to necessary input-fertilizer, seed, pesticide chemicals and other agricultural implements are very indispensable (Sabila 2014). Although, the current government has made a significant progress particularly in road development, the sector is still weak even compared with the African average. World Bank (2007) reported that due to lack of proper and on time transportation facilities post-harvest total production loss reached up to 30%.

Generally, as indicated in many literatures, inadequate infrastructures and social services development such as road, transportation, communication, electrification, education and health services and agricultural services would be major challenges to sustain the growth of agricultural production and food security.

3. METHODOLOGY

3.1. Data Source

The data for this study was taken from Household Consumption and Expenditure (HCE) survey and which were conducted by CSA in 2015/2016.

For this surveys, the list of households obtained from the 2001/2 Ethiopian Agricultural Sample Enumeration (EASE) was used as a frame to select Enumeration Areas (EAs) from the rural part of the country and the 2004 Ethiopian Urban Economic Establishment Census (EUEEC) was used as a frame in order to select sample EAs from the urban part of the country. A fresh list of households from each rural and urban EAs was prepared at the beginning of the survey period. This list was, thus, used as a frame in order to select households from sample EAs.

For the purpose of the survey the country was divided into three broad categories. That is; rural, major urban centers and other urban center categories. The first category consists of the rural areas of eight regional states and two administrative councils (Addis Ababa and Dire Dawa) of the country. Each region was considered to be a domain (Reporting level) for which major findings of the survey are reported. This category comprises 10 reporting levels. A stratified two-stage cluster sample design was used to select samples in which the primary sampling units (PSUs) were EAs. Twelve households per sample EA were selected as a second stage sampling unit (SSU) to which the survey questionnaire were administered. The second category includes all regional capitals and four other urban centers that have relatively larger population sizes. Each urban center in this category was considered as a reporting level. A stratified two-stage cluster sample design was also adopted in this instance. The primary sampling units were EAs of each urban center. Sixteen households from each sample EA were finally selected as a secondary sampling unit. The third category includes other urban centers in the country other than those in the second category. Unlike the above two categories a stratified three-stage cluster samples design was adopted to select samples from this category.

Totally 2,106 EAs and 30,240 households were selected for HCE survey. Sample EAs of each reporting level, were selected using probability proportional to size (PPS) with systematic sampling techniques, size being number of households from the 2001/2 EASE. Twelve households per EA were systematically selected from the fresh list of households prepared at the beginning of the survey.

The rural part of Amhara region is one of the reporting levels of the first category. Hence, from rural Amhara 168 EAs and 2015 households (HHs) were selected for HCE survey. Hence, the sample size of this study is 2015 households (that is, n=2015).

3.2. Variables in the study

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

3.2.1. The Dependent variable

The dependent variable in this study is Household Food Security (HFS) status. Consumption based rather than income-based measure of HFS status is used in this study. This is because consumption better captures long-run welfare, and it better reflects household's ability to meet their basic needs. Consumption is preferable to measure HFS than income because it is less vulnerable to seasonality and life-cycle, less vulnerable to measurement errors because respondents have less reasons to lie, it is closer to the utility that people effectively extract from income, and for the poor most of income is consumed.

The HFS status was determined using the consumption approach based on the 2015/16 HCE survey conducted by CSA. Following this approach, household food security status was set on the basis of the caloric content of consumed food items.

$$Y_i = \begin{cases} 1, & HFS_i < 1950kcal/d/p(foodinsecure) \\ 2, & 2100kcal/d/p > HFS_i \geq 1950kcal/d/p(marginalyfoodsecure) \\ 3, & HFS_i \geq 2100kcal/d/p(foodsecure) \end{cases}$$

WhereY_i is food security status of the i^{th}household,$i = 1,2, \dots ,2015$

3.2.2. Explanatory Variables

Based on the reviewed literatures, some of the common predictors that are expected to influence rural household's food security in the study area could be categorized into Demographic and Socio-Economic variables.

Table3.1 Explanatory Variables

Variables	Description	Values
x_1	Age	Number
x_2	sex	1=Male 2=Female
x_3	Household size	1=≤ 4 2= >4
x_4	Marital Status	1=Never married 2=Married 3=other
x_5	Religion	1=Orthodox 2=Islam 3=Other
x_6	Educational of household	1=educated 2=not educated
x_7	Disabilities of a household	1=Yes 2=No
x_8	Income	1=Yes 2=No
x_9	Employment	1=employed 2=unemployed
x_{10}	Ecology	1=dega 2=weyna dega 3=kola

3.3. Logistic regression

Regression is a statistical procedure which attempts to predict the values of a given variable,(termed the dependent, outcome, or response variable) based on the values of one or more variables (called independent variables, predictors, or covariates). Regression analysis is model building for the relationship between a dependent and one and/or more independent variables. In the regression if the response variable is continuous we can use the usual linear regression model where as when the response variable is discrete, taking on two or more possible values the appropriate regression model is logistic regression which was proposed as alternative method in the late 1960s and early 1970s(Menard 2011). Such a technique was developed by McCullough and Nelder (1989) and is called generalized linear model (GLM), one of its application is logistic regression (Fox 1984). The problem of non-normality and hetroscadasticity lead to the model estimation method to be maximum likelihood after natural logarithm transformation of the odd ratio of the response because in logistic the relationship between the response with the set of explanatory variables is not linear hence the procedures used in the linear regression is extended to logistic regression. Logistic regression models are classified according to the type of categories of response variable as follows:-binary logistic regression model, multinomial logistic regression model and ordinal logistic regression models (Hosmer, Lemeshow et al. 2000). The binary logistic regression model is used to model the binary response variable, whereas the multinomial logistic regression is a simple extension of the binary logistic regression model where the response variable has more than two unordered categories. Ordinal logistic regression models are used to model the relationship between independent variables and an ordinal response variable when the response variable category has a natural ordering.

3.3.1. Binary Logistic Regression Model

Logistic regression analysis extends the techniques of multiple regression analysis in which the outcome variable is categorical. Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of predictor variables that may be continuous, discrete, dichotomous, or a mix of any of these (Gellman and Hill 2007).

Generally, when the dependent variable is dichotomous (such as presence or absence, success or failure and etc) binary logistic regression is used. The logistic regression is also preferred to multiple regression and discriminant analysis as it results in a meaningful interpretation, it is mathematically flexible and easily used distribution and it requires fewer assumptions (Hosmer, Jovanovic et al. 1989).

Unlike discriminant analysis, logistic regression does not have the requirements of the independent variables to be normally distributed, linearly related, nor equal variance with in each group (Cokluk 2010). Logistic regression has a peculiar property of easiness to estimate logit differences for data collected both retrospectively and prospectively (McCullagh 1983).

The two main uses of logistic regression are predicting the group membership, since logistic regression calculates the probability of success over the probability of failure, and providing knowledge of the relationships and strengths among the variables.

3.3.1.1. Model Description

Logistic regression model is used to investigate the effect of predictors on the probability of having diarrhea among under five children The response variable is dichotomous and denoted by $\mathrm{Y}_i, i = 1,2, \dots, n$ which is Bernoulli random variable with two possible values, $\mathrm{y}_i = 1$ with probability of having diarrhea $\mathrm{P}_i = \mathrm{P}(y_i = 1/\mathrm{X}_i)$ and $\mathrm{y}_i = 0$ with probability of having no diarrhea $1 - \mathrm{P}_i = 1 - \mathrm{P}(y_i = 1/\mathrm{X}_i)$.

The logistic model is defined as follows. Let $Y_{n\times 1}$ be a dichotomous outcome random variable with categories 1 (presence of diarrhea) and 0 (absence of diarrhea) in the two weeks prior to the survey. Let $X_{(n\times(k+1))}$ denote the collection of k-predicator variables of the response, where

$$\mathbf{X} = \underbrace{\begin{pmatrix} 1 & X_{11} & X_{12} & \dots. & X_{1k} \\ 1 & X_{21} & X_{22} & \dots. & X_{2k} \\ . & . & . & \dots. & . \\ . & . & . & \dots. & . \\ . & . & . & \dots. & . \\ 1 & X_{n1} & X_{n2} & \dots. & X_{nk} \end{pmatrix}}_{n\times(k+1)} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_n \end{bmatrix}$$

Where X is called regression matrix, and without the loading column of 1's, is termed as predictor data matrix. Then, the conditional probability that the i^{th} child has diarrhea given the vector of predictor variables $\boldsymbol{x_i}$ is denoted by $\mathrm{P}_i = \mathrm{P}(y_i = 1/\mathrm{X}_i)$. The expression P_i in logistic regression model can be expressed in the form of:

$$\mathrm{P}_i = \mathrm{P}(y_i = 1/\mathrm{X}_i) = \frac{\mathrm{e}^{\mathrm{X}_i\beta}}{1 + \mathrm{e}^{\mathrm{X}_i\beta}} \quad , i = 1,2, \dots, n$$

Where $\mathrm{P}(y_i = 1/\mathrm{X}_i)$ is the probability of i^{th} child having diarrhea given his/her individual characteristics $\boldsymbol{x_i}$ and $\beta = (\beta_0, \beta_1, \dots, \beta_\mathrm{k})^\mathrm{T}$ is a vector of unknown coefficients with dimension of $(\mathrm{k} + 1) \times 1$.

