ADDIS ABABA UNIVERSITY
GRADUATE STUDIES PROGRAMME
COLLEGE OF COMPUTATIONAL AND NATURAL SCIENCES
DEPARTMENT OF STATISTICS

RISK FACTORS FOR ANAEMIA LEVELS AMONG WOMEN OF REPRODUCTIVE AGE IN ETHIOPIA: A PARTIAL PROPORTIONAL ODDS MODEL APPROACH

Birhane Zelalem

A Thesis submitted to the Department of Statistics presented in partial fulfillment of the requirements for the Degree of Master of Science in Statistics

Addis Ababa, Ethiopia
September, 2014

ADDIS ABABA UNIVERSITY
GRADUATE STUDIES PROGRAMME
COLLEGE OF COMPUTATIONAL & NATURAL SCIENCES
DEPARTMENT OF STATISTICS

RISK FACTORS FOR ANAEMIA LEVELS AMONG WOMEN OF REPRODUCTIVE AGE IN ETHIOPIA: A PARTIAL PROPORTIONAL ODDS MODEL APPROACH

Birhane Zelalem

A Thesis submitted to the Department of Statistics presented in partial fulfillment of the requirements for the Degree of Master of Science in Statistics

Advisor: Dr Shibru Temesgen

Addis Ababa, Ethiopia
September, 2014

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared by Birhane Zelalem, entitled: Risk factors for anaemia levels among women of reproductive age in Ethiopia: A partial proportional odds model approach is used and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Statistics complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Examiner ____________________________ Signature __________ Date _________

Examiner ____________________________ Signature ___________Date _________

Advisor _____________________________ Signature __________ Date _________

____

Chairman of Department or Graduate Program Coordinator

Acknowledgment

First, and foremost, I would like to thank GOD, the almighty for giving me the opportunity to pursue my study at the Department of Statistics in Addis Ababa University.

Secondly, I owe the deepest gratitude to my thesis advisor Dr. Shibru Temesgen, for his valuable and constructive comments and encouragements throughout my study.

Thirdly, my deepest thanks go to all staff members of statistics department at Addis Ababa university especially, my instructors Dr. Shibru Temesgen, Prof. Eshetu Wencheko, Prof. M. K. Sharma and Dr. Butte Gotu.

Fourthly, my heartfelt thanks go to my mother and brother for their support in the period of study and my thesis work.

Finally, I would like to acknowledge all my friends and classmates especially, Sebsib Maunenda, Belete Adelo, Leul Mekonnen, Amare Wubishet, Berhanu Teshome, Kidane Aragaw, Yared Seyoum, Worku Biyadgie, Eskedar Girma, Garuma Wakuma, Abdulaziz Shifa, for their encouragement and support that they gave me during my study period and this thesis work.

Abstract

Birhane Zelalem

Addis Ababa University, 2014

RISK FACTORS FOR ANAEMIA LEVELS AMONG WOMEN OF REPRODUCTIVE AGE IN ETHIOPIA: A PARTIAL PROPORTIONAL ODDS MODEL APPROACH

Anaemia is defined as a lower than normal level of hemoglobin in the blood. Normal levels of hemoglobin are different for pregnant women, non pregnant women, men, child etc. The prevalence of anaemia in the world according to WHO, based on the studies conducted from 1993 to 2005 was 24.8 percent and it affected 1.62 billion people worldwide. It is one of the global widespread public health and nutritional problems affecting both developing and developed countries, and occurs at all stages of life cycle but it is prominent in pregnant women, young and other women of reproductive age (WHO, 2008). According to the 2011 EDHS data the prevalence of anaemia is 17 percent among Ethiopian women aged 15-49. The objective of this study was to identify determinants of anaemia levels among women of reproductive age (between 15 and 49) in Ethiopia using the 2011 EDHS data. 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, demographic and health related variables as explanatory variables and anaemia levels of reproductive age of women as the response variables. The result of the analysis revealed that the variables education level, region, parity, pregnancy status, body mass index, place of residence, contraceptive methods and marital status were found to be significant determinants of anaemia levels among women in the reproductive age group in Ethiopia and from the result it also suggested that pregnant women were more likely to be moderate anaemic and severe anaemic than none pregnant women. Women who had a large number of children were found to be more likely of being mild anaemic than those who had no child. Rural resident women were more likely of being mild, moderate and severe anaemic than urban resident.

Table of Contents

List of Tables

List of Figures

Acronyms

AIC	Akaike information criteria
BIC	Bayesian information criteria
BMI	Body mass index
EDHS	Ethiopian demographic and health survey
GLM	Generalized linear model
GOLOGIT2	Generalized ordered logit two
GOM	Generalized ordered logit model
Hb	Hemoglobin threshold
LR	Likelihood ratio
OLOGIT	Ordered logit
POM	Proportional odds model
PPOM	Partial Proportional odds model
SNNP	Southern Nation, Nationality and peoples
WHO	World health organization
Chi2	Chi-square
LL	Log-likelihood
OR	Odds ratio
CI	Confidence interval
ML	Maximum likelihood
UNFPA	United Nations Population Fund
IPHN	Institute of Public Health Nutrition

CHAPTER ONE
INTRODUCTION

1.1. Background

Anaemia is defined as a lower than normal level of hemoglobin in the blood. "Lower than normal level" has been debatable for many years and also in different books and research works but this lower than the normal level of hemoglobin are different for pregnant women, non pregnant women, child etc. It is also corrected for different factors like altitude, smokers according to the WHO. It is one of the global widespread public health and nutritional problems affecting both developing and developed countries and occurs at all stages of life cycle, but its prominent in pregnant women, young and other women's of reproductive age (WHO, 2008).The prevalence of anaemia in the world according to WHO, based on the studies conducted from 1993 to 2005 was 24.8 percent and it affected 1.62 billion people worldwide. The estimated prevalence was 41.8 percent in pregnant women and 30 percent in non pregnant women, in number,56 million pregnant women and 468 million non pregnant women were affected (WHO,2008) and also according to WHO the estimated prevalence of anaemia in developed and developing countries in pregnant women is 14 percent and 51 percent, respectively. Recent findings suggest that there is a decline in the prevalence of iron deficiency anaemia among industrialized regions but the global prevalence is still higher in the developing world compared to the expected decline rate (Numbare et al, 2012).WHO estimated that worldwide 41 percent of women are anaemic (Stoltzfus, 2011).The causes of anaemia is multi-factorial where about half of the global burden of anaemia is due to iron deficiency. Iron deficiency, in turn, is largely due to an inadequate dietary intake of bioavailable iron, inadequate dietary iron during periods of increased iron requirements (such as pregnancy and infancy), increased blood loss due to hookworm infestation, and infections such as malaria. Nutritional anaemia includes anaemia due to deficiency in iron plus deficiencies in folate, vitamins B and B12, and certain trace elements involved with red blood cell production. The highest prevalence of anaemia exists in Low Income Countries (LIC) predominantly in South East Asia and African. In South East Asia 48.2 percent pregnant and 45.7 percent non pregnant women are anaemic. In Africa 57.1 percent of the pregnant and 47.5 percent of non pregnant women were anaemic (WHO, 2008).

Anaemia plays an important role in socio-economic development of nations. The effect of anaemia could be attributed to reduction in school performance, physical weakness, fatigue, and reduction in work capacity among affected individuals (Stoltzfus, 2001 and 2003). Women are at increased risk of becoming anaemic due to several physiologic or socio-cultural factors. Significant iron loss occurs in women during menstruation. Median monthly loss of blood during menstruation is estimated at 35 ml which is equivalent to more than 12 mg of iron (INACG, 2002). Because of women's greater iron requirements, and also because they usually consume less food than men, women's daily iron intake tends to be marginal. This is especially true in many developing countries where general food consumption is reduced because of poverty (Sharman,2000).The prevalence of anaemia as a public health problem is categorized as follows: <5%, no public health problem; 5–19.9%, mild public health problem; 20–39.9%, moderate public health problem; ≥40%, severe public health problem (WHO,2008). There is an established relationship between anaemia and maternal mortality. Iron deficiency anaemia alone was estimated to be an underlying factor for 22% of maternal deaths around the world. Severe anaemia is a major contributor to maternal deaths. In addition, mild and moderate anaemia can also increase the risk of death when compared with normal haemoglobin levels (A2Z, 2006; Kalaivani, 2009; Brabin et al, 2001; Geelhoed et al., 2006). Anaemia mostly occurs in women of reproductive age due to menstrual, closed pregnancy and other socio-economic and demographic factors.

In 1998, for the first time, a relatively-representative sample of pregnant and lactating women in Ethiopia was studied, and the prevalence rate of iron deficiency was 18.7 percent (Haider et.al, 1999). Subsequent studies among 196 lactating women from urban slum (an overcrowded area of a city in which the housing is typically in very bad condition) communities of Addis Ababa, the capital city of the country, showed a prevalence of 22.3 percent, suggesting the problem to increase in its magnitude and reaching to a level of moderate public-health significance (Haider et.al, 2003).The study was done to compare the risk factors for anaemic and non-anaemic women and to identify the potential causes of anaemia in among Ethiopian women. In 2006, Micronutrient Initiative (MI) Ethiopia estimated 27.0 percent and 30.6 percent prevalence of anemia among women of reproductive age and pregnant women, respectively (Micronutrient Initiative, Ethiopia country profile (2006)).

Identifying the magnitude of anaemia and its determinants in high-risk groups, such as women of child bearing age, would be essential for evidence-based intervention modalities, particularly in developing countries, such as Ethiopia, where the social conditions pose serious challenges to women. The nutritional status of women in Ethiopia, as in other developing countries, is low, and their daily workload is often enormous because of reproducing and ensuring the survival of their children (Berhane et.al, 2001). To improve the nutrition situation of Ethiopian women, there have been several interventions by the Ministry of Health through its Essential Nutrition Action (ENA) plan, comprising the supplementation of three major nutrients (vitamin A, iron, and iodine) and other primitive activities, such as exclusive breastfeeding, appropriate complementary feeding, and improved maternal and child nutrition (Gibson and Mace, 2006).

Despite the efforts of the line ministry and its stakeholders, the demographic health survey report (EDHS) of 2005 showed that 27 percent of women, aged 15-49 years, were chronically malnourished and about the same proportion suffered from anaemia with significant regional variations. Other nutritional data, such as iron status, however, were not documented in the EDHS 2005. The available information on the causes of iron-deficiency anaemia is limited in their capacity to be representative for the entire country, and some are misleading and non-conclusive, despite the problem being among the 10 top morbidities (Federal Ministry of Health, 2004). Although data on iron-deficiency anaemia are partial, the results show that iron-deficiency anaemia is a mild to moderate public-health problem in the country (Haider, 1999).

According to the 2011 EDHS 17 percent of Ethiopian women age 15-49 are anaemic, with 13 percent having mild anaemia, 3 percent having moderate anaemia, and 1 percent having severe anaemia. A higher proportion of anaemic women were pregnant (22 percent) than women who are breastfeeding (19 percent) and women who are neither pregnant nor breastfeeding (15 percent) (EDHS, 2011). The prevalence of anaemia is not the same across the world countries due to different factors. This study focused on social, economical, geographical determinants for prevalence of anaemia and to rank the significant determinants based on the anaemia levels defined as severe anaemia, moderate anaemia, mild anaemia and non anaemic on women's of reproductive age by using the ordinal logistic regression model so that, the outcomes of the findings can help in evidence-based decision to develop and control intervention strategies to improve the health status of the women of reproductive age.

