

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF STATISTICS

Statistical Analysis on the Determinants of Income Inequality in Ethiopia

BY:

NAME	ID NO
1. TESFAYE WUBESHET	NCSR/334/09
2. FENET JIFARA	NCSR /132/09

ADVISOR: MOHAMMED D. (MSc)

A Senior research paper Submitted to Statistics Department in Partial Fulfillment of the Bsc degree in Statistics, Wolkite University

June, 2019

Wolkite, Ethiopia

ACKNOWLEDGEMENT

First of all, we would like to express our deepest and warmest gratitude to the God, who help us from the beginning to the end of this research. Secondly, we are also greatly interested to express hearted our thanks to our advisor Mohammed D. (MSC.); for his great assisting, contribution, useful suggestion and constructive criticism on this excellence advising and limitless effort in encouraging us in our work, correcting and giving comments by devoting a deal of his time from the beginning to the end of this research. In addition to this, we would like to thanks Wolkite University, Department of statistics. Lastly, we would like to express our heartfelt and appreciation to our friends for their psychological, moral supports during research development.

ACKNOWLEDGEMENT ii
ABSTRACT v
ACRONYMS vi
List of table vii
CHAPTER ONE 1
Introduction 1
1.1. Background of the study 1
1.2. Statement of the problem 4
1.3. Objectives of the study 5
1.3.1. General objective 5
1.3.2. Specific objectives 5
1.4. Significance of the study 5
CHAPTER TWO 6
LITEATURE REVIEW 6
CHAPTER THREE 9
DATA AND METHODOLOGY 9
3.1. Study Area 9
3.2. Study Population 9
3.2.1. Data Source and Method of Data Collection 9
3.3. Variables of the study 9
3.3.1. Dependent variable: 9
3.3.2. Independent variable 9
3.4. Method of data analysis 10
3.4.1 .Chi-square test 11
3.5. Logistic regression model 12
3.5.1. Ordinal logistic regression 12
3.5.2. The Model of Ordinal Logistic regression 12
3.5.3. Assumptions of ordinal logistic regressions 13
3.6 Model Building Using Logistic Regression 13
3.6.1 Parameter Estimation 13
3.6.2. Model adequacy checking 14
3.6.2.1. The Wald Statistic 14
3.6.2.2 Parallel Lines test Assumption 15
3.6.2.3. Goodness of Fit of the Model 15
CHAPTER FOUR 16

4. DATA ANALYSIS AND RESULT 16
4.1 Descriptive statistics 16
4.2. Ordinal Logistic Regression 20
4.2.1. Interpretation for coefficients of significant variables 24
4.3 Discussion 26
CHAPTER FIVE 26
CONCLUSION AND RECOMENDATION 27
5.1 CONCLUSION 27
5.2. Recommendations 28
REFERENCES 29

ABSTRACT

Income inequality is mainly concerned with the relative position of different individuals within the income distribution. The objective of this study was to identify the determinant factors of income inequality in Ethiopia. In this study the data source was the Ethiopian Demographic and Health Survey conducted in 2016 (EDHS 2016) from central statistical agency. In the data analysis method we applied ordinal logistic regression model. From chi-square test of association age, sex of house hold head, number of children, number of hectares of agriculture, Main floor material, has electricity, sources of drinking water, own live stock for farm, place of residence, types of toilet facility, locations of source for water, numbers of own cattle, Locations of toilet facility, Time to get water source (minute) had a significant association with house hold income. In this study the result of ordinal logistic regression model indicates that the selected variables place of residence, main floor material, sex of house hold , has electricity , own live stock for farm , age, source of drinking water ,numbers of children , were found to have significance effect on house holed income level. The government should be the services of electricity broadly distributed to rural, encouraging the female participation , the expansion of urbanization for the areas of rural as much as possible, formulates suitable policy in family planning for reducing the number of children in order to decrease income inequality.

Key word: house hold income, ordinal logistic regression, Ethiopia

ACRONYMS

UNICEF	united nation children fund
EDHS	Ethiopian Demographic Health Survey
MOFED	Ministry of Finance and Economic Development
SNNP	South nations, nationalities and people
CSA	Central Statistical agency.
SPSS	Statistical Package for Social Science

List of table

Table 4.1: cross tabulation of house hold income level with explanatory variable 16

Table 4.2 Model Fitting Information 20

TABLE 4.3 Goodness-of-Fit 20

Table 4.4 Pseudo R- Square 21

TABLE 4.5 Test of Parallel Line 21

Table 4.6 Results of the ordinal logistic regression model 22

CHAPTER ONE

1. Introduction

1.1. Background of the study

Income inequality is mainly concerned with the relative position of different individuals within the income distribution. Inequality can be considered as a case of different people having different degrees of income or consumption. Income distribution can be observed at a personal level or a functional level. Where the functional distribution of income considers the distribution between groups in society who own different factors of production, i.e. the proportion of income going to employees, landowners, and owners of capital respectively. On the other hand, the personal income distribution is concerned with the national distribution of income without paying too much attention to the factors of production. These studies were focus more on the personal distribution of income. A number of techniques to measure inequality in a population have notably been developed and employed over time such as the Gini coefficient, the coefficient variation of income, the logarithm of income and generalized entropy class of inequality indices, the Gini coefficient and the Atkinson index (Zelda *et al.*, 2012)

The measures of inequality consider the whole of the population. As for measures of poverty, as well as inequality can be calculated from an indicator such as income or expenditure, (Boccanfuso D., and Kabore T.S (2000). That is the poverty measures depend on the average level of income or consumption in a country, and the distribution of income or consumption. Based on these two elements, poverty measures then focus on the situation of those individuals or households at the bottom of the distribution. It focuses only on those whose standard of living falls below an appropriate threshold level (such as a poverty line). This threshold may be set in absolute terms (based on an externally determined norm, such as calorie requirements) or in relative terms. (Boccanfuso D., and Kabore T.S ,2000))

Absolute poverty (income less than $1 per day) in sub-Saharan Africa increased from 288 million in 1981 to 516 million in 2001, increasing from 42 % to 47 %which is about13% of the world's total poor population. In 2006, 34 of the 50 nations on the UN list of least developed countries are in Sub-Sahara Africa and a more so bering statistic is that about 14.6 million children (simply, one in every five) live in absolute poverty as at 2007 data.