However, the relationship between the probability of i^{th} child having diarrhea and his/her characteristics are non linear. In order to make meaningful interpretation, the probability of i^{th} having diarrhea should to be written as linear combinations of predictors. This is computed using the logit transformation of the probability of i^{th}child having diarrhea which is given by:

$$logit[\mathrm{P_i}] = log\left(\frac{\mathrm{P_i}}{1-\mathrm{P_i}}\right) = \sum_{j=0}^{k} \beta_i \mathrm{X}_{ij} \ , i = 1,2, \dots n; j = 0,1, \dots , k$$

Where$\mathrm{X}_{i0} = (1,1, \dots ,1)^T$

3.3.2. Ordinal logistic regression

Ordinal logistic regression is an extension of binary logistic regression for analyzing ordinal response variable having more than two categories by considering the ordering of the response variable categories. For more than two categories of response we can build multinomial logistic regression model without considering the natural order of categories. Ordinal logistic regression is used to build a predictive model for ordinal response variable with a set of explanatory variables. It is applicable in biomedical research, epidemiological, biology etc. Ordinal logistic regression models with terms that reflect ordinal characteristics such as monotone trend have improved model parsimony and power. There are different types of ordinal logistic regression models, the most commonly used are: the adjacent-category, the continuation-ratio, the proportional odds models, the unconstrained partial-proportional odds model, the constrained partial-proportional odds model (Hosmer, Lemeshow et al. 2000)

3.3.2.1. Proportional odds model (ordered logit model)

The proportional odds model was originally proposed by Walker and Duncan (1967) as the constrained cumulative logit model and later called proportional odds model (Peterson and Harrell Jr 1990). Proportional Odds Model is used for modeling the response variable that has more than two levels with K set of explanatory variables by defining the cumulative probabilities, cumulative odds and cumulative $logit$ for the $J-1$ categories of the response, this model simultaneously use all cumulative $logits$. Let $j = 1,2, \dots , J$ are the ordinal categories of the response variable Y, and the vector of explanatory variable X, and denoted by vector form $X = (X_1, X_2, \dots , X_K)$ (Peterson and Harrell Jr 1990) For Y, the response with the J ordinal categories given that of K explanatory variables the individual probabilities are defined as follow;

$P(Y = j/X) = P_j$, for$j = 1,2, \dots , J$, and the cumulative probability can be defined as

$$\pi_j(X) = P(Y \leq j/X) = P_1 + P_2 + \cdots + P_j, for j = 1,2, \ldots, J-1 \quad (3.1)$$

$\pi_j(X)$, is the probability of being at or below category j, given that of K set of predictors. The odds of the cumulative probabilities of the response variable for the $J-1$ categories

$$Odds[\pi_j(X)] = \frac{\pi_j(X)}{1-\pi_j(X)}, j = 1,2, \ldots, J-1 \quad (3.2)$$

The logarithm of the odds first $J-1$ cumulative probabilities

$$ln(Odds[\pi_j(X)]) = ln\left(\frac{\pi_j(X)}{1-\pi_j(X)}\right), j = 1,2, \ldots, J-1 \quad (3.3)$$

The relationship between the response variable and the set of predictors is not linear in ordinal logistic regression model. The logistic regression function uses the logit transformation of $\pi_j(X)$ cumulative probabilities of the response,

$$\pi_j(X) = P(Y \leq j/X) = \frac{\exp(\alpha_j - (\beta_1 X_1 + \cdots + \beta_K X_K))}{1 + \exp(\alpha_j - (\beta_1 X_1 + \cdots + \beta_K X_K))}$$

$$ln\left(\frac{P(Y \leq j/X)}{1-P(Y \leq j/X)}\right) = ln\left(\frac{\pi_j(X)}{1-\pi_j(X)}\right) = \alpha_j - (\beta_1 X_1 + \cdots + \beta_K X_K)$$

Equivalent to:

$$logit[P(Y \leq j/X)] = \alpha_j - \sum_{K=1}^{K} \beta_K X_K, j = 1,2, \ldots, J-1 \quad (3.4)$$

Equation 3.4 is called the proportional odds model (POM) to predict cumulative $logits$ across $J-1$ response categories. This model estimates $ln(Odds)$ of being at or below the j^{th} category and assume that there is a linear relationship between the logits and the parallel regression lines and hence this model estimates simultaneously multiple equations of cumulative probability. The model is solved for each category of the dependent variable except the last category.

In the model each $logit$ has its own α_j term called the threshold value and their values do not depend on the values of the independent variables and the $\beta_K's$ are the logistic regression coefficients and the estimated values of these parameters show the direction and the strength of the relationship between the explanatory variables and the $logit$ ($logodd$) of the dependent variable. However, these regression coefficients interpretations are a little different from the usual regression coefficients and the interpretation for categorical explanatory variable is the effect (more likely and less likely) of the estimated category of the independent variables relative to the reference category on the log odds being in higher levels of the categories of the dependent variable. If the effect of each explanatory variable is the same in each $logit$ model then the model is called proportional odds model. In the POM, cumulative $logits$ are simultaneously modeled using the maximum likelihood estimation method. Prior to fitting a POM, it is important to check whether the assumption of proportionality is satisfied by each of the explanatory variables in the model.

Testing parallel lines

For fitting an ordinal logistic regression using the proportional odds model the assumption is that the relationship between independent variables and the $logits$ is the same for all the $logits$. That means this results are test of parallel lines or planes one for each category of the response outcome.

The test of parallel lines or planes has two log-likelihood functions; $-2log - likelihood$ for the model that assumes the lines or planes are parallel and $-2log - likelihood$ for the model that assumes the lines or the planes are separated.

For testing parallel lines for POM, the appropriate test statistic used is a chi-square statistic. This is the deference between the log-likelihood for the two models. A non significance test is evidence that the $logit$ surfaces are parallel and that the odds ratio can be interpreted as constant across all possible cut point of the response. The intercept term in the equations may vary, but the parameters would be identical for each model. If the lines or planes are parallel, the observed significance level for the change should be large, since the general model doesn't improve the fit very much.

If the proportional odds model is not fulfilled there are several options:

- Collapse two or more levels, particularly if some of the levels have small number of observations
- Do bivariate ordinal logistic analyses, to see if there is one particular independent variable that is operating differently at different levels of the dependent variable
- Use the partial proportional odds model
- Use multinomial logistic regression

3.3.2.2. Likelihood function and parameter estimation

In the model:

$$logit[P(Y \leq j/X)] = \alpha_j - \sum_{K=1}^{K} \beta_K X_K, j = 1,2, \dots , J - 1$$

The above model can use all $J - 1$ cumulative $logits$ in a single parsimonious model that means its model fit is not the same as fitting separate $logit$ models for each j. For estimating the parameters of the model define the binary indicator of the response variable for each observation or subject i. Therefore, the likelihood function is defined as follows:

$$l = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} \pi_j(X_i)^{y_{ij}}\right] = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} \pi_1(X_i)^{y_{i1}} \pi_2(X_i)^{y_{i2}} \times \dots . \times \pi_j(X_i)^{y_{ij}}\right] \quad (3.5)$$

Where, $y_{ij}'s$ the response variable indicators for fixed i and $j = 1, \dots , J$.

$$\pi_j(X_i) = P(Y \leq j/X_i) - P(Y \leq j - 1/X_i)$$

And the cumulative probabilities can be written as follows

$P(Y \leq j/X_i) = \frac{\exp(\alpha_j - \sum_{K=1}^{K} \beta_K X_{iK})}{1+\exp(\alpha_j - \sum_{K=1}^{K} \beta_K X_{iK})}$ And $P(Y \leq j - 1/X_i) = \frac{\exp(\alpha_{j-1} - \sum_{K=1}^{K} \beta_K X_{iK})}{1+\exp(\alpha_{j-1} - \sum_{K=1}^{K} \beta_K X_{iK})}$

Having these equations the likelihood becomes

$$l(\alpha, \beta) = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} [P\big((Y \leq j/X_i)\big) - P(Y \leq j - 1/X_i)]^{y_{ij}}\right]$$

$$l(\alpha, \beta) = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} [\frac{\exp(\alpha_j - \sum_{K=1}^{K} \beta_K X_{iK})}{1 + \exp(\alpha_j - \sum_{K=1}^{K} \beta_K X_{iK})} - \frac{\exp(\alpha_{j-1} - \sum_{K=1}^{K} \beta_K X_{iK})}{1 + \exp(\alpha_{j-1} - \sum_{K=1}^{K} \beta_K X_{iK})}]^{y_{ij}}\right]$$

$$l(\alpha, \beta) = \prod_{i=1}^{n} \left[\pi_1(X_i)^{y_{i1}} \pi_2(X_i)^{y_{i2}} \times \dots . \times \pi_j(X_i)^{y_{ij}}\right]$$

Therefore the log-likelihood function is:

$$l(\alpha, \beta) = log(l(\alpha, \beta))$$

$$log(l(\alpha, \beta)) = \log(\prod_{i=1}^{n} \left[\pi_1(X_i)^{y_{i1}} \pi_2(X_i)^{y_{i2}} \times \dots . \times \pi_j(X_i)^{y_{ij}}\right])$$

$=\prod_{i=1}^{n} \left[y_{i1} log\pi_1(X_i) + y_{i2} log\pi_2(X_i) + \dots + y_{iJ} log\pi_J(X_i)\right]$

Hence

$$l(\alpha,\beta)) = \prod_{i=1}^{n}\left[y_{i1}log\pi_1(X_i) + y_{i2}log\pi_2(X_i) + \cdots + y_{iJ}log\pi_J(X_i)\right] \tag{3.6}$$

In general, the method of maximum likelihood estimation produces values of the unknown parameters that best match the predicted and observed probability values. (McCullagh 1980) provided a Fisher scoring algorithm for ML fitting of all cumulative link models. Hence, it is often used as very effective method to obtain ML estimates for ordinal logistic regression parameters.

3.3.2.3. The Generalized ordered $logit$ model

In the case where the proportional odds assumption is violated, the proportionality constraint may be completely or partially relaxed for the set of explanatory variables. Generalized ordered logit model is an ordinal logistic regression which considers order of category of the response variable with k set of explanatory variables. This model results J-1 logits without constrained the effect of each explanatory variable is equal across the logits.