1.2. Statement of the problem

Many researches have been done on the prevalence of anaemia on children and women of reproductive age in the developed countries. In developing country like Ethiopia few researches were conducted on anaemia relatively to its prevalence, burdens and social health problems. Anaemia is a health problem that increases the risk of mortality, under weight infants and decrease physical work capacity of women and children. The prevalence of anemia among women who had a live birth in the five years preceding the surveys decreased from 20% in 2005 to 13% in 2011 and the data for 2011 revealed a much wider gap in the prevalence of anemia between pregnant (29.9%) and non-pregnant women (10.8%)(UNFPA, 2012). Due to the above health consequences on individuals lead to become social and economic problem for the country that means if the women in the period of child bearing are not healthy they cannot be productive in economic and social sectors on the country and also there child may not be healthy. Most research in Ethiopia about prevalence of anaemia among children and women did not give much attention to social, economic and geographical factors mean while only taking the biological factors. Some previous researches were conducted with the objective of identifying the determinants of anaemia by comparing the health status of the women of reproductive age as anaemic and non anaemic. The current study classifies the health status of women into four categories according to the anaemia severity as non anaemic, moderate, mild and sever anaemic based on the concentration of haemoglobin in the individual's blood. This study was an attempt to show the social, economical and geographical determinants of anaemia on women of reproductive age and to fill the gap of knowledge of the contributions and relationship of socio-economic status, the education of the women, and place of residence, nutritional status, occupation of the women, the number of children the women has on/with anaemia levels.

1.3. Objectives of the study

General objective of the study

The general objective of this study is to identify the possible determinants of anaemia among women of reproductive age in Ethiopia.

Specific objectives of the study

1. To study the effect of determinants of anaemia levels among women of reproductive age in Ethiopia.

2. To assess the association between maternal and socio-demographic characteristics with anaemia among women of reproductive age in Ethiopia.

3. To compare anaemia status in pregnant women and non-pregnant women.

1.4. Significance of the study

The significant of this study is to create awareness about anaemia and to identify the common risk factors associated with anaemia and in addition to understand socio-economic and demographic differentials on anaemia levels among womens of reproductive age in Ethiopia then report to the concerned body.

1.5. Limitations of the study

The data used in this research were obtained from EDHS 2011. On this investigation the anaemia levels of the individuals are classified according to their hemoglobin concentration in the blood that means there is no other way of knowing an individual (a person) in different anaemia levels. There were missing values on the variables like anaemia status and body mass index. This may affect the conclusions of the study. Also since the data were collected in 2011this is before three years therefore, we cannot know the current prevalence and correlates of anaemia on women of reproductive age in Ethiopia.

CHAPTER TWO
LITERATURE REVIEW

2.1. Theoretical Literature review

2.1.1. Definition and causes of anaemia

Anaemia is a condition in which the number of red blood cells (and consequently their oxygen-carrying capacity) is insufficient to meet the body's physiologic needs. Specific physiologic needs vary with a person's age, gender, residential elevation above sea level (altitude), smoking behavior, and different stages of pregnancy. Iron deficiency is thought to be the most common cause of anaemia globally, but other nutritional deficiencies (including folate, vitamin B12 and vitamin A), acute and chronic inflammation, parasitic infections, and inherited or acquired disorders that affect haemoglobin synthesis, red blood cell production or red blood cell survival, can all cause anaemia. Haemoglobin concentration alone cannot be used to diagnose iron deficiency. However, the concentration of haemoglobin should be measured, even though not all anaemia is caused by iron deficiency. The prevalence of anaemia is an important health indicator and when it is used with other measurements of iron status, the haemoglobin concentration can provide information about the severity of iron deficiency (WHO, 2008).

Haemoglobin threshold

Normal Hb distributions vary with age, sex, physiological status, pregnancy status etc. During pregnancy Hb level becomes less than non pregnancy women (Koller, 1982). WHO Hb thresholds were used to classify individuals living at sea level as anaemic (Table 2.1) (WHO, 2001). Statistical and physiological evidence indicate that Hb distributions vary with smoking (Nordenberg et.al, 1990) and altitude (Hurtado et.al, 1945) and therefore, the prevalence of anaemia corrected for these factors was used when provided by the survey.

Table 2.1: Hb thresholds used to define anaemia

Groups	Haemoglobin threshold(g/l)
Children (0.50–4.99 yrs)	110
Children (5.00–11.99 yrs)	115
Children (12.00–14.99 yrs)	120
Non-pregnant women (≥15.00 yrs)	120
Pregnant women	110
Men (≥15.00 yrs)	130

Source: WHO, 2001

2.1.2. Consequence of anaemia

Anaemia has detrimental physical, social and economic effects. Even mild to moderate anaemia affects the sense of well-being, resulting in fatigue, stress and decrease in work capacity. Severe anaemia that occurs in developing countries is a major cause of maternal mortality and morbidity (Harrison & Rossiter, 1985; WHO, 1992). Anaemia is attributed as a direct or indirect cause of about 26% of maternal deaths in Africa. Severe anaemia may cause cardiac failure and death, whereas chronic anaemia is considered to be contributory, especially in cases of haemorrhage and infection. Furthermore, anaemic women are poor anaesthetic and operative risks, as anaemia may lead to poor healing of the wound and to increased susceptibility to infections (Brabin, 1985). The extent to which anaemia may cause maternal deaths is also dependent on availability and quality of obstetric services. Availability of blood transfusion is often lifesaving in cases of severe anaemia (Harrison & Rossiter, 1985).

Anaemia during pregnancy is also associated with an increased risk of intrauterine growth retardation, premature delivery and low birth weight, resulting in an increase in perinatal mortality (Brabin et al., 1990; Fleming, 1989b). Infants of anaemic women are born with reduced iron stores and are at risk of anaemia during infancy and increased risk of infant morbidity and mortality (Cook, 1994; Scholl et al., 1992). Reduction of anaemia during pregnancy is therefore a key component of safe motherhood.

Anaemia is a serious obstacle to survival, health and development for individuals as well as the country which affects prominently women of reproductive age and children's more generally when a person is anaemic their heart has to work harder to pump the quantity of blood needed to get adequate oxygen around the body (IPHN, 2007). During heavy exercise, the cells may not be able to carry enough oxygen to meet the body's need and the person is exhausted having the above consequence: It is not a disease in itself, but it is a result of a malfunction somewhere in the body therefore anaemia devastating effect on the health and physical and mental productivity and affect the quality of life of human being. Anaemia also plays an important role in the economic development of a nation. This effect of anaemia could be because of the fact that it causes reduced school performance in children and women, causes weakness, fatigue, and reduced physical ability to work among affected individuals (Brick and Peters, 2014).

2.1.3. Anaemia in women and maternal health

There is an established relation between anaemia and maternal mortality. Iron deficiency anaemia alone was estimated to be an underlying factor for 22% of maternal deaths around the world (see Table 2.2 below) (A2Z FANTA project, 2006). Based on the study in Kwale district in Kenya reported anaemia related maternal deaths of 82/100,000 live births, similar or higher figures were reported from other countries in Africa and Asia (WHO, 1993).Severe anaemia has been shown to be major contributor to maternal deaths though mild and moderate anaemia also increases the risk of death when compared with non-anaemic mothers (A2Z FANTA project, 2006; Kalaivani, 2009; Brabin, 2001; Geelhoed, 2006).

Table 2.2: Women and maternal anaemia

Causes of maternal death	% Contribution
Anaemia	8
Iron deficiency anaemia	22
HIV	3
Hemorrhage	9
Hypertensive Disorder	10
Sepsis	11
Obstructed Labor	7

Unsafe Abortion	5
Indirect Causes	14
Other Direct causes	5
Unclassified	6

Source: A2Z FANTA project, 2006

2.2. Empirical literature review

2.2.1. Determinant of anaemia in women of reproductive age

According to the result obtained from WHO, the prevalence of anaemia is high in developing countries due to the socio-economic and health development. Africa and South East Asia countries are highly affected by anaemia (WHO, 2008).

A quantitative cross-sectional study were carried out for correlates of anaemia among women of reproductive age based on secondary data obtained from Ethiopian DHS 2005 and using binary logistic regression. This method was employed to control potential confounders and to explore associations between the dependent variable (anaemia status) and a wide range of the aforementioned independent variables. Rural residence, poor educational and economic status, 30-39 years of age and high parity were key factors predisposing women to anaemia. Utilizing maternity services, taking iron and vitamin A supplement during pregnancy and postpartum period, didn't have a significant effect in reducing the burden of anaemia (Gebremedhin and Enquselassie, 2011).

Using EDHS 2005 data, the study by Wondu and Bijlsma (2012) has shown that women's educational status, grouped altitude of residential places and household wealth index categories have significant impact on the prevalence of anaemia. The prevalence of anaemia was positively associated with past five year's fertility level. Unavailability of toilet facilities, being resident of rural area and not using contraceptive methods were also associated with prevalence of anaemia among women

A cross-sectional community-based study was conducted by Haider in nine of the 11 regions of Ethiopia to assess the magnitude of anaemia, deficiencies of iron and folic acid and compare the factors responsible for anaemia among anaemic and non-anaemic women of childbearing age

(15-49 years) (Haider, 2010).To identify the effect and predict the most important determinants of anaemia, a stepwise logistic regression analysis was performed. Women having two or more children, using open field as toilet, suffered from chronic illnesses, and who had intestinal parasites were positively associated with anaemia. Women with no formal education and who did not use contraceptives were negatively associated with anaemia. The major determinants identified for anaemia were chronic illnesses (AIDS, diabetes, cancer).

The first large nutrition study from representative sample of women in Ethiopia was conducted from June to July 2005. The results revealed that intake of vegetables less than once a day and meat less than once a week were common, and associated with increased anaemia, and nutrition related and chronic illnesses are the most common causes of anaemia. The researchers concluded that moderate nutritional anaemia in the form of iron deficiency anaemia is a problem in Ethiopia and therefore, the need for improved supplementation to vulnerable groups is warranted to achieve the United Nation's Millennium Development Goals (Haider and Pobocik, 2009).

Based on the data obtained from socio-demographic and socio-economic survey on the anaemia and associated risk factors among pregnant women in gilgel gibe dam area, southwest Ethiopia; place of residence has significant factor for being anaemic, rural women were highly affected due to lack of knowledge of anaemia and due to smaller number health facilities in rural residences (Getachew et al., 2012).

Based on the study conducted on prevalence and predictors of paternal anemia during pregnancy in Gondar, northwest Ethiopia a multiple logistic regression analysis, controlling the possible cofounders, low monthly family income, large family size, hookworm infection and HIV seropositivity were identified as significant predictors of anaemia (Melku et al., 2014).

A study in Bushulo health center, southern Ethiopia, showed that Age, residence, occupation, income family, religion, marital status have significance on the prevalence of anaemia (Tadege, 2009).