Poverty is the economic condition in which people lack sufficient income to obtain certain minimal levels of health services, food, housing, clothing and education which are necessities for standard of living". The various definitions/measures of poverty lead to two perspectives which are "income poverty" and "lack of basic need poverty". Income poverty occurs when an individual does not have enough money to meet up with the a certain standard of living while lack of basic need poverty occurs when one is unable to meet some of the basic needs such as food, shelter and clothing as identified by United Nations, Children's Fund (World Bank ,2009).

There is also a growing of concern that income inequality (at least) within and between countries has been increasing and why academic, public, policy makers and development agencies are highly concerned with inequality. According to McKay, A. (2002), the reasons are; inequality matters of poverty, growth, individuals own rights, behind crimes and social unrest and inequality is critical in attainment of Millennium Development Goals (MDGs). Nevertheless, the available literature on urban poverty and income inequality in Ethiopia has been limited, reflecting the lack of an appropriate and reliable household survey data that would allow the comparison of welfare across a time.(McKay, A. ,2002))

To explain, inequality can take different forms. Some societies may be divided between a rich elite and everyone else. Others may have relatively small numbers of rich and poor and a large middle class. These variations may determine the relationship between inequality and growth for two main reasons. First, inequality may affect how different income groups behave. Second, it may affect how different social groups interact. Some examples: If inequality affects how income groups behave that is the poor: If there are large numbers of poor people, economic growth may be affected by their inability to invest in education and their lower health levels, among other factors. The middle classes: If inequality "squeezes" the middle class, it may reduce its demand for goods and services. The rich: If inequality means rising incomes among the rich, it could see them accumulate savings, which banks can then lend out, so increasing investment in the economy. Or, the rich may use their economic power to lobby against policies that don't serve their needs, for example investment in public health and education (Keeley, Brian, 2015).

Inequality can have many dimensions. Economists are concerned specifically with the economics or monetarily measurable dimension related to individual or household income and consumption. However, this is just one perspective and inequality can be linked to inequality in skills, education, opportunities, happiness, health, life expectancy, welfare, assets and social mobility. The effects of inequality in non-income factors on earnings can be summarized variously. Inequality in education explains a minor fraction of differences in cross-country earnings inequality. The impact decreases by the level of education and depends on the economic development and skill-intensive nature of production technologies. It also negatively affects the Investment rate and growth rate of income. There is no direct link from income inequality to ill Health measured as mortality, but a range of mechanism and social arrangements indicate the Presence of an indirect link. Unlike in the case of income inequality, within country health inequality is a dominating source of inequality. The non-income dimension of inequality is beyond the scope of this paper. Heshmati (2004) reviews the recent advances in the measurement of inequality and gives attention to the interrelationship between income inequality and the non-income inequality dimensions. (Heshmati ,2004).

1.2. Statement of the problem

In Ethiopia, income growth reduces poverty and inequality increase; the income-poverty elasticity lies in the range of -1.7 to -2.2. Growth occurred in urban areas but the rise in inequality in urban areas wiped out the poverty-reducing effect that this growth might boast (Tassew et al., 2008).

There are improvements in the poverty situation of Ethiopia from time to time yet the income inequality, as measured by Gini coefficient (used to measure dispersions of income inequality), has increased despite the continuous growth of the Ethiopian economy for more than five years in double digits, the poverty situation in the country increased; people below poverty line in 1995/06 was 33.5 percent which was increased to 36.9 in 1999/2000, 35.1 percent in 2004/5 and lastly reached 38.7 percent in 2008(MoFED,2006 and 2008).

The basic questions to be addressed in this study are:

- ✓ What is the association between income inequality and its explanatory variables?
- ✓ What are the factors that affect income inequality in Ethiopia?
- ✓ What are the frequencies of income inequality in Ethiopia?

1.3. Objectives of the study

1.3.1. General objective

The main objective of this study was to identify the factors that affect income inequality in Ethiopia.

1.3.2. Specific objectives

The specific objectives of our study were;

- ✓ To asses the frequency of income inequality in Ethiopia.
- ✓ To explore the association between income inequality and Explanatory variables
- ✓ To identify the factors that affect income inequality in Ethiopia.

1.4. Significance of the study

This study was important to know factors that affect income inequality against in Ethiopia. In order to give appropriate measures for those factors and to improve the prevalence of income inequality. This study also would have the purpose to increase the income equality about their principles. This study estimating the prevalence and identify the key risk factors of income inequality in Ethiopia. The result of the study would be providing some important strategies or procedure to solve the problem that can affect income inequality.

Limitation of the study

- There is no sufficient source done by similar model for our study.
- There are no sufficient internet accesses.

CHAPTER TWO

2. LITEATURE REVIEW

Regarding to Ethiopia, Little. P.G. Sanchez, A. and Lndrerch, R.C. (World Bank, 2009 Pp.2-3) analyzed the disappointing performance poverty reduction in Ethiopia was accompanied by a surge in urban inequality, with the Gini Coefficient increasing by 1 percentage points in urban areas from 0.34 in 1995 to 0.44 in 2004. In rural areas the coefficient remained stable at around 0.27. The MOFED report estimated that without this adverse distributional shift urban poverty would have been reduced by 12.6 percentage points, but that the positive impact of growth on poverty reduction was muted by 14.6 percentage point increase in the headcount due to distributional factors.