The model can be expressed as proposed by (Fu 1998) and(Williams 2006)as follows:

$$Logit[P(Y > j/X)] = ln\left(\frac{P(Y>j/X)}{P(Y\leq j/X)}\right) = \alpha_j + (\beta_{1j}X_1 + \beta_{2j}X_2 \ldots + \beta_{Kj}X_K),\ j = 1,2,\ldots,J\text{-}1 \tag{3.7}$$

where,α_j are the intercept or cut points and;$\beta_{1j}, \beta_{2j,}\beta_{kj}$ are $logit$ coefficients. This model estimates the odds of being beyond a certain category relative to being at or below that category.

A positive $logit$ coefficient indicates that an individual is more likely to be in a higher category as opposed to a lower category of the outcome variable. Generalized ordered $logit$ model estimates the regression parameters for each explanatory variable on $J-1\ logit$ of the probability being beyond the j^{th} category in every $logit$ to have different estimated values. Hence, this model has too many parameters and different interpretation to the K^{th} explanatory variable in the$J-1\ logit$. As discussed above the generalized ordered logit model that relaxes the proportionality assumption for all explanatory variables, which is less parsimonious model due to the above listed problems so, another model that allows some variables to have proportional across all $logits$ and the other variables to vary across $logits$ this model is called Partial proportional odds model.

3.3.2.4. Partial proportional odds model

The partial proportional odds model (Williams 2006)is a natural extension of the proportional odds model, which allows β's to vary across $logit$ equations. Suppose one set of predictors X_1 has p_1 parameters that satisfy the parallel line assumption or equal slope assumption and the remaining set of predictors X_2has p_2 parameters that do not satisfy parallel line assumption but they have unequal slopes and also depend on the j^{th} category of the response. PPOM is obtained by modifying Equation 3.7 and written as follow

$$Logit[P(Y > j/X)] = \alpha_j + \sum_{K=1}^{p_1} \beta_K X_{iK} + \sum_{r=1}^{p_2} \beta_{rj} X_{2r} \text{ , } j = 1,2,\ldots,J\text{-}1 \tag{3.8}$$

Equation 3.8 is PPOM; it should also be modified for the cases where, if the explanatory variables are categorical with more than two categories, some of the estimated categories may varies across the $J - 1$ $logits$ while other to be equal in such a case the proportionality is tested related to the categories of the explanatory variables. Generally speaking the generalized ordinal logistic regression model constrained for all explanatory variables estimates equal for $J - 1$ $logits$ called proportional odds model and partial proportional odds model is generalized ordinal logistic regression constrained for some of the variables to be equal across the $J - 1$ $logits$ and relaxed for the other which violate the parallel line assumption.

3.3.2.5. Test of overall model fit

Likelihood ratio test

After the model is selected the first step is to check whether a model fits the data well or not. The null hypothesis is their all the regression parameters are zero, and under the alternative hypothesis at least one regression coefficient (parameter) is not zero .To keep use of the selected mode the null hypothesis must be rejected and possibility for examining the significance for the individual parameters. In binary and ordinal logistic regression models the overall model fit can be based on the change in$-2\ log - likelihood$ when the variables are added to a model that contains only the intercept (Harrell 2015). The difference between the $-2\ log - likelihood$ for the model with only the intercept and the –2 log-likelihood for the selected model this difference follows chi-square distribution under the null hypothesis. Moreover models could be compared by the $-2\ loglikelihood$, a model which has small $-2LL$ are more preferred than for model that has large $-2LL$ value.

The likelihood-ratio test statistic is given by (Min and Agresti 2002)

$$G^2 = -2\ log\Lambda = -2(LL_0 - LL_1), G^2 \sim X^2{}_{p-(J-1)} \tag{3.10}$$

where, P and J are the number of parameter and number of category of the response variable respectively.

Where LL_0 and LL_1 are the maximized log-likelihood functions of the null model and the selected model respectively.

Pseudo R^2 measures

In the linear regression model, the coefficient of determination, $\boldsymbol{R^2}$, summarizes the proportion of variance in the dependent variable associated with the predictor (independent) variables, with larger $\boldsymbol{R^2}$ values indicating that more of the variation is explained by the model. For regression models with a categorical dependent variable, it is not possible to compute a single $\boldsymbol{R^2}$ statistic that has all of the characteristics of $\boldsymbol{R^2}$ in the linear regression model, so these approximations are computed instead. McFadden's pseudo R-squared statistic is based on the log likelihood for the model with predictors compared to the log likelihood for the model without predictors.

However, with categorical outcomes, it has a theoretical maximum value of less than one, even for a "perfect" model. (Smith and McKenna 2013):

$$\boldsymbol{R^2}_{MC} = \frac{LL_0 - LL_1}{LL_0} \tag{3.11}$$

where LL_0 and LL_1 are the maximized log-likelihood functions of the null model and the selected model respectively.

3.3.2.6. Test of a single predictors

Wald test

The Wald test is used to see the significance of a single explanatory variable in the model. The Wald test statistic is the square of the ratio of the estimated coefficient to its standard error and is defined as:

$$W = \left[\frac{\widehat{\beta_1}}{SE(\widehat{\beta_1})}\right]^2 \qquad € \qquad (3.12)$$

Under the null hypothesis $..H_0 : \beta_j = 0$, for $i = 1,2, \dots, k$ and W has a chi-square distribution with one degree of freedom.

3.3.2.7. Goodness-of-Fit Measures

As in linear regression, goodness of fit in logistic regression attempts to get at how well a model fits the data. It is usually applied after a "final model" has been selected. Much of the goodness of fit literature is based on the following hypothesis:

H_0: The model fit the data well Vs H_A: The model does not fit the data well

The measure of goodness of a fit is done by testing whether a model fits is to compare observed and expected values. From the observed and expected frequencies, we can compute the usual Pearson and Deviance goodness-of-fit measures. For a sample of n independent observations, the deviance and Pearson chi-square for a model with p degrees of freedom, both $\boldsymbol{X^2}$ and D has chisquare distribution with $(n - p)$ degrees of freedom.

The Pearson goodness-of-fit statistic is:

$$\mathrm{X}^2 = \sum\sum\left(\frac{\boldsymbol{O_{ij}-E_{ij}}}{\boldsymbol{E_{ij}}}\right)^{\boldsymbol{2}} \qquad (3.13)$$

The deviance measure is:

$$\mathrm{D} = 2\sum\sum \mathrm{O_{ij}} \ln\left(\frac{\mathrm{O_{ij}}}{\mathrm{E_{ij}}}\right) \qquad (3.14)$$

Where $\mathrm{O_{ij}}, \mathrm{E_{ij}}$ are the observed and expected frequencies from $\mathrm{i^{th}}$ row and $\mathrm{j^{th}}$ columns of the cross tabulation. The observed frequency is obtained from the data on the response but the expected frequency is obtained from the estimated probabilities of the response.

Both goodness-of-fit statistics should be used only for models that have reasonably large expected values in each cell. If we have a continuous independent variable or many categorical predictors or some predictors with many values, we may have many cells with small expected values. If our model fits well, the observed and expected cell counts will be similar, the value of each statistic will be small, and the observed significance level will be large. We shall reject the null hypothesis that the model fits the data well if the observed significance level for the goodness of- fit statistics is small. Good models have large observed p- values.

Hosmer-Lemeshow goodness of fit test

The recommended test for overall fit of a binary logistic regression model is the Hosmer Lemeshow test (Hosmer, Lemeshow et al. 2000). This test is preferred over classification tables when assessing model fit. The Hosmer-Lemeshow goodness of fit test divides subjects into deciles based on predicted probabilities, then computes a chi square from observed and expected frequencies. Then a probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model. If the $p-$ value of $H-L$ goodness-of-fit test statistic is greater than .05, as we want for well-fitting models, we do not reject the null hypothesis that there is no difference between observed and model predicted values, implying that the model's estimates fit the data at an acceptable level. Note that the number of groups, g, can be smaller than 10 if there are fewer than 10 patterns of explanatory variables. There must be at least three groups for the Hosmer-Lemeshow statistic to be computed. The Hosmer-Lemeshow goodness-of-fit statistic is obtained by calculating the Pearson chi-square statistic from the $2 \times g$ table of observed and expected frequencies, where g is the number of groups. The statistic is written

$$X^2{}_{HL} = \sum_{i=1}^{g} \frac{(O_i - N_i\overline{\pi_i})^2}{N_i\overline{\pi_i}(1-\overline{\pi_i})} \tag{3.15}$$

where • $N_i$is the total frequency of the subjects in the $\boldsymbol{i^{th}}$ group, O_iiG is the total frequency of the event outcomes in the ith group, and • $\overline{\pi_i}$ is the average estimated predicted probability of an event outcome for the ith group. Under the null hypothesis the $H-L$ test statistic has $X^2{}_{HL}$ distribution with $(g-2)$ degree of freedom. Large values of $X^2{}_{HL}$ (and small p-values) indicate lack of fit of the model.

3.3.2.8. Odds Ratio

In logistic regression the relationship between the response variable and the set of explanatory variables is not linear. Let the logistic probabilities from a model containing one dichotomous covariate coded 0 and 1, the odds of the response being present among individuals with x=1 and x=0 given below respectively (Hosmer, Lemeshow et al. 2000)

$$Odds(X=1) = \frac{P(Y/X=1)}{1-P(Y/X=1)} \text{and} Odds(X=0) = \frac{P(Y/X=0)}{1-P(Y/X=0)} \tag{3.9}$$

The $oddsratio$, denoted OR, is the ratio of the odds for x=1 to the odds for x=0, given as follow

$$OR = \frac{Odds(X=1)}{Odds(X=0)}$$

The odds of the response are multiplied by $OR = e^{\beta}$ for change from reference category to the estimated category of the given explanatory variable and odds less than one indicate the occurrence is less likely than non-occurrence and if the odds greater than one indicate the occurrence is more likely than non-occurrence.