A study conducted in India to determine the prevalence of anaemia among ever-married women of reproductive ages and to explore some factors commonly associated with anaemia. Background characteristics such as age, place of residence, nutritional status, number of children ever born, pregnancy status, educational achievement, and economic status were considered in the study. As a response variable, anaemia level was taken as a dichotomous variable. The findings of the study revealed that the predictors such as pregnancy status, nutritional status, economic status, education level and the habit of cigarette smoking/pan/bidi/gutka were found to be statistically significant. About 49.6% of the women were anaemic. Women in the age groups 20-24 years were at high risk of anaemia. Women who were pregnant and undernourished were at high risk of being anaemic; urban women and with high education level were at low risk of anaemia. The habit of cigarette smoking/pan/bidi/gutka etc. also increased the risk of anaemia (Sanku et.al, 2010).

The study on the burden of anaemia among women in India based on the National Family Health Survey 1998/99 was conducted to investigate the prevalence and determinants of anaemia among women in Andhra Pradesh examined differences in anaemia related to social class, urban/rural location and nutrition status body mass index (BMI). It was found that prevalence of anaemia was high among all women, and poor urban women had the highest rates of being anaemic. Fifty-two percent women with low BMI, fifty percent of normal BMI, and forty-one of overweight women were anaemic (Bentley and Griffiths, 2003).

A study was conducted in Jaipur city, India on the prevalence of anaemia and socio-demographic factors associated with anaemia among pregnant women attending antenatal hospital. The results showed that overall prevalence of anaemia among pregnant women was found to be sixty-three percent. Factors such as level of education and socio-economic status were significantly associated with prevalence of anaemia (Priyanka et.al, 2011).

According to the study conducted on the effect of a health education program on prevalence of anaemia among pregnant women in Tafila, Jordan, anaemia was associated with age of mother, family size, number of deliveries, and abortions. On the other hand, the relationship between anemia and average monthly family income, iron/vitamin pill intake as well as knowledge and consumption of iron rich foods had a negative association with anaemia. Furthermore, no significant relationship was detected between anemia and each of the following: employment, level of education, and use of family planning (Hamad et.al, 2001).

A study conducted in Bali, Indonesia showed that iron-deficiency anemia in pregnant women was 46.2% and that most of the cases of anaemia were mild. The study identified that length of gestation, level of education, antenatal intake of iron pills were significant determinants (Suega et.al, 2002).

A community-based study in Udupi district in the Karnataka state of India showed that the prevalence of anaemia was found to be 50.14 %. Prevalence was high among young women, women belonging to low socioeconomic status and women with short pregnancy intervals and higher parity (Judith et.al, 2008).

A study based on the data obtained from primary healthcare clinic attendees was undertaken in Trinidad and Tobago to investigate the relationship of anaemia with abortion, parity and child spacing. Logistic regression showed relationships between anaemia and the variables parity, gravidity hence, previous spontaneous abortions were directly. Clinic attendees, age were not associated to the severity of anaemia (Uche-Nwachi et al., 2010).

A study was conducted to see the relationship between anaemia and BMI among unmarried girls in Bhodia, Fatehabad district, Haryana. A negative association was observed between BMI and Hb concentration among overweight and obese groups (Peter et.al, 2012). Age was directly and highly correlated with Hb concentration. Incidences of anaemia increased with the age and were highest among the period of increasing growth and adolescents. A similar study on Iraqi adolescent females showed a negative correlation between age and BMI (Thomsen et al., 1986). The study revealed that most of girls of higher age group had lower levels of hemoglobin (<12 gm/dl).

A study in the Niger Delta, Nigeria showed that prevalence of anaemia was 66.7% among pregnant women. None of the women had severe anaemia. Age and occupation were significantly associated with the risk of anaemia. Marital status, educational level, social class and parity did not significantly affect the incidence of anaemia in pregnancy (Ibrahim et.al, 2012).

A cross-sectional study was carried out on the prevalence and socio-demographic factors associated with anaemia during pregnancy based on the data obtained from primary health centre in Rivers State, Nigeria. A Chi-square test showed that anaemia was observed to be least prevalent among women within the extremes of reproductive age (≤ 20 years and 36–40 years). There was no statistically significant association between age, educational level and marital status. The association of anaemia with social class was statistically significant. Severe anaemia was significantly associated with educational status and socio-economic status. The association of anaemia with social class was statistically significant. Severe anaemia was significantly associated with low educational status and low socio-economic status (Ndukwu and Dienye, 2012).

A hospital-based study conducted in Tibet to study the levels of Hemoglobin and anaemia on pregnant women in the highlands of Tibet. The result showed that gestational age, ethnicity, residence and income were significantly associated with the hemoglobin concentration and prevalence of anaemia. Specially, the hemoglobin concentration of pregnant women decreased with increase in gestational age (Yuan et.al, 2009).

A cross-sectional study using nationally representative data over the 7-year period showed that anaemia prevalence increased significantly from 51.3% to 56.1% among Indian women. This corresponded to a 1.11-fold increase in anaemia prevalence after adjustment for age and parity, and 1.08-fold increase after adjustment for wealth, education and caste. There was marked state variation in anaemia prevalence in only 4 of the 25 states was a decline in anaemia prevalence. Anaemia was socially patterned, being positively associated with lower wealth status, lower education and belonging to scheduled tribes and scheduled castes. In this context socioeconomic inequalities in anaemia by wealth, education and caste have narrowed significantly over time of the study (Balarajan et.al, 2013).

A study conducted on prevalence and risk factors of anaemia among women of reproductive age in Bursa, Turkey revealed that more than 2 sanitary pads during menstruation, and more than five days of menstrual bleeding was found to be risk a factor for anaemia. There were no significant association between anaemia and age, education, marital status, job, parity, BMI, regularity of cycle and length of cycle (Kayihan and Nilgun, 2008).

A study conducted in Sanandaj, Iran showed that the prevalence of low hemoglobin was significantly higher in mothers with lower level of education, pregnancies and childbirths more than 3 times, a history of anaemia, and irregular prenatal care. Furthermore, mothers with low hemoglobin encountered to more pregnancy complications (Fariba et al., 2009).

A cross-sectional study on girls in Jamshoro, Pakistan showed that out of 75 girls (17–19 years), 60% were anaemic, out of these 55.55% had mild anaemia (<12 gm/dl), 22.22% had moderate anaemia (8–10 gm/dl), and an equal percentage (22.22%) had severe anaemic (<8 gm/dl). Among 75 females between 14–16 years, prevalence of anaemia was the highest (93.33%) in this age group. out of these, 57.14% were found to have mild anaemia, 14.28% had moderate anaemia and 28.57% had severe anaemia. It was observed that severe anaemia with decreased BMI; all of those suffering from severe anaemia had BMI (<17). An association was found between anaemia and dietary patterns. All girls were unaware about taking iron containing diet (Aabroo et.al, 2012).

CHAPTER THREE
DATA AND METHODOLOGY

3.1. Data

This study was based on secondary data obtained from the 2011 Ethiopia demographic and health survey (EDHS, 2011) which was conducted by the Central Statistical Agency with support of the Ministry of Health and other donor agencies. The survey was primarily designed to collect data on marriage, fertility, family planning, maternal and child health, HIV/AIDS, malaria, anaemia, nutrition and gender. It was conducted as a nationally representative survey for a population of Ethiopia where a sample of 17,817 households was selected. In the sample 17,385 women aged 15-49 identified for interview; complete interviews were conducted for 16,515, yielding a response rate of 95 percent. This study focused on factors related to anaemia on women of reproductive age in Ethiopia based on 15,351 women of reproductive age after removing incomplete data from the survey.

3.2. Variable of the study

Dependent variable

The response variable of this study is anaemia status of women aged between 15 to 49, which is categorized into four ordinal categories: non-anaemia, mild anaemia, moderate anaemia and severe anaemia are given below in Table 3.1.

Table 3.1: Description of the dependent variable

Response variable	Value of the levels	Type
Anaemia levels (ANA)	1=non-anaemia, 2=mild anaemia, 3=moderate anaemia, and 4=severe anaemia	Ordinal

Independent variables

Independent variables related to anaemia among women of reproductive age based on various literature reviews and theoretical aspect are given with their respective coding in Table 3.2 below.

Table 3.2: Description of explanatory variables

Predictor's name	Value of level Predictor's	Type
Age of women (AGE)	1=15-19,2=20-24,3=25-29,4=30-34,5=3539,6=40-44, 7=45-49	Ordinal
Region of women(REG)	1=Addis Ababa,2=Afar,3=Amara,4=Oromiya, 5=Somali,6=Benishangul-Gumuz,7=SNNP, 8=Gambela,9= Harari,10= Tigray,11= Dire Dawa	Nominal
Place of residence(PR)	0= Rural and 1= Urban	Nominal
educational level(WEL)	1=no education, 2=primary and 3=secondary and higher	Ordinal
Wealth index(WI)	1=poor,2= Middle and 3=rich	Ordinal
Parity(TCEB)	1= No child, 2=1-2,3=3-5, 4= 6 and above	Ordinal
Pregnancy Status(PS)	0= No or unsure, 1=yes	Nominal
Contraceptive methods(CM)	1= No method,2= Folkloric and traditional methods, 3= Modern methods	Nominal
Body Mass Index(BMI)	1= Less than 18.5, 2=18.5-24.9, 3=25.0 and above	Ordinal
Marital status(MS)	1=single,2=married,3= Widowed, 4=divorced	Nominal
Habit of Cigarette smoking(HC)	0=no, and 1=yes	Dichotomous
Occupation type(WOT)	1= Not working, 2= Agricultural sector, 3= Non-agricultural sector and/or others	Nominal

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 (Cabrera, 1994). 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 and Lemeshow, 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. 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 and Lemeshow, 2000).

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 (McCullagh, 1980; Hosmer and Lemeshow, 2000). 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 $j = 1, .., J$ are the ordinal categories of the response variable Y, and the vector of explanatory variable X, and denoted by vector form $X = (X_1, X_2, \dots, X_K)'$. For Y, the response with the J ordinal categories given that of K explanatory variables the individual probabilities are defined as follow;

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

$$\pi_j(X) = P(Y \leq j|X) = P_1 + P_2 + \cdots + P_j, \text{ for } j = 1, \dots, J-1 \qquad (3.1)$$

$\pi_j(X)$, 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

$$odds[\pi_j(X)] = \frac{\pi_j(X)}{1-\pi_j(X)}, j = 1, \cdots, J-1 \qquad (3.2)$$

The logarithm of the odds first j-1 cumulative probabilities

$$ln(odds[\pi_j(X)]) = ln\left(\frac{\pi_j(X)}{1-\pi_j(X)}\right), j = 1, \cdots, J-1 \qquad (3.3)$$

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:

$$logit[P(Y \leq j|X)] = \alpha_j - \sum_{k=1}^{K} \beta_k X_k \;, j = 1, \cdots, J - 1 \tag{3.4}$$

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 β_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 (log odd) 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 models 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

3.3.1.1.1. Likelihood function and parameter estimation

In the model:

$$logit[P(Y \leq j|X)] = \alpha_j - \sum_{k=1}^{K} \beta_k X_k \ , j = 1, \cdots, J-1$$

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:

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

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

$$\pi_j(X_i) = P\big((Y \le j|X_i)\big) - P\big((Y \le j-1|X_i)\big)$$

And the cumulative probabilities can be written as follows

$$P\big((Y \le j|X_i)\big) = \frac{exp(\alpha_j - \sum_{k=1}^{K}\beta_k X_{ik})}{1+exp(\alpha_j - \sum_{k=1}^{K}\beta_k X_{ik})} \; And \; P\big((Y \le j-1|X_i)\big) = \frac{exp(\alpha_{j-1} - \sum_{k=1}^{K}\beta_k X_{ik})}{1+exp(\alpha_{j-1} - \sum_{k=1}^{K}\beta_k X_{ik})}$$

Having these equations the likelihood becomes

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

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

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

Therefore the log-likelihood function is:

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

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

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

Hence

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

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.