In the same way, Tassew W, Hoddinott, J and Dercon S. (2009), in their poverty and inequality analysis in Ethiopia, found that while inequality remained unchanged in rural areas, there was a substantial increase in urban inequality. In Ethiopia, income growth reduces poverty and increases in inequality increase poverty; the income-poverty elasticity lies in the range of -1.7 to -2.2. In rural Ethiopia, the increase in consumption has led to a reduction in headcount poverty.

Daniel *et al.* (2015) study the relationship between economic growth and income inequality using Single and multiple linear regression models across a country were formed using data from 2010. This study found a positive relationship between economic growth and income inequality. Found that unemployment, natural resource, marital statuses, gross saving .Statistical inference tests supported the all variable are significance but unemployment are not significant.

Almas and biwei(2013) study on the determinants of income and urban-rural income gap to shed light on the problem of urban-rural income inequality in China using OLS, conditional quintile Regression and Blinder-Oaxaca decomposition methods. Found that level of education, gender, age, marital status, type of occupation. From these levels of education and type of occupation are essential determinants of households' income level.

Skare and Stjepanovic (2014) examined the determinants of income distribution and Income inequality in 200 countries. They found that many macroeconomic variables Like exports, unemployment, labor force and population have strong impact on income distribution regardless of difference in their income level.

Odedokun and Round (2004) investigated the relationship between growth and inequality as well as determinants of income inequality in African countries. They Found that share of labor force in agricultural sector, regional factors, level of economic development, size of government budget as well as land and human resource are main determinants of income inequality. They also found the negative relation between inequality and growth.

A study conducted by Mduduzi in 2017 on determine the factors that influence poverty and household welfare in South Africa using ordinary least squares and probit/logit models found that of education of the household head, some province dummies, race of the household, dependency ratio, gender of the household, employment status of the household and marital status of the household are statistically significant determinants of household welfare.

A study conducted by Bereket and markos in 2018 Identifying major factors affecting poverty status of a society is important to decide what action should be taken to alleviate the poverty and assess the factors that affect the poverty status of rural Residence in Gamo Gofa zone, Southern Regional State of Ethiopia .applies binary logistic regression model found that Saving culture, access to credit, resource base, land fertility, use of agricultural inputs, use of improved tools, availability of rain, land topography, labor availability and dependency attitude have significant association with the poverty status of a households.

A study conducted by Araya in 2010on the incidence of poverty and its determinants and income inequality in Wukro town, Ethiopia using logistic regression model found that productive members, access to electricity, marital status, telephone subscription, educational level, family size and sex of the household. The incidence of poverty allied with this income inequality calls for urgent interventions aimed at curbing the problems of the people through creating employment opportunities, family planning, provision of market based short term trainings, empowering females and distribution of social service.

Gabrielet .al in 2018 analyzes the incidence and determinants of poverty and income inequality in pastoral and agro-pastoral communities in Afar Regional State, Ethiopia. Applies logistic regression model found that education level membership in local institutions, distance to market area and remittances l, Gender of the household head, family size, access to credit &mobility from these Gender of the household head, family size, access to credit & mobility is affecting poverty.

A study conducted by Abe be et .al in 2016 on the determinants income inequality among sampled households who find themselves at the bottom and top of the income/consumption distribution in urban centers in South Wollo Administrative Zone, Ethiopia using quantile regression analysis found that family size, household head main employment status or income sources, quality of houses, household energy sources, durable goods/assets, water and sanitation and place of residence are the main determinants of expenditure/ income inequality of per adult equivalent consumption expenditure across all quantiles distribution. Whereas, the household years of schooling and housing occupancy are the main determinants of expenditure/ income inequality at the bottom and higher quantiles distribution of per adult equivalent consumption expenditure.

CHAPTER THREE

3. DATA AND METHODOLOGY

3.1. Study Area

Ethiopia is one of the African countries which were located in the eastern part of Africa around the place called horn of Africa. This place located in the north eastern extension of the continent. The country lies west of Somalia, north of Kenya, east of Sudan, south of Eretria and Djibouti. The country has 9 region (Oromiya, Amhara, Tigray, Affar, Somali, Ben-Gumuz, SNNPR, Gambela, and Harari) and two city administration that is (Dire-Dawa and Addis-Ababa).

3.2. Study Population

The target populations of this study would be those house hold incomes who were affected by explanatory variable from Ethiopian Demographic and Health Survey (EDHS 2016) data and the case area under consideration of Ethiopia.

3.2.1. Data Source and Method of Data Collection

Basically we have two main categories of data source mainly primary source of data and secondary source of data**.** Depending on the source, Primary method, consist of obtaining data or information by any one of the following ways: direct personal interview, in direct oral interview, information from correspondents, mailed questionnaire and so on .In case of our study was conducted by using secondary source of data .The data for this study would be obtained from the Ethiopian Central Statistical authority (CSA)'collected with every five interval and our data is Ethiopian demographic and health survey of 2016.