3.3.2.9. Model adequacy checking

Model building is not the final goal in regression analysis. The model adequacy checking is the main step of regression analysis after a model fit. It can measure based on diagnosing residuals and measure of influence.

Residuals

Residuals are the difference between the observed and predicted value of the response variable Residuals are useful in identifying observations that are not explained well by the model. For logistic regression diagnostics the residuals are calculated in a similar way as usual. However, since the variables are categorical we have to consider contingency tables. The pattern of lack of fit revealed in cell-by-cell comparisons of observed and fitted (expected) counts may suggest a better model. For a model with categorical predictors, the residuals are computed from the observed and expected counts of the contingency table. Let Y_i denote the binomial variate for n_i trials at setting i of the explanatory variables, $i = 1, \dots, N$. Let $\bar{\pi}_i$ denote the model estimate of $p(Y = 1)$.Then$n_i\hat{\pi}_i$ is the fitted number of successes.

The Pearson residual is defined by (Min and Agresti 2002):

$$e_i = \frac{Y_i - n_i\hat{\pi}_i}{[\widehat{var}Y_i]^{1/2}} = \frac{Y_i - n_i\hat{\pi}_i}{\sqrt{[n_i\hat{\pi}_i(1-\hat{\pi}_i)]}} \qquad (3.16)$$

With$\hat{\pi}_i$ replaced by π_i in the numerator of the Pearson residual,e_i is the difference between a binomial random variable and its expectation, divided by its estimated standard deviation; for large $n_i \geq 30$ e_i has an approximate $N\ (0, 1)$ distribution. Since π_i is estimated by $\hat{\pi}_i$and ‘ $\hat{\pi}_i$ depend on Y_i, The Pearson residuals do not have unit variance since no allowance has been made for the inherent variation in the fitted value. A better procedure is to further adjust the Pearson residuals by their estimated standard deviation that contains variation due to the effect leverage value is called standardized Pearson residual.

The Standardized Pearson residual is slightly larger in absolute value then e_i, and is proximately $N\,(0,1)$ when the model holds. It's similar to the Pearson residual the only difference is standardized residuals uses the leverage from an estimated hat matrix that means for an observation i with leverage value $\hat{h}_i$. Observations with absolute standardized residual values in excess of 3 may indicate lack of fit (Scokaert and Rawlings 1998). The standardized Pearson residual is given (Min and Agresti 2002):

$$r_i = \frac{e_i}{\sqrt{1-\hat{h}_i}} = \frac{Y_i - n_i\hat{\pi}_i}{\sqrt{[n_i\hat{\pi}_i(1-\hat{\pi}_i)(1-\hat{h}_i)]}} \tag{3.17}$$

Deviance residuals are used to check for lack of fit by considering the $\boldsymbol{i^{th}}$observation. Logistic regression is a type of generalized linear model, if the model fits poorly based on the overall goodness-of-fit test, examination of residuals highlights where the fit is poor. This residual uses the components of the deviance statistic. The deviance residual for observation $\boldsymbol{i}$ is defined as:

$$\sqrt{d_i} \times sign(Y_i - n_i\hat{\pi}_i \tag{3.18}$$

Where

$$d_i = 2\left(Y_i log\frac{Y_i}{n_i\hat{\pi}_i}\right) + (n_i - Y_i)log\frac{n_i - Y_i}{n_i - n_i\hat{\pi}_i}$$

The deviance residual can have negative sign when $n_i\hat{\pi}_i$exceeds Y_i and negative sign, if Y_i exceeds$n_i\hat{\pi}_i$, Observations with absolute deviance residual values greater than 3 may indicate lack of fit (Scokaert and Rawlings 1998), each squared deviance residual is a component ofD^2, deviance statistic test for goodness of fit is given by

$$D^2 = \sum_{i=1}^{N} d_i{}^2 \text{ ‰}$$

Measure of influence

An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism. 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 (Belsley 1980). Diagnostics are certain quantities computed from the data with the purpose of pinpointing influential points after which these influential points can be removed or corrected. The standard logistic regression model, we should check for the effect of individual observations on model estimates and fit. We are interested in identifying subjects with high leverage, large residuals, or a large degree of influence on the model estimates.

In linear regression the diagonal elements of the hat matrix are called the leverage values and are proportional to the distance of X_i to the mean of the data$\bar{X}$. Similarly for logistic regression leverage values are the diagonal element of the hat matrix. These values show as the distance between individual observations to the mean, if the distance is large or as individual observation are far from the mean it may have considerable influence on the values of the estimated parameters.

Pregibon (1981) derived a linear approximation to the fitted values which yields a hat matrix for logistic regression. This matrix is

$$H = V^{1/2} X (X'VX)^{-1} X'V^{1/2}$$

Where V is $n \times n$ diagonal matrix with general diagonal element

$$V_i = m_i \hat{\pi}(x_i)[1 - \hat{\pi}(x_i)]$$

Leverage values for logistic regression are the diagonal elements of the hat matrix and denoted by h_i is given below (Hosmer, Lemeshow et al. 2000).

$$h_i = m_i \hat{\pi}(x_i)[1 - \hat{\pi}(x_i)] x_i'(x_i'VX)^{-1} x_i \qquad (3.19)$$

Where $\widehat{V_1} = m_i \hat{\pi}(x_i)[1 - \hat{\pi}(x_i)]$

And $x_i' = (1, x_{1i}, x_{2i}, \dots, x_{pi})$ is the vector of the covariate values defining the i^{th} covariate pattern.

The hat matrix for the logistic regression as a $n \times n$ matrix the diagonal element is bounded from the above by$1/m_i$, where m_i is the total number of subject with the same covariate pattern. When the hat matrix is based upon data grouped by covariate pattern, the upper bound for any diagonal element is one that means the centered leverage values ranges from 0 to $(n-1)/n$ and the leverage value greater than one for the i^{th} observation indicates that observation is influential (Belsley 1980).

Another useful diagnostic statistic is one that examines the effect that deleting all subjects with a particular covariate pattern has on the value of the estimated coefficients and the overall summary measures of fit $X^2$and D^2 . The change in the value of the estimated coefficients is analogous to the measure proposed by Cook (1977, 1979) for linear regression (Hosmer and

Lemeshow, 2000). It is obtained as the standardized difference between the estimated coefficient with i^{th} observation and without the i^{th} observation, where this represents the maximum likelihood estimates computed using all i covariate patterns and excluding the m_i subjects with pattern x_i respectively, and standardizing via the estimated covariance matrix of the estimators.

The Analog Cook's influence statistic for logistic regression is given as follow $\Delta\widehat{\beta_1} = \left(\hat{\beta} - \widehat{\beta_{-1}}\right)'(X'VX)\left(\hat{\beta} - \widehat{\beta_{-1}}\right)$.

Computationally, the i^{th} Cook's distance, CD_i, is more easily obtained as:

$$CD_i = \frac{r^2{}_i h_i}{(1-h_i)} \tag{3.20}$$

Where r_i is the standardized residual and h_i is the i^t diagonal element of H matrix computed from the full logistic regression with K explanatory variables.

Cook's distance is the difference between the estimated coefficient with the i^{th} observation and after deleting the i^{th} observation. This is based on the squared value of standardized Pearson residual and leverage value. If Cook's distance is large for i^{th} observation it is considered to be influential.The suggested cut off values for i^{th} observation to be influential such as outlier, if the CD_i is greater than "one" (CD_i>1) (Hosmer and Lemeshow, 2000, Rawlings, 1998).

DFBETA(S) is a diagnostic that measure the effect of the i^{th} observation on the estimates of the logistic regression coefficients. These are computed by dropping the i^{th} observation. 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 i^{th} observation greater than 1.0, implies the observation is an outlier (Cook and Weisberg 1982)and the formula DFBETA is the change of the coefficient estimates(i^{th} explanatory variable) from the deletion of a case i. It is computed

$$\text{DFBETA}_{K_i} = \frac{(X'VX)^{-1}X_i'e_i}{1-h_i} \tag{3.21}$$

Where e_i and h_i are the Pearson residual and leverage value respectively.

3.3.3. Multilevel logistic regression model

Reflecting the usefulness of multilevel analysis and the importance of categorical outcomes in many areas of research, generalization of multilevel models for categorical outcomes has been an active area of statistical research. For dichotomous response data, several approaches adopting either a logistic or probit regression model or various methods for incorporating and estimating the influence of the random effects have been developed (Hutcheson and Sofroniou 1999).The developments have been mainly in terms of logistic and probit regression models. Because the proportional odds model, which is based on the logistic regression formulation, is a common choice for analysis of ordinal data, many of the multilevel models for ordinal data are generalizations of this model (Hedeker, Demirtas et al. 2009).

A multilevel logistic regression model also referred as a hierarchical model in the literature. It can account for lack of independence across levels of nested data. Hierarchical models are statistical models that can be used to analyze nested sources of variability in hierarchical data, taking account of the variability associated with each level of the hierarchy. These models have also been referred to as multilevel models, mixed models, random coefficient models, and covariance component models (Khan and Shaw 2011).

4. STATISTICAL DATA ANALYSIS RESULT

4.1. Summary of descriptive statistics

The purpose of this chapter is to analyze the effect of different socio-economic and demographic determinants of household food security statues in Amhara region the data from the 2015/16 house hold consumption and expendicher survey by CSA.

Table 4.1 descriptive statistics of response variable

Food security	Freq.	Percent
food insecure	829	41.14
Marginally food secure	274	13.60
food secure	912	45.26
Total	2,015	100.00

From the above descriptive summery statistics out of the total sample 45.26% is food secure, 13.60 is marginally food secure and the remaining 41.14% are food insecure based on use of kilocalorie per person per day.