3.3.1.2. 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:

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

where, α_j are the intercept or cut points and $\beta_{1j}, \beta_{2j}, \ldots, \beta_{Kj}$ 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.

3.3.1.3. Partial proportional odds model

The partial proportional odds model (Peterson and Harrell, 1990; Fu, 1998; 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_2 has 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

$$logit[P(Y > j|X)] = \alpha_j + \sum_{k=1}^{p_1} \beta_k X_{1k} + \sum_{r=1}^{p_2} \beta_{rj} X_{2r}, j = 1, \cdots, J-1 \quad (3.8)$$

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.

3.4. 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 and Lemeshow, 2000)

$$odds(\mathrm{x} = 1) = \frac{P(Y|X=1)}{1-P(Y|X=1)} \quad \text{And} \quad odds(\mathrm{x} = 0) = \frac{P(Y|X=0)}{1-P(Y|X=0)} \quad (3.9)$$

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

$$OR = \frac{odds(\mathrm{x} = 1)}{odds(\mathrm{x} = 0)}$$

The odds of the response are multiplied by $\mathrm{OR} = e^{\beta}$ 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.

3.5. Model Selection Criteria

In regression analysis fitting a model is the main issue and we should give more care for selecting model that well fit the data. To achieve this task selection criteria's such as R-square, Adjusted R-square, Pseudo $\boldsymbol{R^2}$, BIC, AIC, etc should be considered. It is much better to compare models based on their results, reasonableness, and fit as measured; we can make comparisons among the possible models using the above selection criteria. In the case of logistic regression the model selection criteria will be taken as AIC. The AIC computation is based on the likelihood of the fit and the number of parameters in the model is considered. Therefore, if the model contains many variables there will be many parameters to be estimated; therefore, this may penalize the AIC criteria. If we fit a model that contains all the possible variables under study it needs much computation time and resources, the collinarities of the variables may affect the model fit and also less important variables might be included in the model or if the model contains few variables, it may not well explain the outcome (response) and the error of the model becomes large due to exclusion of important variables. The variables included in the model should be selected based on their significance and relationship with the outcome variables or response variable. Therefore, the issue of inclusion and exclusion of explanatory variables are called variable selection problem. Methods such as forward, backward and stepwise selection are commonly used.

In this thesis the model selection criteria used is AIC (Akaki information criterion) and the model with small value of this criterion is the optimal model, that means a model that close to actual one (Agresti, 2002) and the model which have few parameters to be estimated.AIC is defined as

$$AIC = -2\big(maximized\ L(\alpha, \beta)\big) - P)$$

Where $L(\alpha, \beta)$= log likelihood, P=the number of parameters of the model

3.6. Test of overall model fit

3.6.1. 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. 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 log-likelihood, 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 (Agresti, 2002):

$$\mathrm{G}^2 = -2\mathrm{Log}\Lambda = -2(LL_0 - LL_1), \mathrm{G}^2 \sim \chi^2_{P-(J-1)} \tag{3.10}$$

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.

3.6.2. Pseudo R^2 measures

In the linear regression model, the coefficient of determination, R^2, summarizes the proportion of variance in the dependent variable associated with the predictor (independent) variables, with larger 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 R^2statistic that has all of the characteristics of 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. McFadden's pseudo R-squared statistic is given by (McFadden, 1974):

$$R_{Mc}^2 = \frac{LL_0 - LL_1}{LL_0} \tag{3.11}$$

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

3.7. 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:

$$W = \left[\frac{\widehat{\beta_i}}{SE(\widehat{\beta_i})}\right]^2 \tag{3.12}$$

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

3.8. 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_O: 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 χ^2 and D has chi-square distribution with (n-p) degrees of freedom.

The Pearson goodness-of-fit statistic is:

$$\chi^2 = \sum\sum\left(\frac{O_{ij}-E_{ij}}{E_{ij}}\right)^2 \tag{3.13}$$

The deviance measure is:

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

where O_{ij}, E_{ij} are the observed and expected frequencies from i^{th} row and 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 and Lemeshow, 1980; Hosmer and Lemeshow, 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

$$\chi^2_{HL} = \sum_{i=1}^{g} \frac{(O_i - N_i\bar{\pi}_i)^2}{N_i\bar{\pi}_i(1-\bar{\pi}_i)} \tag{3.15}$$

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

3.9. 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.

3.9.1. 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 $\hat{\pi}_i$ 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 (Agresti, 2002):

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

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 approximately 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 (Rawlings, 1998). The standardized Pearson residual is given (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 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 i is defined as:

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

Where

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

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 (Rawlings, 1998), each squared deviance residual is a component of D^2, deviance statistic test for goodness of fit is given by

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

3.9.2. 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 et al., 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 and Lemeshow, 2000).

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

$$\text{Where} \quad \hat{\mathrm{v}}_\mathrm{i} = \mathrm{m_i}\,\hat{\pi}(\mathrm{x_i})[1-\hat{\pi}(\mathrm{x_i})]$$

And $\mathrm{x}_\mathrm{i}' = (1, \mathrm{x}_{1\mathrm{i}}, \mathrm{x}_{2\mathrm{i}}, \cdots, \mathrm{x}_{\mathrm{pi}})$ 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 et al., 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 χ^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

$$\Delta\widehat{\beta}_\iota = \left(\hat{\beta} - \widehat{\beta_{-\iota}}\right)'(X'VX)(\hat{\beta} - \widehat{\beta_{-\iota}}).$$

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

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

where r_i is the standardized residual and $h_i$is the i^{th} 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(k^{th} explanatory variable) from the deletion of a case *i*. It is computed as

$$DFBETA_{\,k(i)} = \frac{(\mathrm{X'VX})^{-1} x_{\mathrm{i}}' e_i}{1-h_i} \tag{3.21}$$

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

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

4.1. Descriptive statistics

A total sample of 15,351 women of reproductive age (15-49) was included in this study from these 10,633(69.7%) lived in rural and 4,718(30.7%) lived in urban. The prevalence of anaemia levels varies by place of residence: 77.6%, 15.9%, 5.4% and 1.1% are non-anaemic, mild, moderate and severe anaemic levels for rural women of rural residence for urban women while 85.6%, 11.4%, 2.8% and 0.4% are non-anaemic, mild, moderate and severe anaemic showing that the proportions of mild, moderate and severe anaemia levels in urban residence women are smaller than for those women who were rural residence. The proportions of mild, moderate and severe anaemic women were lower with increased levels of education and body mass index. Anaemia levels varied according to type of occupation, women who were not working have higher proportion of anaemia levels than working women. The distribution of anaemia levels by socio-economic and demographic characteristics of women of reproductive age is presented in Table 4.1.

Table 4.1: Distribution of women's anaemia levels by socioeconomic and demographic variables

Variables	Anaemia level of women								Total
	Not-anaemia		Mild		Moderate		Severe		
	Count	%	Count	%	Count	%	Count	%	
Anaemia level(Y)	12291	80.1	2215	14.4	710	4.6	135	.9	15351
Place of residence									
Rural	8252	77.6	1688	15.9	576	5.4	117	1.1	10633
Urban	4039	85.6	527	11.2	134	2.8	18	0.4	4718
Women's Education level									
No education	5818	74.8	1358	17.5	490	6.3	107	1.4	7773
Primary	5735	85.2	770	11.4	203	3.0	26	0.4	6734
Secondary and higher	738	87.4	87	10.3	17	2.0	2	0.2	844
Parity (Total children ever born)									

no child	4340	84.6	599	11.7	169	3.3	25	0.5	5133
1 – 2	2842	80.3	503	14.2	160	4.5	36	1..0	3541
3 -5	2791	77.7	566	15.7	203	5.6	44	1.2	3604
6 and above	2318	75.4	547	17.8	178	5.8	30	1.0	3073
Pregnancy Status									
No or unsure	11427	80.8	2054	14.5	545	3.9	110	0.8	14136
Yes	864	71.1	161	13.3	165	13.6	25	2.1	1215
Contraceptive methods									
No method	9874	78.3	1948	15.4	665	5.3	127	1.0	12614
Folkeric and traditional methods	113	82.5	20	14.6	4	2.9	0	0.0	137
Modern methods	2304	88.6	247	9.5	41	1.6	8	0.3	2600
Body mass index									
Less than 18.5	3261	76.4	715	16.8	241	5.6	49	1.1	4266
Between 18.5 and 24.5	8017	81.0	1369	13.8	427	4.3	83	0.8	9896
25.0 and above	1013	85.2	131	11.0	42	3.5	3	0.3	1189
Marital status									
Single	4312	85.4	588	11.6	126	2.5	26	0.5	5052
Married	6887	77.3	1393	15.6	532	6.0	102	1.1	8914
Widowed	403	74.9	105	19.5	27	5.0	3	0.6	558
Divorced	689	81.3	129	15.2	25	3.0	4	0.5	847
Women's occupation	**Type**								
Not working	5701	76.4	1205	16.1	466	6.2	93	1.2	7465
Agricultural sector	2627	82.8	427	13.5	102	3.2	18	0.6	3174
Non agriculture sector and others	3963	84.1	583	12.4	142	3.0	24	0.5	4712

The prevalence of mild anaemia levels among regions shows that the proportion of mild anaemia were 24.2%, 23.8%, 18.0% and 15.9% in Affar, Somalia, Dire Dawa and Gambela respectively (see Table 4.2) and lower proportion of mild anaemia in Addis Ababa, Tigray and followed by southern nation nationality people. The proportion of moderate anaemia were found to be high in Somali (15%), Dire Dawa (11.2%) and Affar (10.4%) compared to the other regions. The proportion of severe anaemia in Somali region (4.7%), Dire Dawa (1.8%) and Affar (1.7%) region were found to be higher compared to other regions (See Table 4.2).