3.3. Variables of the study

3.3.1. Dependent variable:

The response variable for this study was household income level (1= poor level of household,2= medium level of house hold, 3= rich levels of house hold)

3.3.2. Independent variables

Explanatory variables are variable that affects the response variable. These are continuous variables that is age , number of children, hectare of land for agriculture, time to get water source in (minute), number of own cattle. The categorical are listed below

Table 3.1 Categorical variable coding

variable	categories
Place of residence	0=rural,1=urban
Source of drinking water(SDW)	borehole=0,protectedwell=1,unprotectedwell=2, protectedspring=3,unprotectedspring=4,river=5,rain water=6,tankertrunk=7,cartwithsmalltank=8,piped water=9.
Types of toilet facility	Flush to pit latrine=0,flush to some were =1,ventilater improved pit latrine=2,flush donot know were=3, pit latrine with slab=4,pit latrine without slab=5,compositing toilet=6,other =7,flush toilet=8
Has electricity	yes=0,no =1
Own live stock for farm land	yes=0,no=1
Has telephone	yes=0,no=1
Sex of house hold head	0=female, 1=male
Main floor material	Dung=0,wood plank=1,bamboo=2,asphalt strip=3,ceramic =4,cement=5,carpet=6,earth=7
Locations of toilet facility	In own dwelling=0,in own yard/plot=1,else where=2
Locations of source of water	In own dwelling=0,in own yard /plot=1,else where=2

3.4. Method of data analysis

By using spss .16 both descriptive and inferential statistics was used. The study use tables from descriptive statistics. Inferential statistics the study use chi-square test for the purpose of association checking between variables and logistic regression model under consideration.

3.4.1 .Chi-square test

Chi-square(x^2) is a statistical measure with the help of which it is possible to access the significance of the different between the observed frequencies and expected frequencies obtained from some hypothetical universe. It is a measure of the different between the observed and expected counts. In addition it helps us to determine whether two variables are associated or not. In order to that chi-square test may be applicable both the frequencies must be grouped in the same way and the theoretical distribution must adjust to give the same total frequency, which is equal to that of observed frequency.

The Assumption of Chi-square Test

- ✓ The observation must be independent of each other.
- ✓ The sample must be randomly selected from the population.
- ✓ The expected frequency of each category must be at least 5.

The Statistical Test:

$$X^2_{cal}=\sum_{j=1}^{r}\sum_{i=1}^{c}\frac{(Oij-Eij)2}{Eij}$$

The degree of freedom associated with contingency table possessing r-row and c-columns=, X^2(r-1) (c-1) the test x^2_{Cal} with $x^2_{\alpha/2}$(c-1)* (r-1)

Where, O_{ij}=is the observed class frequency

E_{ij}=is expected cell frequency, r= number of row variables, c = number of column variable

Decisions: - if $x^2_{Cal}>x^2_{tab}$ then reject H_o.

If $x^2_{Cal}<x^2_{tab}$ then fail to reject H_o.

P-value is the smallest level of the test for which the null hypothesis (H_o) is rejected

That is when p-value greater than the significance level, Ho is not rejected.

Test of Hypothesis

H_0: There is the association between the response variable and explanatory variable Response and explanatory variable are independent.

H_1: There is no association between the response variable and explanatory variable.

Decision Rule: Reject the null hypothesis and not reject the alternative hypothesis if p-value is less than α-value

Conclusion: based on the decision**.**

3.5. Logistic regression model

Logistic regression is a special case of generalized linear models in which the mean of the response variable is related to explanatory variables through a regression equation. The elements of such a model are a distribution for the response variable and a function that links the distribution to the explanatory variables called a link function. Logistic regression can be binary, multinomial and ordinal logistic regression. The response variable is usually dichotomous for the response taken as success and failure. Logistic regression model in a single explanatory variable x for binary response variable(y =1, as probability of success and y = 0, as probability of failure) and when is probability of success at value x. The binary logistic regression is a type of regression which used to when the dependent variables dichotomous and the independent variables are any type.

3.5.1. Ordinal logistic regression

Ordinal logistic regression is a type logistic regression analysis that when the response variable is categorized in to more than two categories with having natural order or rank. Ordinal logistic regression is similar with multinomial logistic regression, but in multinomial there is no order which it has meaningful in ordinal regression. For our study we used ordered logistic regression, because of the response variable income is categorized into three categories, which is poor, medium and rich. So the category can be ordered, that is rich greater than medium and medium greater than poor.

3.5.2. The Model of Ordinal Logistic regression

Logit $[P(yi\leq j)]=\log[P(yi\leq j)/P(yi>j)]=\alpha j+\beta x, j=1,2,\ldots\ldots J-1.$

There is a very clear and intentional ordering to these responses like as poor medium and rich , If we know that a category is ordinal then there are special models that tell us how independent variables relate to someone being higher or lower on the scale..

Different links lead to proportional odds models or ordered probit models. The model cannot be consistently estimated using ordinary least squares; it is usually estimated using maximum likelihood. Typical response functions that were modeled are cumulative logits, adjacent-category logits, or continuation-ratio logits resulted in ordinal logistic models known as the cumulative logit model, the adjacent-category logit model and the continuation-ratio logit model. The proportional odds model and the partial proportional odds model were special cases of the cumulative logit model.

3.5.3. Assumptions of ordinal logistic regressions

Dependent variable should be measured at the ordinal level, Independent variables are continuous, ordinal or categorical (including dichotomous variables). However, ordinal independent variables must be treated as being either continuous or categorical; there is no multi co linearity. Multi co linearity occurs when two or more independent variables that were highly correlated with each other. We had proportional odds, which was a fundamental assumption of this type of ordinal regression model.

3.6 Model Building Using Logistic Regression
Variable Selection

The are several computational techniques for generating subset regression models and illustrate criteria for evaluating the model are all possible regression which requires that the analyst if all regression equations involving one candidate regressor, two candidate regressor and so on. And the second is step wise-type procedures which consists forward selection, back ward elimination and step-wise regress or. In our study we use back ward elimination method.