Table 4.2 descriptive statistics of explanatory variable and its cross tabulation response variable

variables	Categories	Food security statues							
		Food insecure		Marginal food secure		Food secure		total	
		Count	%	count	%	count	%	count	%
Religion	Orthodox	660	41.16	190	11.93	742	46.61	1592	79.01
	Islam	168	40	84	20	168	40	420	20.84
	Other	1	33.33	0	0	2	66.67	3	0.05
Disability	Yes	42	42.86	11	11.22	45	45.92	98	4.86
	No	787	41.05	263	13.72	867	45.27	1917	95.14
marital status	Nevermerrid	411	53.59	99	12.91	257	33.51	767	38.06
	Marrid	309	35.56	125	14.38	435	50.06	869	43.13
	other	109	61.07	50	28.18	220	10.74	379	18.81
Income	Yes	528	45.13	163	13.93	479	40.94	1,170	58.06
	No	301	35.62	111	13.14	433	51.24	845	41.94
Ecology	Dega	231	39.9	71	12.26	277	47.84	579	28.73
	weyna dega	448	40.47	158	14.27	501	45.26	1,107	54.94
	Kola	150	45.59	45	13.68	134	40.73	329	16.33
sex of household head	Male	643	40.75	209	13.2	726	46.01	1,578	78.31
	Female	186	42.56	65	14.87	186	42.56	437	21.69
household Educational level	Educated	262	29.54	101	11.39	524	59.08	887	44.02
	Noteducated	567	50.27	173	15.34	388	34.40	1,128	55.98
Size of household	<=4	637	41.35	258	16.83	638	41.62	1,533	76.08
	>4	192	39.83	16	3.32	274	56.85	482	23.92
Age of households		829	41.14	274	13.60	912	45.26	2015	100
Employment	Employed	732	41.64	243	13.82	783	44.84	1,758	87.25
	unemployed	97	37.34	31	12.06	129	50.19	257	12.75

From table 4.2 the descriptive summery statistics out of the total sample 79.01 % are orthodox, 20.84% are islam and the remaining 0.15% are other religion based on use of kilocalorie per person per day.

In the same way out of the total sample 78.31 % house hold head are male and the remaining 21.69% house hold head are female based on use of kilocalorie per person per day.

Table 4.2 reveals that house hold food security differs by educational attainment of head of household. For instance, 29.54% of educated house hold are food insecure, 11.39% are marginally food secure and the remaining 59.08%are food secure. On the other hand 50.27 % not educated house hold are are food insecure, 15.34%are marginally food secure and the remaining 34.40%are food secure. From this we show than non educated house hold are more food insecure than educated house hold.

Sex of house hold is also important factor of food security. In the above table house hold head is male more food secure than house hold head is female. It shows that 46.01% of male is food secure and 40.75% food insecure. On the other hand 34.4% and 42.56% female are food secure and food insecure respectively.

In the other hand ecology an important factor for determinant of food security. The output shows that 47.84% house hold in dega, 45.26% in weyna dega and 40.73% in kola are food secure in the same way 12.26% house hold in dega, 14.27% house hold in weyna dega and 13.68% house hold in kola are marginally food secure in the other hand 39.9% house hold in dega 40.47% house hold in weyna dega and 45.59% house hold in kola are food in secure.

Marital states also factors for determinant of food security. The output shows that 53.59%, 12.91% and 33.51% house hold of never married are food insecure, marginally food secure and food secure respectively. In the same way 35.56%, 14.38% and 50.06% house hold of married are food insecure, marginally food secure and food secure respectively.

4.2. INFERENTIAL STATISTICS RESULT

4.2.1. Binary logistic regression

The result of the binary logistic regression model is presented in table A1. Food security was assigned a value of 1 if the respondents reported food insecure and 2 otherwise. The reference category of each dichotomously measured independent variable has a value of one and the values for other categories are compared to that of the reference category. A value less than one implies that individuals in that category have a lower probability of food secure than individuals in the reference category. The Wald Chi-Square statistic, which tests the unique contribution of each predictor, holding the other predictors constant, that is, eliminating any overlap between predictors. Each predictor (except currently working) must meet the conventional .05 standard for statistical significance.

4.2.2. Ordinal logistic regression

Ordinal logistic regression is an appropriate model for a response variable with more than two categories (ordinal) these model is simply an extension of binary logistic regression (only two category). This model is based on the estimation of log (odds) cumulative probability for the response which has a linear relationship to the set of explanatory variables. Proportional odds model is a set of logit model estimated simultaneously by assuming the effects of explanatory variables equal in all logits.

Univariate analysis

The variables in this study are house hold food security as the response and education, religion, disability, age, employment, house hold size, ecology, sex of head, income and marital status are the explanatory variables that related to food security on different literatures. Before building the logistic regression model for analyzing the categorical data we first checked the association of each explanatory variable with response using Pearson chi-square test. Consequently, it was found that the explanatory variables education, religion, ecology, age, employment, house hold size, income and marital status are significantly associated at 25% level of significance (see Table A2). Hence, all these explanatory variables except disability and sex of household head will be entered into the proportional odds model (Hosmer, Lemeshow et al. 2000) since all the explanatory variables are significantly associated with house hold food security.

One of the assumptions underlying ordinal logistic regression is that the relationship between each pair of outcome groups is the same. This is commonly referred to as the test of parallel lines because the null hypothesis states that the slope coefficients in the model are the same across response categories (and lines of the same slope are parallel). If we fail to reject the null hypothesis, we conclude that the assumption holds.

Table 4.3 Test of parallel lines

	Chi2	df	P>Chi2
Wolfe Gould	157.7	10	0.000
Brant	146.6	10	0.000
Score	172.2	10	0.000
likelihood ratio	132.4	10	0.000
Wald	215.8	10	0.000

From table 4.3 shows parallel line test for general model with chi square value 146.6 and p-value=0.000 it shows that the assumptions of the parallel-lines model are violated. Due to the violation of proportional odds assumption gologit2 can overcome this limitation by fitting partial proportional odds models.

4.2.2.1. Result of partial proportional odds model (PPOM)

This model can be fitted using the GOLOGIT2 with AUTOFIT option of STATA user written command (Williams, 2006). Using AUTOFIT option to estimate a model in which some variables are constrained to meet the parallel lines assumption while others are not. In simple words PPOM is a model that relaxes the constraints for those variables that violate the assumption of POM. Based on the above PPOM with AUTOFIT option by a series of Wald tests are also used to check the assumption of proportionality for all categories of each explanatory variable and finally test all the categories of the explanatory variables that pass the Wald test using global wald test with degrees of freedom equal to the number of parameters that pass the assumption of proportional odds model.

When the model is fitted using STATA 12 for categorical explanatory variables the first category of each explanatory variables are considered as reference category. Results of the fitted PPOM are given in Table 4.4. The categories, religion orthodox, marital status never married, income yes, employment employer, ecology dega, age <=19 years education educated and HHsize<=4 were used as reference categories.

Table 4.4 PPOM model parameter estimates

variables	food insecure			Marginally food secure		
	Coef.	P>z	Odds Ratio	Coef.	P>z	Odds Ratio
Religion						
Islam	0.086	0.486	1.090	-0.257	**0.039**	0.774
Other	-13.729	0.992	0.000	12.923	0.993	409791.700
Marital states						
Married	1.366	**0.000**	3.920	1.235	**0.000**	3.439
Other	2.000	**0.000**	7.388	1.302	**0.000**	3.676
Income						
No	0.759	**0.000**	2.137	0.742	**0.000**	2.101
Ecology						
weyna dega	-0.103	0.364	0.902	-0.148	0.183	0.862
Kola	-0.304	**0.042**	0.738	-0.347	**0.021**	0.707
HH Education level						
not educated	-1.081	**0.000**	0.339	-1.134	**0.000**	0.322
Size of HH						
>4	-0.201	0.125	0.818	1.029	**0.000**	2.798
Employment						
Unemployed	-0.434	**0.014**	0.648	-1.111	**0.000**	0.329
age_head	-0.004	0.181	0.996	-0.001	0.775	0.999
_cons	-0.181	0.537	0.835	-0.648	0.029	0.523

A global Wald test is then performed for the final model with constrained versus the original unconstrained model. The test indicates that the final model does not violate the parallel lines assumption. As the global Wald test shows, 14 constraints have been imposed in the final model, the chi2 (8) = 5.54, with P = 0.6990 which not a significant value indicating that the final model does not violate the proportional odds or parallel lines assumption.

Interpretation of partial proportional odds model

The results in Table 4.4 regarding the partial proportional odds model provide estimated coefficients, standard errors and p-values of the explanatory variables categories. The coefficients of the explanatory variables in the model are interpreted as the log odds of the response variable being in higher categories as opposed to the lower categories. In logistic regression the interpretation of the model estimates are based on odds ratios and their confidence interval (see Table 4.4). On the basis of Table 4.4 the interpretations are given as follow.

In this study religion is a significant determinant of food security in Amhara region, when marginally food secure is compared to food secure showing that Islam had 23% less risk of being food secure compared with Orthodox.

In this study marital status is a significant determinant of food security in Amhara region, married were 3.920 times more likely of being in the marginally food secure or food secure (as opposed to food insecure) compared to never married. Married were 3.439 times more likely of being in the food secure (as opposed to food insecure and marginally food secure) compared to never married.

Others (divorced and Windowed) household were 7.388 times more likely of being in the marginally food secure or food secure (as opposed to food insecure) compared with never married house hold. Others (divorced and Windowed) household were 6.37 times more likely of being in the food secure (as opposed to food insecure and marginally food secure).

Income is a significant determinant of food security in Amhara region, no income house hold were 2.137 times more likely of being in the marginally food secure or food secure (as opposed to food insecure) compared with there is income. No income households were 2.101 times more likely of being in the food secure (as opposed to food insecure and marginally food secure) compared with there is income.