Table 4.2: The distribution of anaemia levels by regions

			Anaemia level				Total
			Not-anaemic	Mild	Moderate	Severe	
Region	Tigray	Count	1455	162	38	7	1662
		%	87.5	9.7	2.3	0.4	100.0
	Affar	Count	796	302	130	21	1249
		%	63.7	24.2	10.4	1.7	100.0
	Amhara	Count	1608	288	50	8	1954
		%	82.3	14.7	2.6	0.4	100.0
	Oromiya	Count	1660	311	68	15	2054
		%	80.8	15.1	3.3	0.7	100.0
	Somali	Count	452	191	120	38	801
		%	56.4	23.8	15.0	4.7	100.0
	Benishangul-Gumuz	Count	963	175	49	5	1192
		%	80.8	14.7	4.1	0.4	100.0
	SNNP	Count	1697	170	42	7	1916
		%	88.6	8.9	2.2	0.4	100.0
	Gambela	Count	854	170	40	4	1068
		%	80.0	15.9	3.7	0.4	100.0
	Harari	Count	777	140	42	7	966
		%	80.4	14.5	4.3	0.7	100.0
	Addis Ababa	Count	1347	128	20	5	1500
		%	89.8	8.5	1.3	0.3	100.0
	Dire Dawa	Count	682	178	111	18	989
		%	69.0	18.0	11.2	1.8	100.0
Total		Count	12291	2215	710	135	15351
		%	80.1	14.4	4.6	0.9	100.0

The proportions of non-anaemia were found to be different among the wealth index category; 4403(75.2%) of poor women, 1706 (79.4%) of middle income women and 6182(84.1%) of rich women are non anaemia this shows anaemia prevalence may be related to the women economic status (Table 4.3).

The proportions of mild, moderate and severe anaemia were higher for womens with poor economic status as compared to women whose economic statuses were middle and rich.

Table 4.3: The distribution anaemia levels by wealth index

			Anaemia level				Total
			Not-anaemia	Mild	Moderate	severe	
Wealth index	Poor	Count	4403	1011	369	73	5856
		%	75.2	17.3	6.3	1.2	100.0
	Middle	Count	1706	321	98	23	2148
		%	79.4	14.9	4.6	1.1	100.0
	Rich	Count	6182	883	243	39	7347
		%	84.1	12.0	3.3	0.5	100.0
Total		Count	12291	2215	710	135	15351
		%	80.1	14.4	4.6	0.9	100.0

Table 4.4: The distribution of anaemia levels by habit of cigarette smoking

			Anaemia level				Total
			Non-anaemia	Mild	Moderate	severe	
Habit of Cigarette smoking	No	Count	12242	2210	704	134	15290
		%	80.1	14.5	4.6	0.9	100.0
	Yes	Count	49	5	6	1	61
		%	80.3	8.2	9.8	1.6	100.0
Total		Count	12291	2215	710	135	15351
		%	80.1	14.4	4.6	0.9	100.0

The anaemia levels for women according to habit of smoking cigarettes showed that 8.2% of mild anaemia, 9.8% moderate anaemia and 1.6% severe anaemia were associated with smoking cigarette (see Table 4.4).

4.2. Results of the 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 anaemia levels of women of reproductive age as the response and education levels, occupation type, place of residence, region, age, parity (total children ever born), wealth index, pregnancy status, body mass index, contraceptive methods, habit of smoking cigarette and marital status are the explanatory variables that related to anaemia based 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 all the explanatory variables are significantly associated at 15% level of significance (see Table A1). Hence, all these explanatory variables will be entered into the proportional odds model (Hosmer and Lemeshow, 2000) since all the explanatory variables are significantly associated with anaemia level.

The proportional odds model can be estimated for all significantly associated variables by descending option in logistic procedure using SAS 9.0. Results of the fit showed that education level, place of residence, region, pregnancy status, body mass index, contraceptive methods and marital status are the significant variables at 5% level of significance while the others are not significant variable at this level. It is also possible to use variable selection option in fitting POM using SAS 9.0 similar variables were significantly included in the final iteration to that of the model which includes all possible variables, cases where variable selection procedures may discard biologically important variables or variables which are directly related to the objective of the study based on Hosmer and Lemeshow (2000) it allows to include these variables. Therefore, in this case the variable parity is included in the final model in spite of the fact that it was not statistically significant at 5% level but biologically important predictor.

The fit of POM model for the selected explanatory variables was done using SAS 9.0, by including “aggregate scale=none” option to test the overall goodness fit test by the Pearson and deviance chi-squares (see Table A2).This goodness of fit tests supports that the model well fit the data that means the p-values for both goodness of fit test is greater than 5% level of significance while fitting POM perform test of the proportionality assumption using score test which has p-value less than 0.05 this showed that the assumption is violated (see Table A2). In addition, the model proportionality assumption can be tested using parallel line test in SPSS 20 which also confirms that violation of this assumption (see Table A3). Similarly the generalized ordered logit model can be fitted using STATA 12 by the command GOLOGIT2 by default. GOM is fitted as alternative model when the proportionality assumption of POM is not fulfilled by pretending the ordinarity nature of the response categories (see Table A4).

If the POM assumption is violated there are different alternatives that have been discussed in chapter three, in this study the alternative model to meet the objective of the study is partial proportional odds model. PPOM is a model that relaxes the assumption of proportionality for those variables that did violate assumption to have different effect in all logits and constrained for those explanatory variables where smaller differences are across the logits were observed (those which does not violate the assumption). The Brant test and the Wald test are used to check the assumption of proportionality for all categories of each explanatory variable in the ordinal logistic regression. Here is also another problem with the test of proportionality of individual explanatory variables using Brant test, although the given explanatory variable fulfill the assumption of proportionality, some of the categories of the explanatory variable may violate. Therefore, in order to overcome this problem the partial proportional odds model fitted by the GOLOGIT2 with option AUTOFIT (Williams, 2006) provides a good alternative model than the other models like POM (with strict assumption of parallel line for all explanatory variables) and GOM(with relaxed the assumption parallel line for all explanatory variables).

4.3. 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.

The following is the format of PPOM of anaemia levels and the estimated model is given in Table 4.5.

$$ln\left[\frac{\pi_j(X)}{1-\pi_j(X)}\right] = \alpha_j + \sum_{r=2}^{11} \beta_{1jr}\mathrm{REG}_r + \beta_{2ju}PR_u + \sum_{w=2}^{3} \beta_{3jw}\mathrm{WEL}_w + \sum_{t=2}^{4} \beta_{4jt}\mathrm{TCEB}_t + \beta_{5jp}\mathrm{PS}_p + \sum_{c=2}^{3} \beta_{6jc}\mathrm{CM}_c + \sum_{b=2}^{3} \beta_{7jb}\mathrm{BMI}_b + \sum_{m=2}^{4} \beta_{8jm}\mathrm{MS}_m$$

where, $\pi_j(X) = P(Y > j|X)$ is the sum of all probabilities of the response above the j^{th} category.

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 4.5. The categories, Addis Ababa, urban residence, no education, no child, not pregnancy, no use of contraceptive method, marital status single and BMI less than 18.5 were used as reference categories.

Table 4.5 and 4.6 provide a variety of PPOM estimates (estimates, p-value, odds ratios, 95% CI for odds ratios).

Table 4.5: PPOM model parameter estimates

Predictors		Non-anaemic		Mild		Moderate	
		Coef.	P>z	Coef.	P>z	Coef.	P>z
Region	Afar	.790	0.000	.790	0.000	.790	0.000
	Amhara	-.0563	0.636	-.412	0.017	-.440	0.254
	Oromiya	.052	0.652	.052	0.652	.052	0.652
	Somali	1.181	0.000	1.606	0.000	2.054	0.000
	Benishangul-Gumuz	-.012	0.925	-.012	0.925	-.012	0.925
	SNNP	-.564	0.000	-.564	0.000	-.564	0.000
	Gambela	.110	0.388	.110	0.388	.110	0.388
	Harari	.360	0.003	.360	0.003	.360	0.003
	Tigray	-.478	0.000	-.478	0.000	-.478	0.000
	Dire Dawa	1.005	0.000	1.491	0.000	1.448	0.000
Residence	Urban	-.448	0.000	-.671	0.000	-1.020	0.000
Education level	Primary	-.221	0.000	-.221	0.000	-.221	0.000
	Secondary and higher	-.204	0.095	-.204	0.095	-.204	0.095
Parity	1-2 children	.215	0.004	.001	0.993	.466	0.114
	3-5 children	.190	0.015	-.025	0.832	.325	0.266
	6 and above	.207	0.011	-.060	0.631	.014	0.965
Pregnancy	Yes(pregnant)	.312	0.000	1.047	0.000	.602	0.009
Contraceptive methods	Folkloric and traditional	-.010	0.965	-.010	0.965	-.010	0.965
	Modern methods	-.571	0.000	-.571	0.000	-.571	0.000
BMI	Between 18.5 and 24.9	-.178	0.000	-.178	0.000	-.178	0.000
	25.0 and above	-.456	0.000	-.456	0.000	-.456	0.000
Marital status	Married	.138	0.055	.421	0.000	.158	0.531
	Widowed	.320	0.009	.320	0.009	.321	0.009
	Divorced	.065	0.551	.065	0.551	.065	0.551
	Cons	-1.383	0.000	-3.104	0.000	-5.061	0.000

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, thirty (30) constraints have been imposed in the final model, the chi2 (30) = 25.66, with P = 0.6922 which not a significant value indicating that the final model does not violate the proportional odds or parallel lines assumption.

Table 4.6: Odds ratio estimates of PPOM for the determinants of anaemia levels

Predictors		Non-anaemic		Mild		Moderate	
		Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Region	Afar	2.203	1.749- 2.775	2.203	1.749- 2.775	2.203	1.749- 2.775
	Amhara	.945	.748-1.194	.662	.472-.929	.644	.302-1.371
	Oromiya	1.054	.839-1.323	1.054	.839-1.323	1.054	.839-1.323
	Somali	3.259	2.566- 4.140	4.983	3.792-6.548	7.803	4.969-12.254
	Benishangul-Gumuz	.988	.771 -1.266	.988	.771-1.266	.988	.771-1.266
	SNNP	.569	.445-.728	.569	.445-.728	.569	.445-.728
	Gambela	1.117	.869 -1.434	1.117	.869 -1.434	1.117	.869 -1.434
	Harari	1.433	1.126- 1.824	1.433	1.126- 1.824	1.433	1.126 - 1.824
	Tigray	.620	.484 -.794	.620	.484-.794	.620	.484-.794
	Dire Dawa	2.731	2.183-3.416	4.442	3.392-5.818	4.256	2.438-7.428
Residence	Urban	.639	.561-.727	.511	.417-.626	.360	.213-.611
Education level	Primary	.801	.721-.891	.801	.721-.891	.801	.721-.891
	Secondary and higher	.815	.641-1.036	.815	.641-1.036	.815	.641-1.036
Parity	1-2 children	1.240	1.070-1.436	1.001	.791-1.266	1.594	.894-2.842
	3-5 children	1.21	1.037-1.411	.975	.770-1.234	1.384	.781-2.455
	6 and above	1.230	1.048-1.443	.942	.737-1.203	1.014	.549-1.873
Pregnancy	Yes(pregnant)	1.366	1.184-1.577	2.849	2.365-3.431	1.825	1.160-2.871
Contraceptive methods	Folkloric and traditional	.990	.629-1.558	.990	.629-1.558	.990	.629-1.558
	Modern methods	.565	.492-.648	.565	.492-.648	.565	.492-.648
BMI	Between 18.5 and 24.9	.837	.763-.918	.837	.763-.918	.837	.763-.912
	25.0 and above	.634	.523-.767	.634	.523-.767	.634	.523-.767
Marital status	Married	1.148	.997-1.322	1.523	1.225-1.894	1.171	.715-1.920
	Widowed	1.378	1.084-1.752	1.378	1.084-1.752	1.378	1.084-1.752
	Divorced	1.067	.863-1.319	1.067	.863-1.319	1.067	.863-1.319
	Cons	.251	.198-.318	.045	.034-.059	.006	.004-.011

4.4. Marginal effects

In PPOM the probability of a single level of the response variable is not possible. The sign of the coefficients does not always determine the direction of the effect of the intermediate outcomes (Washington et al., 2003; Wooldridge, 2002). Therefore, there is a need find another way of finding the contribution of each explanatory variables on the categories of the response which can be done by computing marginal effects.