3.6.1 Parameter Estimation

Maximum Likelihood Ratio Estimation

The goal of logistic regression is to estimate unknown parameters; this parameter estimation involves like maximum likelihood estimation. The logistic regression uses maximum likely-hood estimation after transforming the dependent variable in to a logic variable. It is used to assess the overall goodness fit of the model

The likelihood-ratio test statistic is given by:-

G^2=-2log (L_0/L_1) =-2[log (L_0)-log (L_1)]

Where,L_0 is the log likelihood value of the model which has the intercept term only and L_1 is the log likelihood value of the full model.

Used to explore the extent to which the fitted response obtained from the postulated model compares with the observed data.

The equation to parameter estimation in logistic regression is the likelihood function for given(x_1, x2.....X_n) can be expressed as;

L (β, Y) = ЛI P (Y_i/X_{i1}, X_{i2}…Xn) =ЛI [$exp^{X, \beta}/1+ exp^{X, \beta}$] $^{1-Y}$

The summary measure of goodness of fit test statistics that will be used in this study are the Wald-test, the likely hood ratio test and score test. Agresti, (2008).

3.6.2. Model adequacy checking

Once a model has fit to a given data, it is a good statistical practice to check the adequacy of the model, which is essentially checking the agreement between the observed values under the model. If the agreement between the observed and the corresponding fitted values is good, the model may be acceptable. If not, the current form of the model will certainly not be acceptable and the model will need to be revising. This aspect of the adequacy of a model is widely referred to as goodness off it. An ill fitted model is says to display lack of fit. The discrepancy between observed data and fitted values under an assumed model can be assesses by several statistical measures.

3.6.2.1. The Wald Statistic

The Wald statistic is an alternative test which is commonly used to test the significance of individual logistic regression coefficients for each predictor variable (that is, to test the null hypothesis in logistic regression that a particular logit (effect) coefficient is zero).The Wald test statistic is:

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

The Wald statistic, W under the null hypothesis is approximately chi-square distributed. Each Wald statistic is compared with an X^2 distribution with 1 degree of freedom. For data that

produce large estimates of the coefficient, the standard error is often inflated, resulting in a lower value of the Wald statistic, and therefore the explanatory variable may be incorrectly assumed to be unimportant in the model. Likelihood ratio tests are generally considered to be superior.

Hypothesis testing**:-**

The hypothesis testing for the i^{th} explanatory

Ho: The coefficient is not associated with the predictor is equal to zero ().

H1: The coefficient associated with the predictor is not equal to zero ().

3.6.2.2 Parallel Lines test Assumption

In ordinal logistic regression models there is an important assumption which belongs to ordinal odds. According to this assumption:-

- The correlation between independent variable and dependent variable does not change for dependent variable's categories.
- The dependent variable must be categorical.
- Parameter estimations do not change for cut-off points. In an ordinal logit regression, when the assumption holds for j-1 logit comparison in a J categorized variable, a-1 cut-off points and j–1 parameters are found.
- The dependent variable's categories are parallel to each other. When the assumption does not hold, it means that there are no parallelism between categories. In ordinal logit regression, these tests examine the equality of the different categories and decides whether the assumption holds or not.

3.6.2.3. Goodness of Fit of the Model

After fitting a model, it is nature to enquire about the extent to which the fitted value of the response variables under the model compare with observed values of the response variable under the model compare with the observed values. The goodness of fit of the model in this study is to be tested in either of the following to approaches.

Deviance analysis:-Is based on the likelihood function of the observed for the fitted model (current model), the likelihood function for the true success probability under the assumed perfect model (full or saturated model), and say if Pearson X^2 statistics:-an alternative approach to test goodness of fit if to use person's X^2

CHAPTER FOUR

4. DATA ANALYSIS AND RESULT

4.1 Descriptive statistics

Descriptive statistics is a type of analyzing statistically recommended data that contain description like percentages and frequency of all selected variables. The descriptive statistics summarized in the following table.

Table 4.1: cross tabulation of house hold income level with explanatory variable

variable	category				Pearson chi-square	df	sig
		poor	medium	rich			
		Count (%)	Count (%)	Count(%)			
residence	urban	131(2.8%)	244(3.8%)	4857(87.5%)	1.216E4	2	0.0
	rural	4545(97.2%)	6181(96.2%)	692(12.5%)			
SDW	Bore hole	768(16.4%)	1295(20.2%)	156(2.8%)	1.285E4	28	0.0
	Protected well	458(9.8%)	468 (7.3%)	120(2.2%)			
	unprotectedwell	706(15.1%)	243(3.8%)	25(0.5%)			
	protectedspring	227(4.9%)	909 (14.1%)	176(3.2%)			
	unprotectedspring	706(15.1%)	1300(20.2%)	48(0.8%)			
	river	1067(22.8%)	672(10.5	46(0.8%)			

	Rain water	66(1.4%)	20(0.3%)	1(0%)			
	Tanker trunk	32(0.7%)	30(0.5%)	39(0.7%)			
	Cart with small tank	29(0.6%)	35(0.5%)	157(2.8%)			
	Piped water	617(13.2%)	1452(22.6%)	4781(86%)			
Ownfarm animal for agri	yes	3872(82.8%)	5517(85.9%)	1189(21.4%)	6.379E3	2	0.0
	no	804(17.2%)	908(14.1%)	4360(78.6%)			
Location of toilet facility	Own dwelling	5(0.6%)	35(0.8%)	327(6.1%)	5.859E2	4	0.0
	Own plot	613(69.8%)	3785(82.5%)	4570(85.9%)			
	Else were	260(29.6%)	770(16.8%)	421(7.9%)			
Location of source for drinking	Own dwelling	15(0.3%)	8(0.1)	3(0.2%)	63.149	4	0.0
	Own plot	59(1.3%)	172(2.8%)	58(4.8%)			
	Else were	4569(98.4%)	6051(97.1%)	1149(95%)			
sex	male	3040(65%)	4931(76.7%)	3442(62%)	3.368E2	2	0.0
	female	1636(35%)	1494(23.3%)	2107(38%)			
Has electricity	no	4616(98.7%)	5610(87.3%)	407(7.3%)	1.168E4	2	0.0
	yes	60(1.3%)	815(12.7%)	5142(92.7%)			
Main	dung	381(8.1%)	2249(35%)	303(5.5%)	1.155E4	18	0.0