In this study employment is a significant determinant of food security in amhara region, when food insecure is compared to marginally food secure and food secure showing that households who are unemployed had 35% less risk of being marginally food secure or food secure compared with households who are employed. Household who are marginally food secure compared to food secure showing that households who are unemployed had 67% less risk of being food secure.

Education is a significant determinant of food security, when food insecure is compared to food secures showing that no educated had 66% less risk of being food secure (as opposed to food insecure) compared with educated. Marginally food secure is compared to food secure showing that no educated had 67% less risk of being food secure (as opposed to food insecure or marginally food secure) compared with educated.

House hold size is a significant determinant of food security, households size is greater than four were 2.798 times more likely of being in the food secure (as opposed to food insecure and marginally food secure) compared to households size is less or equal to four.

4.2.2.2. Model adequacy checking

Model adequacy checking includes diagnosing residuals and measures of influence. This is difficult to do in ordinal and multinomial logistic models. In order to reduce the difficulty, the ordinal response variable categories can be changed to binary categories by collapsing two or more categories. Then a binary logistic regression model is fitted after which it is possible to apply model adequacy checking in this study the response has three categories. By collapsing the two categories into one including food secure and marginally food secure these can be called food secure. The other category will be food insecure. Therefore, the diagnostics performed in binary logistic regression model is the same for the partial proportional odds model (ordinal logistic regression). We could calculate residuals, measures of influence and the predicted probabilities of the data. The plots of standardized Pearson residuals, deviance residuals, DFBETA, Cook's distance, leverage value with predicted probability can then be used to see the pattern of all cases using the software SPSS version 20. The residuals and measure of influence plots against the predictive probabilities revealed that the model is adequate.

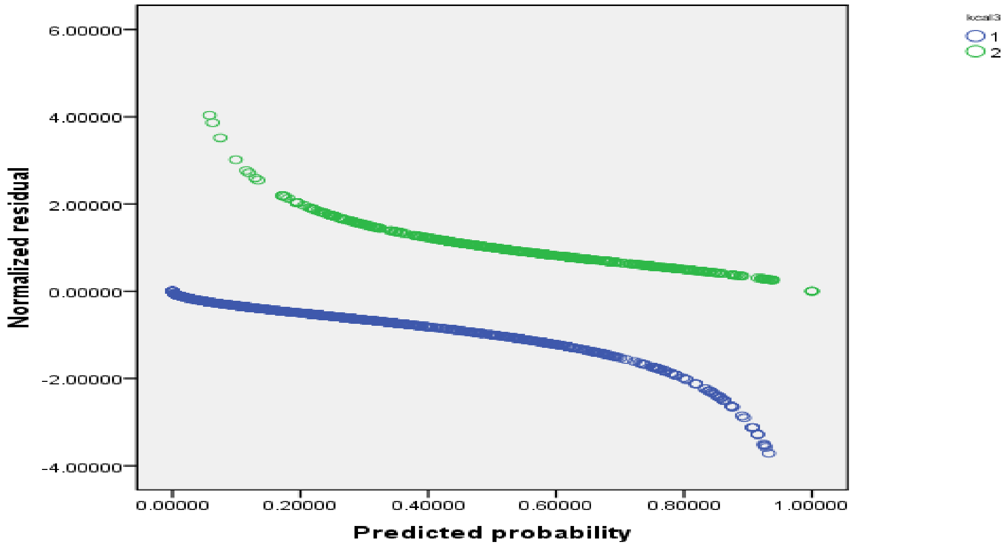

Figure 4. 1Plots of standard residual vs predicted probability.

Figure 4.1 is the plot of standard residuals vs predicted probabilities of all observations. There are few observations far from the others. However, the computed standard residuals do not influencing the model that means all standard residuals are less than three (see from Y- axis).

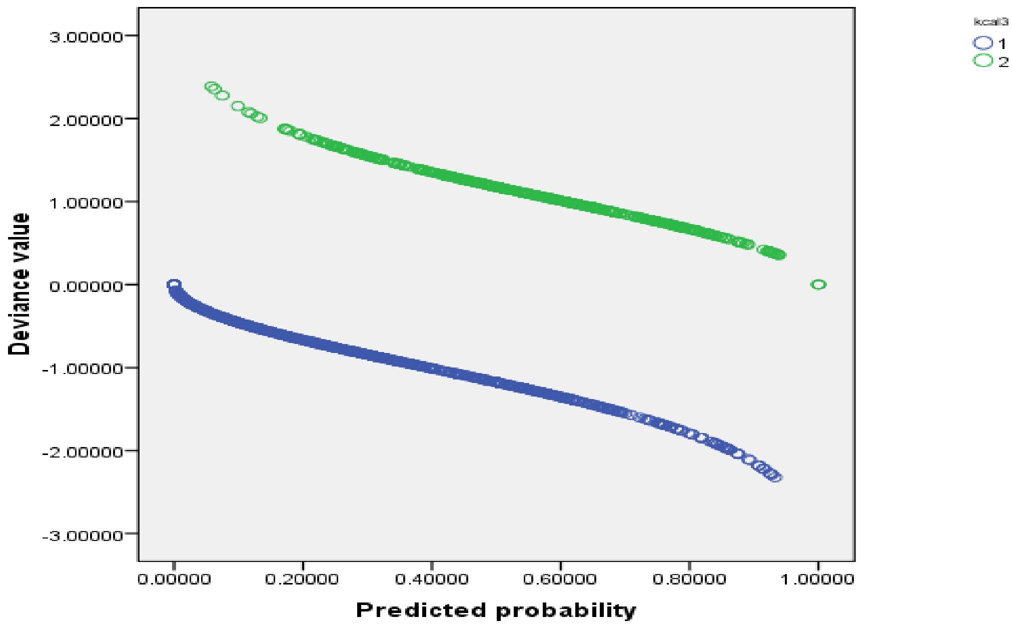

Figure 4. 2 Plots of deviance residual vs predicted probability.

Figure 4.2 above is the two plots of deviance residuals vs predicted probabilities of all observations. Apparently there are few observations that lie far away from the rest but all absolute deviance residuals are less than three. Therefore, there is no lack of fit.

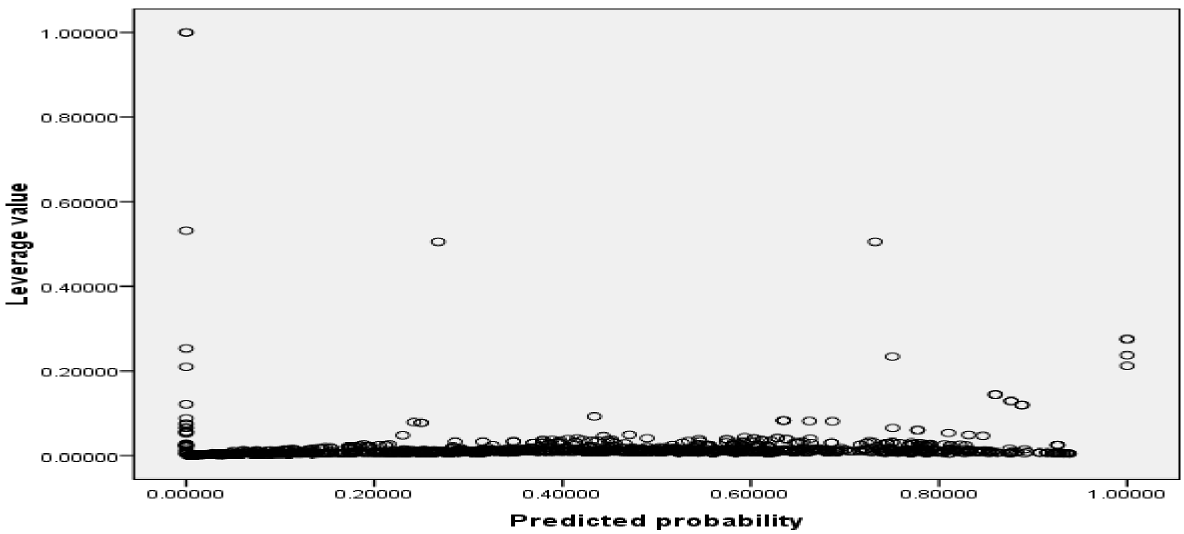

Figure 4. 3 Plots of leverage value vs predicted probability.

Figure 4.3 the plots of leverage value vs the predicted probabilities of all observations. It was observed that leverage values of the above plots are less than one. Therefore, there are no outliers.

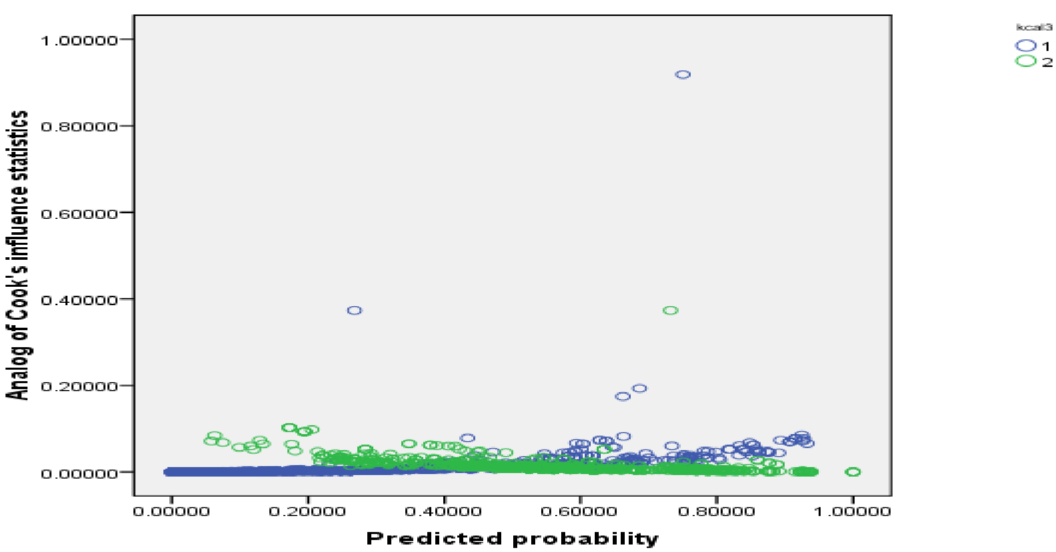

Figure 4. 4 Plots of Analog of Cook’s influence statistic vs predicted probabilities

Figure 4.4 is the plot of Analog of Cook’s influence statistic vs the predicted probabilities of all observations. There are observations a little far away from the others. These are not influential observations since all Cook’s influence statistic are less than one. (See on Y-axis of the graph).