Marginal effect is used to measure the magnitude and types of association between the levels of the explanatory variable on the probability of levels of the response variable. In Table 4.7 the effect of the levels of the explanatory variables on all anaemia levels are given.

Table 4.7: Marginal effect of each predictor in the four anaemia levels probabilities

Predictors		MRI1	P>z	MRI2	P>z	MRI3	P>z	MRI4	P>z
Region	Afar	-.141	0.000	.101	0.000	.034	0.000	.006	0.000
	Amhara	.008	0.632	.005	0.732	-.011	0.010	-.002	0.178
	Oromiya	-.008	0.656	.006	0.656	.002	0.658	.0002	0.659
	Somali	-.230	0.000	.113	0.000	.088	0.000	.030	0.000
	Benishangul-Gumuz	.002	0.925	-.001	0.925	-.0003	0.924	-.0001	0.924
	SNNP	.072	0.000	-.055	0.000	-.015	0.000	-.002	0.000
	Gambela	-.017	0.401	.013	0.399	.004	0.408	.001	0.411
	Harari	-.058	0.007	.043	0.006	.013	0.011	.002	0.016
	Tigray	.062	0.000	-.047	0.000	-.013	0.000	-.002	0.000
	Dire Dawa	-.189	0.000	.086	0.000	.088	0.000	.015	0.003
Residence	Urban	.062	0.000	-.040	0.000	-.018	0.000	-.005	0.000
Education level	Primary	.032	0.000	-.024	0.000	-.007	0.000	-.001	0.000
	Secondary and higher	.028	0.076	-.021	0.078	-.006	0.070	-.001	0.073
Parity	1-2 children	-.033	0.006	.033	0.001	-.003	0.509	.003	0.162
	3-5 children	-.029	0.019	.030	0.005	-.003	0.494	.002	0.306
	6 and above	-.032	0.015	.034	0.003	-.002	0.589	.0001	0.965
Pregnancy	Yes(pregnant)	-.050	0.000	-.009	0.368	.055	0.000	.004	0.042
Contraceptive methods	Folkloric and traditional	.002	0.965	-.001	0.965	-.0003	0.965	-.0001	0.965
	Modern methods	.074	0.000	-.057	0.000	-.015	0.000	-.002	0.000
BMI	Between 18.5 and 24.9	.027	0.000	-.020	0.000	-.006	0.000	-.001	0.001
	25.0 and above	.059	0.000	-.045	0.000	-.012	0.000	-.002	0.000
Marital status	Married	-.020	0.054	.005	0.555	.014	0.000	.001	0.527
	Widowed	-.052	0.016	.038	0.014	.012	0.022	.002	0.028
	Divorced	-.010	0.558	.007	0.557	.002	0.561	.0003	0.563
		Non-anaemia		Mild		Moderate		Severe	

Women from Affar, Somali and Dire Dawa were more likely to be mild anaemia by 10%, 11% and 8.6%, respectively, compared to women in Addis Ababa; they were more likely to be moderate anaemia by 3.4%,8.8% and 8.8% respectively compared to Addis Ababa. Women from

Somali and Dire Dawa were more likely to be severe anaemia by 3% and 1.5% respectively compared to women from Addis Ababa.

Women who did not use of any contraceptive method were more likely to be mild and moderate anaemia by 5.7% and 1.5% respectively compared to those who did use modern contraceptive methods, respectively.

Women who had 1-2, 3-5 and 6 and more children were more likely to be mild anaemic by 3.3%, 3% and 3.4% respectively compared to those women who had no child. Illiterate women were more likely to be mild anaemic by 2.4% compared to women whose education level was primary. Rural women were more likely to be mild, moderate and severe anaemia by 4%, 1.8% and 0.1% respectively compared to urban women. Pregnant women were likely to be moderate anaemic by 5.5% compared to non-pregnant women.

4.5. Result of test of overall model fit

The model fit statistics (AIC, likelihood ratio (LR) test and Pseudo R^2) for the three models POM, PPOM and GOM are given in Table 4.8 and Table 4.9 below.

Table 4.8: AIC, BIC for POM, PPOM and GOM

Model	Obs	DF	AIC	BIC
POM	15,351	27	18,491.86	18,698.11
PPOM	15,351	45	18,346.76	18,690.52
GOM	15,351	75	18,378.52	18,951.44

Table 4.9: Likelihood ratio test and computed Pseudo R^2 for POM, PPOM and GOM

Model	Obs	LL(null)	LL(model)	DF	LR chi2	PROB > chi2	Pseudo R^2
POM	15,351	-9841.885	-9218.931	24	1,245.91	0.0000	0.0633
PPOM	15,351	-9841.885	-9128.382	42	1,427.01	0.0000	0.0725
GOM	15,351	-9841.885	-9114.26	72	1,455.25	0.0000	0.0739

Among the three models POM has smallest number of parameters in the final model, and PPOM model also contains fewer parameters than GOM. A model with small AIC is preferred therefore; PPOM has the smallest AIC which is 18,346.76.

The final fit for the three models are significant compared to their null model (only intercept term) evidenced from deviance LR test all the P=0.0000 and again the Pseudo R^2 POM model is the smallest and for the other two the result is almost similar but slightly larger for GOM.

The Pearson and deviance chi-square goodness of fit tests are be performed for proportional odds model. If the values of the test statistics are greater than 0.05 level of significance showing that a model fit the data well. Partial proportional odds model is a series of logit models estimated simultaneously, the only way to perform goodness of a fit by separately fitting all logits and perform goodness of fit test for each binary model.

For ordinal response variable with J ordinal category we can obtain a J-1 binary logistic regression model which makes a series of binary comparisons. For example, a four-category ordered variable (variable *Y* coded using 4 categories; 1, 2, 3, 4) can be represented as three possible separate binary comparisons.

Ana1: non-anaemia Vs mild or moderate or severe anaemia:- $ln\left[\frac{\pi_1}{1-\pi_1}\right] = \alpha_j + \sum_{r=2}^{11}\beta_{1jr}\text{REG}_r + \beta_{2ju}PR_u + \sum_{w=2}^{3}\beta_{3jw}\text{WEL}_w + \sum_{t=2}^{4}\beta_{4jt}\text{TCEB}_t + \beta_{5jp}\text{PS}_p + \sum_{c=2}^{3}\beta_{6jc}\text{CM}_c + \sum_{b=2}^{3}\beta_{7jb}\text{BMI}_b + \sum_{m=2}^{4}\beta_{8jm}\text{MS}_m$

where, π_1 is the probability of anaemia (at least mild anaemia) given that of the set of explanatory variables.

Ana2: non-anaemia or mild Vs moderate or severe anaemia:- $ln\left[\frac{\pi_2}{1-\pi_2}\right] = \alpha_j + \sum_{r=2}^{11}\beta_{1jr}\text{REG}_r + \beta_{2ju}PR_u + \sum_{w=2}^{3}\beta_{3jw}\text{WEL}_w + \sum_{t=2}^{4}\beta_{4jt}\text{TCEB}_t + \beta_{5jp}\text{PS}_p + \sum_{c=2}^{3}\beta_{6jc}\text{CM}_c + \sum_{b=2}^{3}\beta_{7jb}\text{BMI}_b + \sum_{m=2}^{4}\beta_{8jm}\text{MS}_m$

where, π_2 is the probability of anaemia (at least moderate anaemia) given that of the set of explanatory variables.

Ana3: non-anaemia or mild or moderate Vs severe anaemia:- $ln\left[\frac{\pi_2}{1-\pi_2}\right] = \alpha_j + \sum_{r=2}^{11}\beta_{1jr}\text{REG}_r + \beta_{2ju}PR_u + \sum_{w=2}^{3}\beta_{3jw}\text{WEL}_w + \sum_{t=2}^{4}\beta_{4jt}\text{TCEB}_t + \beta_{5jp}\text{PS}_p + \sum_{c=2}^{3}\beta_{6jc}\text{CM}_c + \sum_{b=2}^{3}\beta_{7jb}\text{BMI}_b + \sum_{m=2}^{4}\beta_{8jm}\text{MS}_m$

where, $\pi_3$the probability of is severe anaemia given that of the set of explanatory variables.

After fitting the above three separate binary models the Hosmer-Lemeshow goodness-of-fit test for those binary logit models were performed using SAS 9.0 (see Table A5). The three logit model results of H-L test are insignificant showing all the three binary logits well fit the data. Since H-L test cannot be used for proportional odds model candidate binary logits were well fitted. Hence, the proportional odds model that contains all binary logits also well fit the data because partial proportional odds model is a set of binary logits which are simultaneously modeled.

4.1. Interpretation of partial proportional odds model

The results in Table 4.5 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 4.6). On the basis of Table 4.6 the interpretations are given as follow.

From PPOM the non-anaemia status is compared to mild, moderate and severe anaemia status, women lived in Dire Dawa, Somali region, Affar region and Harari region had 2.73, 3.26, 2.202 and 1.33 times higher risk of being mild or moderate or severe anaemia status respectively compared with those lived in Addis Ababa. When non-anaemia and mild anaemia status are compared to moderate and severe anaemia status, women lived in Dire Dawa, Somali region, Affar region and Harari region had 4.44, 4.98, 2.202 and 1.33 times higher risk of being moderate or severe anaemia status respectively compared to women lived in Addis Ababa. When non-anaemia, mild and moderate anaemia status are compared to severe anaemia status, women lived in Dire Dawa, Somali region, Affar region and Harari region had 4.26, 7.80, 2.202 and 1.33 times higher risk of being severe anaemia status respectively compared with women lived in Addis Ababa.

In this study pregnancy status is a significant determinant of anaemia status among women of reproductive age, pregnant women were 1.366 times more likely of being in the mild or moderate or severe anaemia (as opposed to non-anaemia status) compared to non-pregnant women. Pregnant women were 2.85 times more likely of being in the moderate or severe anaemia (as opposed to non-anaemia or mild anaemia) compared with non-pregnant women.

Pregnant women were 1.83 times more likely of being severe anaemia status (as opposed to non-anaemia or mild or moderate anaemia status) compared with non-pregnant women.

Women those having 1-2, 3-5 and 6 or above children had 1.24, 1.21 and 1.23 times greater risk of being mild or moderate or severe anaemia status (as opposed to non-anaemia status) respectively compared with women who had no child, holding all other variables constant. This result is also confirmed by marginal effect, women who had 1-2, 3-5 and 6 or above children had positive marginal effect (see Table 4.7) implying that women having 1-2, 3-5 and 6 or above children were more likely of being in the mild anaemia status compared with women who had no child.