flooer material							
	Wood plank	5(0.1%)	3(0%)	20(0.4%)			
	bamboo	3(0.1%)	60(0.9%)	51(0.9%)			
	Asphalt strip	10(0.2%)	129(2%)	1609(29%)			
	ceramics	0(0%)	4(0.1%)	220(4%)			
	cement	14(0.3%)	116(1.8%)	1122(20.2%)			
	carpet	32(0.7%)	163(2.6%)	1346(23.8%)			
	earth	4246(90.8%)	3699(57.6%)	899(16.2%)			
Has telephone	no	4673(99.9%)	6399(99.6%)	4699(84.7%)	1.678E3	2	0.0
	yes	3(0.1%)	26(0.4%)	8509(15.3%)			
Types of toilet facility	Flush to pit latrine	48(1%)	89(1.4%)	332(6%)	1.119E4	24	0.0
	Flush to some where	0(0%)	1(0%)	43(0.8%)			
	VIPL	2(0%)	14(0.2%)	110(2%)			
	Flush to do;t know	6(0.1%)	5(0.1%)	29(0.5%)			
	Pit latrine with slab	95(2%)	399(6.2%)	2560(46.1%)			
	Pit latrine with out slab	680(14.5%)	3974(61.9%)	1696(30.6%)			
	Composite toilet	26(0.7%)	67(1%)	20(0.4%)			
	other	3818(81.7	1867(29.1%	309(5.6%)			

	%)	)				
Flush toilet	1(0%)	9(0.1%)	450(8.1%)			

The chi-square test result indicate that there was a statistically significant association between Household income level and the explanatory variables such as sex , place of residence, types of toilet facility , main flower material ,has telephone , has electricity , Location of source for drinking , Location of toilet facility ,Own farm animal for agriculture , sources of drinking water at 5% levels of significance. Because the p-values of each explanatory variable are less than 0.05.

Table 4.1 showed that the proportion of house hold income differed based on the residence that means 4857 (87.5%) rich levels of households live in urban and 12.5% (692) rich levels of house hold live in rural, 4545(97.2%) poor level of house hold live in rural and 2.8%(131) live in urban, 6181 (96.2%) live in rural and 244 (3.8%) live in urban for medium levels of house hold. That means poor level of house hold are live majorly on rural.

The levels of house hold income differed based on the electricity that is poor level of house hold 60(1.3%) are used electricity and 98.7% not used electricity, 5142 (92.7%) rich level of house hold use electricity and 7.3% not use electricity. Means that poor level of households is majorly not used electricity. And also the levels of house hold income differed based on the sex that is 3442 (62%) rich level of house hold are male and 2107 (38%) are female. And 65% are male and 35 % are female poor levels of house hold .that means poor level of house hold are majorly female.

Household income differed based presence of telephone that is 4699(84.7%) rich levels of house hold use telephone and 15.3 % not used telephone. 4673(99.9%) poor levels of house hold not use telephone and 0.1% use. From main flower material 4246 (90.8%) poor levels of house hold live on earth and not live on ceramics and from rich levels of house hold 1609 (29%) live on asphalt strip and 20 (0.4) live on wood plank.

4.2. Ordinal Logistic Regression

Logistic regression examines the relationship between one or more predictor variables and a ordinal response variable and hence the logistic regression equation can be used to examine how the probability of an event changes as the predictor variable changes.

Table 4.2 Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept only	1675.614			
final	894.955	780.659	76	0.000

From the table 4.2 the model fitting information informs the -2 log likelihood chi square at76 degree of freedom which is equal to 0.000 is less than the level of significance (0.05). Therefore the null hypothesis is rejected and the house hold income is at least affected by one explanatory variable. To know which explanatory variables affect the response variable we look for individual test or Wald test in the below.

TABLE 4.3 Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	2739.453	1952	.000
Deviance	894.955	1952	1.000

State Hypothesis test

Ho: the model fit the data well Vs H1: Not Ho

From the table 4.3 we conclude that the goodness of fit indicates that the model is good fitted model since the p-values of Deviance is greater than 5% or 0.05 (level of significance).

Table 4.4 Pseudo R- Square

Cox and Snell	.536
Nagelkerke	.663
McFadden	.466

Link function: Logit.

From the table 4.4 also we conclude that as like goodness of fit test the pseudo R-square indicate the model is good fitted model since the value of Cox and Snell is found b/n 0 and 1.

TABLE 4.5 Test of Parallel Line

model	-2loglikely hood	Chi-square	df	sig
General	580.38	514.85	18	0.0812

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

A. The log-likelihood value cannot be further increased after maximum number of step-halving

B. The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.

c. Link function: Logit.

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. From the table 4.6 shows parallel line test for model with chi square value 514.85 and p-value=0.0812 which is greater than the 5% level of significance fail to reject the null hypothesis.

Therefore, there is no enough evidence to reject the Null hypothesis for the model. Thus, the proportional odds assumption appears to have held for the general model.