5. DISCUSSION, CONCLUSION AND RECOMMENDATION

5.1. Discussion of the Results

This study carried out to identify the determinate of Food security in rural house hold of Amhara region based on (HCES) dataset taken from (CSA). Consequently, partial proportional odds model were used to identify the model important significant variables that affect the food security in rural house hold of Amhara region.

Size of household: the result showed that the family size of the household is statistically significant at 5% probability level. This negative relationship indicates that odds ratio in favor of the probability of being food secure decreases as family size increases. If all other things are held constant, the odds ratio of 0.818 for size of house hold implies that, the odds ratio in favor of being food secure decreased by a factor of 0.818 as family size increase by one person or one adult equivalent. The result indicated that larger household size tends to be food insecure compared to smaller family size. The result is consistence with the research finding by (Dagne 2016) and (Anyanwu 2014). And oppose to research finding of (Welderufael 2014)

Education: the result showed that the education of the household is statistically significant at 5% probability level. This negative relationship indicates that odds ratio in favor of the probability of being food secure decreases as households are not educated. If all other things are held constant, the odds ratio of 0.339 for education implies that, the odds ratio in favor of being food secure decreased by a factor of 0.339. The result indicated that not educated tends to be food insecure compared to households is educated. The result is consistence with the research finding by (Dagne 2016) and (Bogale and Shimelis 2009).

Employment: the result showed that employment of the household is statistically significant at 5% probability level. This negative relationship indicates that odds ratio in favor of the probability of being food secure decreases as households are unemployed. If all other things are held constant, the odds ratio of .5859 for employment implies that, the odds ratio in favor of being food secure decreased by a factor of .5859. The result indicated that unemployed tends to be food insecure compared to households is employed. The result is consistence with the research finding by (Sultana and Kiani 2011).

Ecology: the result showed that ecology of the household is statistically significant at 5% probability level. This negative relationship indicates that odds ratio in favor of the probability of being food secure decreases as households are live in weyna dega. If all other things are held constant, the odds ratio of .7786 for ecology implies that, the odds ratio in favor of being food secure decreased by a factor of .7786. The result indicated that live in weyna dega tends to be food insecure compared to households is live in dega. This result is consistent with the results obtained by (Asmelash 2014)

Religion: This study showed that religion of the household is statistically significant at 5% probability level, consistent with the results obtained by (Dagne 2016) who found that religion is a significant predictor of food security. The study results obtained by (Bogale and Shimelis 2009) conclude the opposite.

Marital status: the result showed that marital status of households is statistically significant at 5% probability level. This disagrees with the result obtained by (Welderufael 2014) which concluded the contrary. However, our finding is similar to (Leza and Kuma 2015) who showed that Marital status is significant effect of food security.

Income: the result showed that income of households is statistically significant at 5% probability level.

Generally the variables like marital status, education, household size, religion, income and ecology are significant effect of food security. In the same way variables age of household head, sex of household head and disability are not significant in this study.

5.2. CONCLUSION

The major objective of this study was to identify the determinants of food security among the rural households of the Amhara National Regional State. As a result, this study found that household food security in the study region was determined by six key factors. However, this is not a complete study to come up with solid solution to address the food security situation in the region under this study. This is because the range of factors and elements that affect food security are complex and multifaceted in nature and not easy to comprehend. Therefore, effort has been made in this study to see the impact of some demographic and socioeconomic factors on household food security.

In the study region 41.14%, 13.60% and 45.26% of the households were found food insecure, marginally food secure and food secure, respectively. The figures show that the proportion of food secure households is higher than the food insecure households higher than marginally food secure in the year during which the data was collected. However, this result might have been changed if the data had been collected in another year.

The food security related factors studied through the partial proportional odds model analysis revealed that factors such as marital status, education, household size, religion, income and ecology were found the major contributors to the discrimination of the three group households and to the prediction of actual group membership of a household.

5.3. Recommendations

As rural part of Amhara region is constantly facing food insecurity and famines, there is a need for integrating famine relief and prevention strategies at the regional level with the overall development strategy. The strategy should aim at self-sufficiency at the local level and food security at the household level by incorporating the following recommendations.

- ✓ It should be noted that household size is known to be one of the leading causes of food insecurity in the study area. This implies that policy measures directed towards the provision of better family planning to reduce household size should be given adequate attention and priority by the federal and regional governments. Education that encompasses all aspects of training and which brings about attitudinal changes targeting at reducing fertility level is important for rural households in the study area.
- ✓ Based on the study, households with there are income is better in food security status than households with no income in the study region. Therefore, it is recommended that the regional and federal governments should provide access to work and getting own income. In the short term, government should be create opportunity getting their own income for rural house hold of in the study area.
- ✓ Finally, I recommend for further studies to be conducted on the area of food security by considering detail and accurate information on various variables including political, climatic and weather (rainfall and temperature), topography, natural disasters, ecological conditions and other factors that affect food security. It is also recommended to conduct a study that compares status of food security in rural households with urban households. And it is best to use multilevel regression model based on the nature of the data.

.

Reference

Adenew, B. (2004). "The food security role of agriculture in Ethiopia." eJADE: electronic Journal of Agricultural and Development Economics **1**(853-2016-56101): 138-153.

Adgo, E., et al. (2014). "Impact of land certification on sustainable land resource management in the Amhara region, Ethiopia." Retrieved September **21**: 2015.

ALEMU, E. and E. Ayalew (2016). Enset value chain analyis: the case of angecha woreda in kembata tembaro zone, south Ethiopia.

Anyanwu, J. C. (2014). "Marital status, household size and poverty in Nigeria: evidence from the 2009/2010 survey data." African Development Review **26**(1): 118-137.

Aragie, T. and S. Genanu (2017). "Level and Determinants of Food Security in North Wollo Zone (Amhara Region—Ethiopia)." Journal of Food Security **5**(6): 232-247.

Asmelash, M. (2014). "Rural household food security status and its determinants: The case of Laelaymychew Woreda, Central Zone of Tigrai, Ethiopia." Journal of Agricultural Extension and Rural Development **6**(5): 162-167.

Ayelign, A. and S. De Saeger (2020). "Mycotoxins in Ethiopia: Current status, implications to food safety and mitigation strategies." Food Control **113**: 107163.

Barrett, C. B., et al. (2013). Hunger and food insecurity, Oxford University Press Oxford.

Bedeke, S. B. (2012). "Food insecurity and copping strategies: a perspective from Kersa district, East Hararghe Ethiopia." Food Science and Quality Management **5**(3): 19-27.

Belsley, D. A. (1980). "On the efficient computation of the nonlinear full-information maximum-likelihood estimator." Journal of Econometrics **14**(2): 203-225.

Benson, T., et al. (2008). Global food crises: Monitoring and assessing impact to inform policy responses, Intl Food Policy Res Inst.

Bogale, A. and B. Korf (2009). Analysis of poverty and its covariates among smallholder farmers in the eastern Hararghe highlands of Ethiopia.

Bogale, A. and A. Shimelis (2009). "Household level determinants of food insecurity in rural areas of Dire Dawa, Eastern Ethiopia." African Journal of Food, Agriculture, Nutrition and Development **9**(9).

Boon, E. K. (2007). "Food security in Africa: challenges and prospects." Regional Sustainable Development Review, Africa.

Boussard, J.-M., et al. (2005). "Food security and agricultural development in Sub-saharan africa: building a case for more support. Background document. Final report."

Clover, J. (2003). "Food security in sub-Saharan Africa." African security review **12**(1): 5-15.

Cokluk, O. (2010). "Logistic Regression: Concept and Application." Educational Sciences: Theory and Practice **10**(3): 1397-1407.

Cook, R. D. and S. Weisberg (1982). Residuals and influence in regression, New York: Chapman and Hall.

Dagne, G. (2016). "Determinants of food security in farm household in drought prone area of Oromia Region: In case of Dodota District." Journal of Economics and sustainable development **7**: 2222-2855.

Degefa, T., et al. (2006). "Uterine involution in the post-partum Balady goat." Veterinarski arhiv **76**(2): 119-133.

Desalegn, G. and S. N. Ali (2018). "Review of the Impact of Productive Safety Net Program (PSNP) on Rural Welfare in Ethiopia."

Devereux, S., et al. (2007). "An Evaluation of Concern Worldwide's Dowa Emergency Cash Transfer Project (DECT) in Malawi, 2006/07." Lilongwe: Concern Worldwide.

Devereux, S. and I. Sussex (2000). Food insecurity in Ethiopia, Institute for Development Studies.

Endris, F. S. and K. T. Nura (2018). "Determinants of Household Food Insecurity in Kemisse Special Zone: Case Study at Dawachefa Wereda, Ethiopia." International Journal of Multicultural and Multireligious Understanding **5**(3): 192-202.

FIKIRE, Z. and A. P. Bekele (2014). Determinants of food security in the rural households of meskan woreda, gurage zone, snnpr, ethiopia, St. Mary's University.

Fisseha, S. (2017). Food Insecurity As A Disaster Risk Among Rural Households Of Zuway Dugda Woreda, Arsi Zone, Ethiopia.