Marital status is the significant variable for anaemia status among women of reproductive age. When non-anaemia status is compared to mild, moderate and severe anaemia status, widowed women had 1.38 times more likely of being in the mild or moderate or severe anaemia compared with women whose marital status been single. Widowed women had 38% higher risk being in moderate or severe anaemia (as opposed to non-anaemia or mild anaemia status) compared with women whose marital status was single. When non-anaemia, mild and moderate anaemia status are compared to severe anaemia status, widowed women had 1.38 times higher risk of being severe anaemia status compared with women whose marital status was single., holding all other variables constant and the 95% confidence interval for the odds ratio could be as minimum as 1.08 and as maximum as 1.75.

BMI is the significant determinant of anaemia status of women, that means, the estimated odds ratio (OR=0.834) revels that women with BMI between 18.5 and 24.9 had 16% less risk of being mild or moderate or severe anaemia status (as opposed to non-anaemia status) compared to women with BMI < 18.5. Women with BMI between 18.5 and 24.9 had 16% less risk of being moderate or severe anaemia status (as opposed to non-anaemia or mild anaemia status) compared to women with BMI < 18.5. Women with BMI between 18.5 and 24.9 had 16% less risk of being severe anaemia status (as opposed to non-anaemia or mild anaemia or moderate anaemia status) compared to women with BMI < 18.5. The 95% confidence interval also suggests that odds ratio could be as minimum as 0.763 and as maximum as 0.918.

Women with BMI ≥ 25.0 were less likely of being in higher categories of anaemia. The estimated odds ratio (OR=0.634) revels that women with BMI ≥ 25.0 had 36.6% less risk of being mild or moderate or severe anaemia status (as opposed to non-anaemia or mild anaemia status) compared to women with BMI < 18.5. Women with BMI ≥ 25.0 had 36.6% less risk of being moderate or severe anaemia status (as opposed to non-anaemia or mild anaemia status) compared to women with BMI < 18.5. While women with BMI ≥ 25.0 had 36.6% less risk of being severe anaemia status (as opposed to non-anaemia or mild anaemia moderate anaemia status) compared to women with BMI < 18.5. The 95% confidence interval also suggests that odds ratio could be as minimum as 0.52and as maximum as 0.77.

Place of residence is a significant determinant of anaemia status, when non-anaemia status is compared to mild, moderate and severe anaemia status showing that urban women had 36% less risk of being in mild or moderate or severe anaemia compared with rural women. Urban women had 49 % less risk of being in the moderate anaemia or severe anaemia (as opposed to non-anaemia or mild anaemia status) compared with rural women. While urban women were 64% less risk of being severe anaemia status (as opposed to non-anaemia or mild anaemia or moderate anaemia status) compared with rural women, holding all other variables constant. That means rural women were more likely of being in higher levels of anaemia than lower levels of anaemia compared with urban women.

Education level is a significant determinant of anaemia status, when non-anaemia status is compared to mild, moderate and severe anaemia status showed that women who had attained primary school were 20% less risk of being mild or moderate or severe anaemia status compared with women who had no education. Women who had attained primary school were 20% less risk of being moderate or severe anaemia status (as opposed to non-anaemia or mild anaemia status) compared with women who had no education. Women who had attained primary school were 20% less risk of being severe anaemia status (as opposed to non-anaemia or mild anaemia or moderate anaemia status) compared with women who had no education, holding all other variable constant i.e., Not educated women were at higher risk of being in anaemia levels compared to those attained primary school.

In this study contraceptive method is a significant determinant of anaemia levels, when non-anaemia status is compared to mild, moderate and severe anaemia status showing that women who did use modern contraceptive methods had 43% less risk of being mild or moderate or severe anaemia status compared with women who did not use contraceptive methods. Women who use modern contraceptive methods had 43% less risk of being moderate or severe anaemia status compared with women who did not use contraceptive methods while women who use modern contraceptive methods had 43% less risk of being severe anaemia status (as opposed to non-anaemia or mild or moderate anaemia status) compared with women who did not use contraceptive methods, holding all other variable constant i.e., women who did not use contraceptive methods were found to be at risk to higher levels of anaemia. This result is also confirmed by marginal effect. That is, marginal effects of women who did use modern contraceptive methods compared to women who did not use contraceptive methods were negative i.e., women who did not use of contraceptive methods were at higher risk of anaemia status (Table 4.7).

4.2. 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 four categories. By collapsing the three categories into one including mild, moderate and severe anaemic these can be called anaemic. The other category will be non-anaemic. 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.

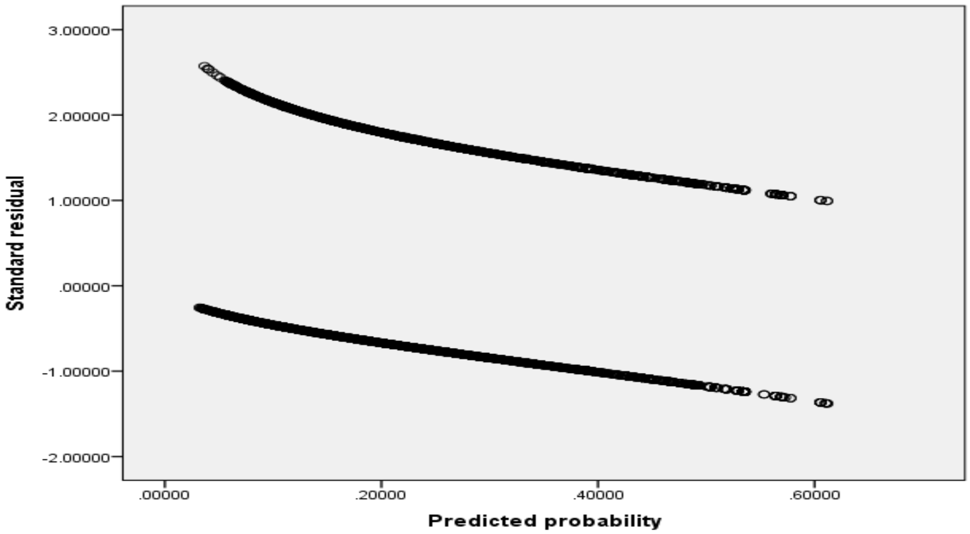

Figure 4.1: Plots of standard residual vs predicted probability.

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).

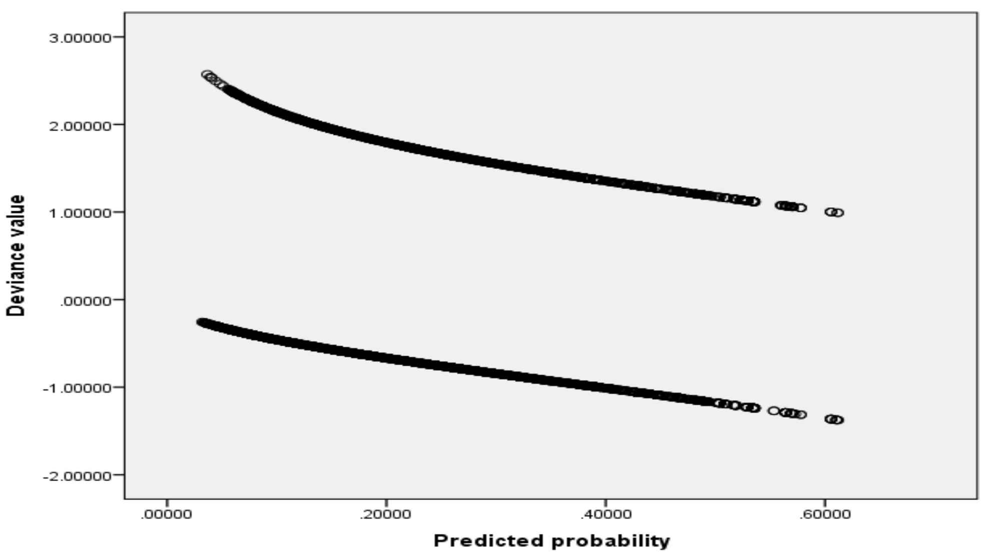

Figure 4.2: Plots of deviance residual vs predicted probability.

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.

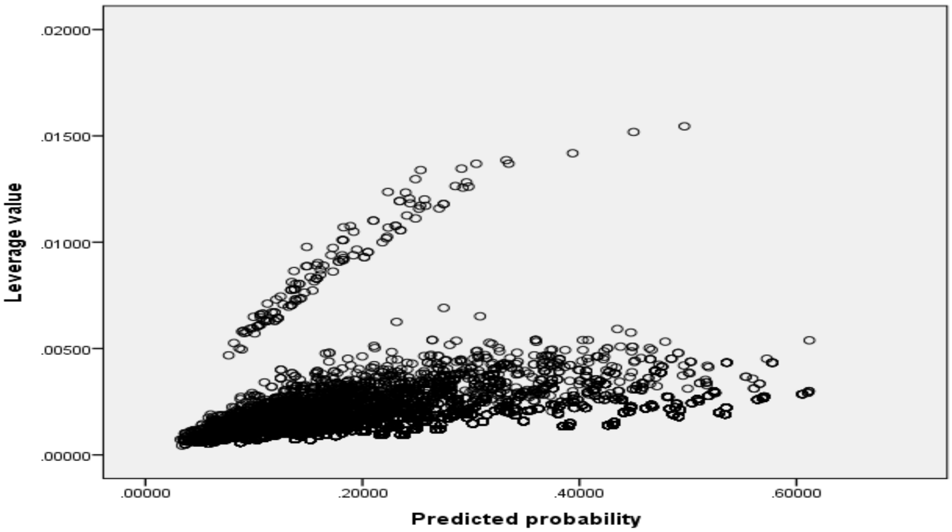

Figure 4.3: Plots of leverage value vs predicted probability.

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.

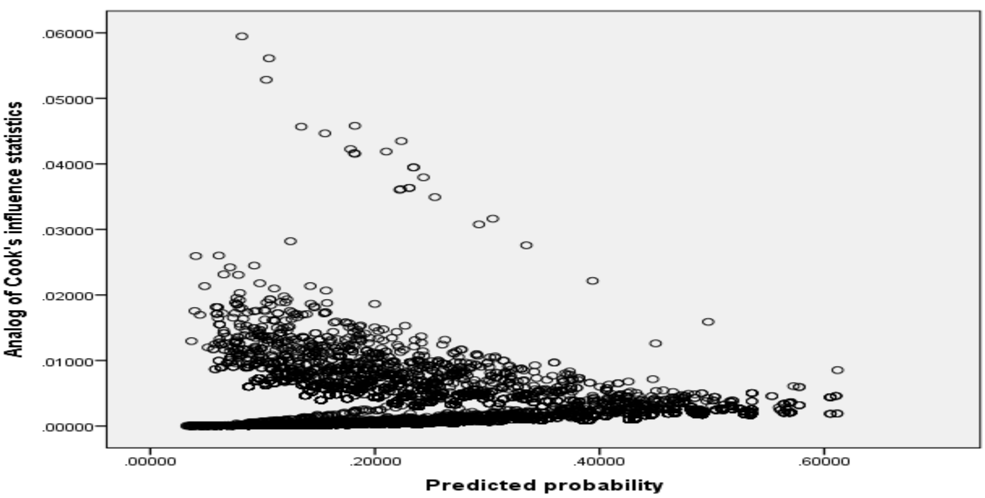

Figure 4.4: Plots of Analog of Cook's influence statistic vs predicted probabilities

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).