Table 4.6 Results of the ordinal logistic regression model

variable	category	coeff	s.e	df	AOR	p-value	95% C.I	
							LB	UB
House hold income level	poor	-10.240	5.5372	1		0.00	-21.13	-0.61
	medium	-3.7607	5.5182	1		0.00	-14.57	7.05
residence	rural	-2.3989	0.4119	1	0.091	0.00	-3.21	-1.59
	urban			0				
Has electricity	No			0				
	yes	2.4879	0.2917	1	1.34	0.00	1.92	3.06
sex	male			0				
	female	-0.696	0.2448	1	0.49	0.004	-1.18	-0.23
Main flower material	dung	1.6466	0.2663	1	5.189	0.00	1.13	2.15
	Wood plank	16.48	2322.9	1		0.994	-4536.44	4569.40
	bamboo	3.365	0.6225	1	28.9	0.00	2.15	4.58
	Asphalt strip	3.052	0.5401	1		0.00	1.99	4.11
	Ceramic	15.67	2322.9	1		0.995	-4537.31	4568.59
	Cement	1.935	0.6974	1	6.93	0.006	0.57	3.30
	carpet	3.348	0.5417	1	28.5	0.00	2.29	4.41
	earth			0				
Own livestock for farm animal	no			0				
	yes	0.680	0.2404	1	1.97	0.005	0.21	1.15
Age		-0.012	0.0057	1	0.98	0.026	-0.2400	-0.0015

Sources of drinking water	Bore hole	-0.440	0.2749	1		0.109	-0.98	0.98
	Protected well	-0.656	0.3521	1		0.062	-1.35	0.04
	Un protectedwell	-1.7503	0.4229	1	0.17	0.00	-2.58	-0.92
	Protected spring	-0.1909	0.3337	1		0.567	-0.85	0.46
	Un protected spring	-0.6837	0.3179	1	0.51	0.032	-1.306	-0.061
	river	-1.7817	0.3382	1	0.17	0.00	-2.444	-1.119
	Rain water	-0.3707	1.1705	1		0.752	-2.665	1.924
	Tanker trunk	-1.608	1.0764	1		0.135	-3.718	0.502
	Cart with small tank	-1.75	1.0857	1		0.106	-3.882	0.374
	Piped water			0				
Number of children		-0.380	0.1065	1	0.68	0.00	-0.589	-0.173

a. This parameter is set to zero because it is redundant

For the household income levels and the explanatory variable from the above table the model can be written as follows.

$$logit\,(p(yi \leq j) = \alpha j + \beta x \ldots\ldots\ldots\ldots\ldots J-1$$

$$logit(p(y \leq poor))$$
$$= -10.2403 - 2.398878X1 + 2.487856X2 - 0.6956877X3 + 1.646584X4 + 3.364826X41 + 3.052113X42 + 1.935451X43 + 3.347914X44 + 0.6802408X5 - 0.0127398X6 - 1.750325X7 - 0.6837328X71 - 1.781731\,X72 - 0.3809076X8.$$

$$logit\big(p(y \leq medium)\big)$$
$$= -3.760659 - 2.398878X1 + 2.487856X2 - 0.6956877X3 + 1.646584X4 + 3.364826X41 + 3.052113X42 + 1.935451X43 + 3.347914X44 + 0.6802408X5 - 0.0127398X6 - 1.750325X7 - 0.6837328X71 - 1.781731\,X72 - 0.3809076X8.$$

Where X_1=residence, X_2=has electricity , X_3=sex of head of house hold,X_4=main floor material(dug),x41= main floor material(bamboo),x42= main floor material(asphalt strip), x43= main floor material(cement),x44= main floor material(carpet),x_5= own live stock, farm animal , x_6 = age,x_7=source of drinking water(un protected well) ,x71= source of drinking water(un protected spring),x72= source of drinking water (river), x8= number of children.

4.2.1. Interpretation for coefficients of significant variables

Wald test

From table 4.6 we seen that coefficients, their standard errors, associated p-values, 95% confidence interval of the variables, and ajusted odd ratios. Since p-values less than alpha level they are statistically significant; otherwise not statistically significance. The exp (βj ̂), is the factor by which the odds of household income change when the jth independent variable increases by one unit. If $\hat{\beta}j$ is positive, adjusted odd ratio will be greater than one, which means the odds of household income increases. If βj is negative, adjusted odd ratio will be less than one, which means the odds of household income decreases.

Generally we conclude that the explanatory variables like place of residence , has electricity, sex of head of house hold head, main floor material , own live stock for farm animal , age , sources of drinking water ,number of children had a significance effect on the house hold income levels.

Odds ratio interpretation

The odds of having poor income levels of household VS middle and rich income levels of household for rural was 0.091 (95%CI: 0.041, 0.2036) times less than that compared to the urban house hold income levels, other variables are constant in the model, similarly the odds of having poor income levels of household and mediums VS rich levels of house hold income was 0.091 (CI: 0.041, 0.2036) times less than that compared to the urban house hold income levels, other variables are constant in the model.

The odds of having the poor income levels of house hold and medium levels of household VS rich levels of house hold having electricity was 5.189(95% CI: 3.096, 8.58) times higher than that compared to not having electricity other variables are constant in the model, And also the odds of the poor levels of house hold VS medium levels of house hold and rich levels of house hold having electricity was 5.189 (95% CI: 3.096, 8.58) times higher than that compared to not having electricity.

The odds of having the poor levels of house hold and medium VS rich levels of house hold the source of drinking water that is use river was 0.17(95%CI:0.087 , 0.332) times less than that compared the people use piped water, other variables are constant in the model. Similarly ,the odds of poor levels of house hold VS medium levels of house hold and rich levels of house hold the source of drinking water that is use river was 0.17(95%CI:0.087 , 0.332) times less than that compared to the people use piped water.

4.3. Discussion

This study is undertaken to identify the factors that affect income inequality in Ethiopia based on 2016 Ethiopian Demographic and Health survey (EDHS) data. Based on the Results of the ordinal logistic regression model (table 4.6) the parameter estimation that we get significant variables. Among the factors that are affected income inequality was age, sex, main flower material, sources of drinking water, number of children, Own livestock for farm animal, Has electricity , place of residence.