Fox, S. R. (1984). The mirror makers: A history of American advertising and its creators, University of Illinois Press.

Fu, W. J. (1998). "Penalized regressions: the bridge versus the lasso." Journal of computational and graphical statistics **7**(3): 397-416.

Gellman, A. and J. Hill (2007). Data analysis using regression and multilevel/hierarchical models, Cambridge: Cambridge University Press.

GEMEDA, M. (2020). DETERMINANTS OF HOUSEHOLD FOOD INSECURITY IN MERTI WOREDA OF ARSI ZONE OROMIA REGION, St. Mary's University.

Haile, H., et al. (2005). Causes of household food insecurity in koredegaga peasant association, Oromiya zone, Ethiopia.

Harrell, F. E. (2015). Ordinal logistic regression. Regression modeling strategies, Springer**:** 311-325.

Hedeker, D., et al. (2009). "A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data." Statistics and its Interface **2**(4): 391.

Hindle, R. E. (1990). "The World Bank approach to food security analysis." IDS Bulletin **21**(3): 62-66.

Hoddinott, J. (1999). "Operationalizing household food security in development projects: an introduction." Technical guide **1**: 1-19.

Hosmer, D. W., et al. (1989). "Best subsets logistic regression." Biometrics: 1265-1270.

Hosmer, D. W., et al. (2000). Applied logistic regression, Wiley New York.

Hutcheson, G. D. and N. Sofroniou (1999). The multivariate social scientist: Introductory statistics using generalized linear models, Sage.

Kalkuhl, M., et al. (2013). Conceptual framework on price volatility and its impact on food and nutrition security in the short term.

Karuhanga, M. (2008). Living with AIDS in Uganda: impacts on banana-farming households in two districts.

Kenneth, O. (2008). Impact of contracted sugarcane farming on household food security in Uriri Division-Kenya.

Khan, H. R. and E. Shaw (2011). "Multilevel logistic regression analysis applied to binary contraceptive prevalence data." Journal of Data Science **9**: 93-110.

Kidane, H., et al. (2005). "Causes of household food insecurity in Koredegaga peasant association, Oromiya zone, Ethiopia." Agrekon **44**(4): 543-560.

Leroy, J. L., et al. (2015). "Measuring the food access dimension of food security: a critical review and mapping of indicators." Food and nutrition bulletin **36**(2): 167-195.

Leza, T. and B. Kuma (2015). "Determinants of rural farm household food security in Boloso Sore District of Wolaita Zone in Ethiopia." Asian Journal of Agricultural Extension, Economics & Sociology: 57-68.

Mannaf, M. and M. T. Uddin (2012). "Socioeconomic factors influencing food security status of miaze growing households in selected areas of Bogra district." Bangladesh Journal of Agricultural Economics **35**(454-2016-36355): 177-187.

Maxwell, D., et al. (1999). "Alternative food-security indicators: revisiting the frequency and severity ofcoping strategies'." Food Policy **24**(4): 411-429.

Maxwell, S. (1994). "Food security: a post-modern perspective."

Maxwell, S. and M. Smith (1992). "Household food security: a conceptual review." Household food security: Concepts, indicators, measurements **1**: 1-72.

McCullagh, P. (1980). "Regression models for ordinal data." Journal of the Royal Statistical Society: Series B (Methodological) **42**(2): 109-127.

McCullagh, P. (1983). "Quasi-likelihood functions." The Annals of Statistics **11**(1): 59-67.

Melese, S. M. (2016). "Effect of land use land cover changes on the forest resources of Ethiopia." International Journal of Natural Resource Ecology and Management **1**(2): 51.

Menard, S. (2011). "Standards for standardized logistic regression coefficients." Social Forces **89**(4): 1409-1428.

Min, Y. and A. Agresti (2002). "Modeling nonnegative data with clumping at zero: a survey."

Mitiku, A., et al. (2012). "Empirical analysis of the determinants of rural household’s food security in Southern Ethiopia: The case of Shashemene District." Basic Research Journal of Agricultural Science and Review **1**(6): 132-138.

Mohamed, A. A. (2017). "Food security situation in Ethiopia: A review study." International Journal of Health Economics and Policy **2**(3): 86-96.

Momsen, J. (2008). Women and development in the third world, Routledge.

Mwangi, M. and S. Kariuki (2015). "Factors determining adoption of new agricultural technology by smallholder farmers in developing countries." Journal of Economics and sustainable development **6**(5).

Mwaniki, A. (2006). "Achieving food security in Africa: Challenges and issues." UN Office of the Special Advisor on Africa (OSAA).

Ogbu, J. U. (1973). "Seasonal hunger in tropical Africa as a cultural phenomenon. The Onicha Ibo of Nigeria and Chakaka Poka of Malawi examples." Africa: 317-332.

Oluoko-Odingo, A., et al. (2018). "Food Security: Remodelling the Equations for Future African Agricultural Sustainability." Tackling Sustainable Development in Africa and Asia: Perspectives from Next Generation Researchers: 156.

Peterson, B. and F. E. Harrell Jr (1990). "Partial proportional odds models for ordinal response variables." Journal of the Royal Statistical Society: Series C (Applied Statistics) **39**(2): 205-217.

Ranganathan, J., et al. (2016). "Shifting diets for a sustainable food future." World Resources Institute: 831878447.1548317259-1387216062.1543582872.

Sabila, S. C. (2014). Factors that influence food security in rural households of mount Elgon sub county, Kenya, University of Nairobi.

Scokaert, P. O. and J. B. Rawlings (1998). "Constrained linear quadratic regulation." IEEE Transactions on automatic control **43**(8): 1163-1169.

Smith, T. J. and C. M. McKenna (2013). "A comparison of logistic regression pseudo R2 indices." Multiple Linear Regression Viewpoints **39**(2): 17-26.

Sultana, A. and A. Kiani (2011). "Determinants of food security at household level in Pakistan." African journal of business Management **5**(34): 12972-12979.

Tolossa, D. (2001). "Causes of Seasonal Food Insecurity in Oromiya Zone of Amhara Region: Farmers' View."

Toulmin, S. and B. Gustavsen (1996). Beyond theory: Changing organizations through participation, John Benjamins Publishing.

Welderufael, M. (2014). "Determinants of households vulnerability to food insecurity in Ethiopia: econometric analysis of rural and urban households." Journal of Economics and Sustainable Development **5**(24): 70-79.

Williams, R. (2006). "Generalized ordered logit/partial proportional odds models for ordinal dependent variables." The Stata Journal **6**(1): 58-82.

APPENDIX

Table A1: Result of the binary logistic regression mode

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	religion			4.353	2	.113			
	religion(1)	.324	1.807	.032	1	.858	1.383	.040	47.759
	religion(2)	.071	1.810	.002	1	.969	1.073	.031	37.256
	desibility(1)	-.121	.228	.284	1	.594	.886	.566	1.385
	maritalstates			129.564	2	.000			
	maritalstates(1)	-1.527	.149	105.095	1	.000	.217	.162	.291
	maritalstates(2)	-.464	.139	11.159	1	.001	.629	.479	.826
	income(1)	-.521	.104	25.233	1	.000	.594	.485	.728
	ecology			6.623	2	.036			
	ecology(1)	.366	.153	5.741	1	.017	1.442	1.069	1.944
	ecology(2)	.145	.139	1.081	1	.299	1.156	.880	1.518
	sex_head(1)	.202	.121	2.777	1	.096	1.224	.965	1.551
	edu(1)	1.238	.102	147.999	1	.000	3.450	2.826	4.212
	HHsize(1)	-.842	.122	47.283	1	.000	.431	.339	.548
	age_head	-.002	.003	.392	1	.531	.998	.992	1.004
	Constant	.440	1.820	.059	1	.809	1.553		

a. Variable(s) entered on step 1: religion, desibility, maritalstates, income, ecology, sex_head, edu, HHsize, age_head.

Table A2: Chi- square test of association between explanatory variables with response variable

Explanatory variables	Pearson chi-square (P-value)
religion	0.002
desibility	0.775
maritalstates	0.000
income	0.000
employment	0.000
Ecology	0.249
HHsize	0.000
edu	0.000
sex_head	0.399
age	0.000

Table A3: Result of Ordered logistic regression

Ordered logistic regression Log likelihood =-1839.0512				Number of obs =2,015 LR chi2(13) =333.80 Prob > chi2 =0.0000 Pseudo R2 =0.0832			
food security	Coef.	Std. Err.	z	OR	P>z	[95% Conf. Interval]	
religion							
islam	-0.091	0.110	-0.820	0.913	0.411	-0.306	0.125
other	-0.193	1.702	-0.110	0.825	0.910	-3.529	3.143
disability							
no	0.292	0.213	1.370	1.339	0.170	-0.125	0.709
marital states							
married	1.172	0.112	10.470	3.229	0.000	0.953	1.392
other	1.592	0.140	11.390	4.911	0.000	1.318	1.866
income							
no	0.753	0.104	7.260	2.123	0.000	0.550	0.956
ecology							
weyna dega	-0.152	0.104	-1.470	0.859	0.142	-0.355	0.051
kola	-0.339	0.139	-2.430	0.713	0.015	-0.612	-0.066
sex_head							
female	-0.087	0.110	-0.790	0.917	0.428	-0.302	0.128
edu							
not educated	-1.166	0.095	-12.310	0.312	0.000	-1.351	-0.980
HHsize							
>4	0.446	0.116	3.850	1.563	0.000	0.219	0.674
employmeent							
unemployed	-0.783	0.156	-5.030	0.457	0.000	-1.088	-0.478
age_head	-0.002	0.003	-0.810	0.998	0.418	-0.008	0.003
/cut1	0.041	0.272		0.041		-0.493	0.574
/cut2	0.676	0.273		0.676		0.142	1.210