In our study place of residence and family size are the main factors that affect the income inequality in Ethiopia, similar results are reported by (Abe be et .al in 2016).

Age is another variable that affect the income inequality in Ethiopia, but the results reported by (almas and biwei ,in 2013) are not the significance variables for those study.

Other variables such as access to electricity, family size and sex of the household affect the income inequality and also similar results are reported by (Araya in 2010)

CHAPTER FIVE

5. CONCLUSION AND RECOMENDATION

5.1 CONCLUSION

The main objective of this study was to investigate the factors affecting household income inequality in Ethiopia. According to result this study was an attempt to examine the impact of some factors on the house hold income in Ethiopia. The finding shows that the household's income is affected by some factors during their life in our country.

According to inferential statistics, sex of house hold head , Main floor material, has electricity, sources of drinking water, own live stock for farm, place of residence, types of toilet facility , locations of source for water, , Locations of toilet facility , had a significant association with house hold income.

From the result of our study we have conclude that the predictors residence, main flower material, sex of house hold, electricity ,own live stock, age, sex of head of house hold head, numbers of children, sources of drinking water were found to have significance effect on the house hold income in Ethiopia using ordinal logistic regression

5.2. Recommendations

Based on the findings of our study, the following recommendations are put forward to bring about reduction in the income inequality.

- ✓ The government should be broadly distributed to rural the services of electricity to reduce income inequality in Ethiopia for future time.
- ✓ The government should be encouraging the female participation for have own sources of income for the highest chance than male to reduce income inequality for the future time.
- ✓ The government formulates suitable policy in family planning for reducing the number of children in order to decrease income inequality
- ✓ The government should be creating the expansion of urbanization for the areas of rural as much as possible in order to decrease income inequality.

REFERENCES

1.Agresti A, 2008. An introduction to categorical data analysis second edition. John Wiley and Soni I

2. Agresti, A (1996).An Introduction to categorical data Analysis. John Wiley and Sons, Inc., New York.

3. Abebe Fentaw Nebebe, CH. Appa Rao Determinants of income inequality in urban Ethiopia: A study of South Wollo Administrative Zone, Amhara National Regional State International Journal of Applied Research 2016; 2(1): 550-563.

4. Araya in 2010 poverty and income inequality in urban areas: (socio –economic analysis of house hold in wukro town, tigray region, ethiopia).

5. Araya M. Teka, Gabriel Temesgen and Zeremariam Fre in 2018 Status and Determinants of Poverty and Income Inequality in Pastoral and Agro-pastoral Communities: Household based Evidence from Afar Regional State, Ethiopia.

6. Almas biwei(2013)Analysis of the Determinants of Income and Income Gap between Urban and Rural China.

7. Bereket Getachew Mamo, Markos Abiso. Statistical Analysis of Factors Affecting Poverty Status of Rural Residence. American Journal of Theoretical and Applied Statistics. Vol. 7, No. 5, 2018, pp. 188-192. doi: 10.11648/j.ajtas.20180705.14.

8. Boccanfuso D, Kabore T.S in 2000 . Economic and Quantitative Poverty Analysis, The Panso Institute West Africa, Senegal, Dakar.

9. Bureau of African Affairs, 2008. Economic, environmental, health, social and politics of Ethiopia, U.S Department of State Bureau of Public Affairs, Washington .

10.Daniel Baden, Daniel Holliday, and Rohan Medhekar IN 2015 Income Inequality and Economic Growth: A Cross-Country Analysis.

11. Skare, M. and Stjepanovic, S., 2014. Income Distribution Determinants and Inequality International Comparison. Amfiteatru Economic, 16(37), pp. 980-993.

12. Tassew W, Hoddinott J, Dercon S. Poverty and Inequality in Ethiopia: 1995/96-2004/0Addis Ababa University, International Food Policy Research Institute and University of Oxfod,Addis Ababa, Ethiopia, 2009.
13. Little PG, Sanchez A, Lndrerch RC. (World Bank, 2009); The Evolution of Urban Inequality in Ethiopia, World Bank, New York, 1984
14. McKay A. Definition and Measuring of Inequality, Overseas Development Institute and University of Nottingham, England, 2002.
15. MoFED(Ministry of Finance and Economic Development),2007.Building on Progress: Plan for Accelerated and Sustained Development to End Poverty: Annual progress report 2006/07, Addis Ababa, Ethiopia.
16 . Mduduzi Biyase and Talent Zwane in2017An Empirical Analysis of the Determinants of poverty and household welfare in South Africa.
17. Heshmati A. (2004), Inequalities and their measurement, Unpublished Manuscript.
18. Keeley, Brian (2015), "How does income inequality affect our lives?", in Income Inequality: The Gap between Rich and Poor, OECD Publishing, Paris.
20. Zelda Okatch , Abu Siddique and Anu Rammohan (2012), determinants of income inequality in botswana. Business School University of Western Australia discussion paper 13.15.
19. Odedokun, M.O. and Round, J.I., 2004. Determinants of Income Inequality and its Effects on Economic Growth: Evidence from African Countries. African Development Review, 16(2), pp. 287-327.

20. Zelda Okatch , Abu Siddique and Anu Rammohan (2012), determinants of income inequality in botswana. Business School University of Western Australia discussion paper 13.15.

.

DECLARATION

we, the undersigned declare that this paper is our original work and has not been presented for a degree in any other university and that all sources of material used for the paper have been duly acknowledged.

Declared by:

Name_____

Signature_____

Date_____

Name_____

Signature_____

Date_____

Confirmed by advisor

Name_____

Signature_____

Date_____