Plots of *DFBETA(S)* of all explanatory variables vs predicted probability are given in Figures A1-A8 where it is shown that all the DFBETA(S) of all explanatory variables are less than one. This is an indication that there is no serious problem with the fitted model.

CHAPTER FIVE

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1. Discussion

Generally speaking, we can say that the PPOM fitted the data very well; with very small P-values, relatively minimum AIC value compared to POM and GOM. A relatively large pseudo-R-squared was associated with PPOM than POM model.

In this study we considered the explanatory variables education level, occupation type, place of residence, region, age, parity (total children ever born), wealth index, pregnancy status, body mass index, and habit of smoking cigarette, contraceptive methods and marital status. However, from among these variables, education level, region, parity (total children ever born), pregnancy status, body mass index, contraceptive methods and marital status were found to be significant determinants of anaemia levels of women of reproductive age in Ethiopia.

Age of women is not significant variable in this study. This result is consistent with the result obtained by Uche-Nwachi et al (2010); Thomson et al. (1986); Kayihan and Nilgun (2008). However, the finding is not consistent with the result obtained by Ibrahim et al. (2012), Tadege (2009); Judith et al. (2008); and Yuan et al. (2009).

Occupation is not as significant determinant of anaemia in this study which is consistent with the study conducted by Tadege (2009).

Wealth index is not a significant predictor of anaemia level. This result is not consistent with the finding by Balarajan et al. (2013) result revealed that wealth index is a significant determinant of anaemia.

Place of residence was found as a significant determinant of anaemia levels. Rural women were more likely to be anaemic than urban women. This result is consistent with studies conducted by, Gebremedhin and Enquselassie (2011) who showed that rural residence was key factor predisposing women to anaemia. Getachew et al. (2012) found that rural women were highly affected by anaemia. Tadege (2009) and Yuan et al. (2009) found that residence is a significant determinant of anaemia. Sanku et al. (2010) showed urban women were at lower risk of anaemia. Wondu and Bijlsma (2012) found the same result.

In the study women who had more than one child were more exposed to severe anaemia than those with no child. This is an indication that parity is a significant determinant of anaemia. The finding is consistent result by Gebremedhin and Enquselassie (2011) who showed that high parity was a key factor predisposing to anaemia. Haider (2010) also showed that women having more than two children are associated with anaemia. Fariba et al. (2009), Uche-Nwachi et a. (2010), Balarajan et al. (2013) and Judith et al. (2008) showed that anaemia is associated with high parity. Contrary to the foregoing Sanku et al. (2010) and Ibrahim et al. (2012) showed that this is not the case.

The current study identified marital status is a significant predictor of anaemia. This finding is similar to the result by Tadege (2009) and not consistent with the result obtained by Kayihan and Nilgun (2008) and Ibrahim et al. (2012).

In this study body mass index was found to be highly related with prevalence of anaemia. Women with low body mass index were more likely to be in higher levels of anaemia (as opposed to lower levels of anaemia). This result is consistent with the result of studies by Melku et al. (2014) and Bentley and Griffiths (2003); and not consistent with the result by Kayihan and Nilgun (2008) and Thomson et al. (1986). Women with BMI <18.5 are 37% more likely to be in severe anaemia than those who have BMI >25. This result is similar to the finding by Aabroo et al. (2012) that showed those suffering from severe anaemia had BMI <17.

Women who did not use any contraceptive method were at higher risk of being in one of the levels of anaemia. This disagrees with the result obtained by Haider (2010) which concluded the contrary. However, our finding is similar to Wondu and Bijlsma (2012) who showed that women who did not use contraceptive methods were at higher risk of being anaemic than who did use modern methods.

This study showed that illiterate women were at higher risk of being anaemic, consistent with the results obtained by Balarajan et al. (2013), Ndukwu and Dienye (2012), Hamid et al. (2001) and Suega et al. (2002) who found that educational level is a significant predictor of anaemia and severe anaemia. The study results obtained by Ibrahim et al. (2012); Kayihan and Nilgun (2008) conclude the opposite.

The result of the current study showed that the chance of being in mild or moderate or severe anaemic (relative to none anaemic women) was higher by 36.6% compared to women who were not pregnant. The chance of being in severe anaemic (relative to non-anaemic or mild or moderate womens) was higher by 83% compared to the non-pregnant women. The finding is similar to those by Getachew et al. (2012); Melku et al. (2014); Tadege (2009); Priyanka (2011); Suega et al (2002); Judith et al. (2008); Ibrahim et al. (2012); Ndukwu and Dienye (2012); Yuan et al. (2009). Sanku et al. (2010) which showed that pregnancy is a significant determinant of anaemia. Fariba et al. (2009) suggested that anaemia risk increased with pregnancy and causal more pregnancy complications.

5.2. Conclusions

The main objective of this study was to identify determinant of anaemia among women of the reproductive age group in Ethiopia. The variables education level, region, parity (total number of children ever born), pregnancy status, body mass index, place of residence, contraceptive methods and marital status were found to be the significant determinants anaemia.

Women from Affar, Somali and Dire Dawa regions were at higher risk of being mild, moderate and severe anaemic. Rural residence, low body mass index and widowed women were found to be at higher risk of mild, moderate and severe anaemic. Those who did not use of any contraceptive method were more likely to be mild and moderate anaemic. Women with a larger number of children and had no education were at higher risk of mild anaemic. Pregnancy was found to be a risk factor for women to have moderate and severe anaemic.

5.3. Recommendations

Based on the results of this study the following recommendations are suggested.

- Since the prevalence of anaemia differs among regions, the Ministry of Health should give special attention to regions Dire Dawa, Somali, Affar and Harari.
- Since women with higher number of children were at higher risk of anaemia, we recommend that health centers should teach women about family planning.
- Uses of modern contraceptive methods decrease the risk of anaemia. Hence, concerted effort should be made to create the awareness and understanding the use of contraceptive methods.
- The government and other concerned bodies should pay attention to the above and make appropriate efforts to tackle problems that contribute to anaemia.

References

A2Z, FANTA project. (2006). Maternal anaemia: A preventable killer [Internet]. Available from: http://www.aed.org/publications/upload/FANTAanaemia2006.pdf

Aabroo T., Khand A., Laghari, A. (2012). Prevalence of anemia in adolescent girls. Department of Physiology, University of Sindh, Jamshoro, Pakistan. Pak J Physiol, 8(Suppl 1) P-BN-01.http://www.pps.org.pk/PJP/8-1S.pdf .

Agresti A. (2002). Categorical data analysis. John Wiley & Sons, New York, 2nd edition.

Balarajan, Yarlini S, Wafaie W Fawzi, and S V Subramanian. (2013). "Changing patterns of social inequalities in anaemia among women in India: cross-sectional study using nationally representative data." BMJ Open 3 (3): e002233.doi:10.1136/bmjopen-2012-002233. http://dx.doi.org/10.1136/bmjopen-2012-002233.

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

Bentley ME and Griffiths PL, (2003). The burden of anemia among women in India. European journal of clinical nutrition .57, 52 – 60. doi:10.1038/sj.ejcn.1601504

Berhane Y., Gossaye Y., Emmelin M., Hogberg H. (2001).Women's health in rural setting in societal transition in Ethiopia. Soc Sci Med; 53: 1525–39.

Brabin BJ. (1985). The epidemiology of infection in pregnancy. Rev Infect Dis; 7:579-603.

Brabin BJ., Ginny M., Sapau J., Galme K., Paino J. (1990), Consequences of maternal anaemia on outcome of pregnancy in a malaria endemic area in Papua New Guinea. Ann Trop Med Parasitol; 84:11-24.

Brabin BJ., Hakimi M., Pelletier D. (2001). An analysis of anaemia and pregnancy related maternal mortality.J Nutr, 131(25-2):6045-145; discussion 6145

Brick T and Peters MJ. (2014). Kathryn Maitland on treating severe anaemia in sub-Saharan Africa; *BMC Medicine* 2014, **12**:68.

Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. Higher Education: Handbook of Theory and Research, Vol. 10, 225-256.

Cook JD. (1994).The effect of endurance training on iron metabolism. Seminars in Hematology 31, 146–154.

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

Cook, R. D. (1977). Detection of influential observations in linear regression. Technometrics.

Cook, R. D. (1979). Influential observations in linear regression. Journal of the American.

EDHS. (2011). Ethiopian demographic and health Survey; Addis Ababa, Ethiopia.

Fariba R., Roonak S., Farzaneh Z. (2009).Prevalence of low haemoglobin and its correlation with maternal factors and pregnancy outcomes in pregnant women in health centers of Sanandaj in 2009.life science journal, 10(2s).http://www.lifesciencesite.com.

Federal Ministry of Health (2004), National guideline for control and prevention of micronutrient deficiencies. Addis Ababa: Family Health Department, Federal Ministry of Health, Government of Ethiopia, p. 26.

Fleming AF. (1989b), Tropical obstetrics and gynaecology. 1. Anaemia in pregnancy in tropical Africa. Trans R Soc Trop Med Hyg; 83:441-8.

Fox J. (1984), Linear statistical methods and related methods with application to social research, Wiley, New York.

Fu Vincent. (1998). Estimating Generalized Ordered Logit Models. Stata Technical Bulletin 44: 27-30, In Stata Technical Bulletin Reprints, vol 8, 160-164. College Station, TX: Stata Press.

Gebremedhin S., Enquselassie F. (2011) Correlates of anemia among women of reproductive age in Ethiopia: Evidence from Ethiopian DHS 2005. Ethiop J Health Dev; 25(1):23-30

Geelhoed D., Agedzi F., Visser L., Ablordeppey E., Asare K., O'Rourke P et al.(2006) Maternal and fetal outcome after severe anaemia in pregnancy in the rural Ghana. Acta Obstet Gynecol Scand.; 85(1):49-55.

Getachew M., Yewhalaw D., Tafess K., Getachew Y., Zeynudin A. (2012) Anaemia and associated risk factors among pregnant women in Gilgel Gibe dam area, Southwest Ethiopia. Parasites & Vectors, 5:296, doi: 10.1186/1756-3305-5-296

Gibson MA, Mace R. (2006). An energy-saving development initiative increases birth rate and childhood malnutrition in rural Ethiopia. PLoS Med 3:e87. [PMC free article] [PubMed]

Haider J. (2010). Prevalence of Anaemia, deficiencies of Iron and Folic Acid and Their Determinants in Ethiopian Women. J Health Popul Nutr; 28(4): 359–368

Haider J., Muroki N.M., Omwega A.M., and Ayana G. (2003), Malnutrition and iron deficiency in lactating women in urban slum communities from Addis Ababa, Ethiopia. East Afr Med J. 80: 191–4.

Haider J., Nekatibeb H., Urga K. (1999). Iron deficiency anemia in pregnant and lactating mothers in rural Ethiopia. East Afr Med J; 76:618–22.

Haider J., Pobocik R. (2009). Iron deficiency anemia is not a rare problem among women of reproductive ages in Ethiopia: a community based cross sectional study. BMC Blood Disorders, 9:7.doi: 10.1186/1471-2326-9-7.

Hamad I., Jamil Q., Iman D.(2001) The Effect of a Health Education Program On Prevalence Anemia among Pregnant Women in Tafila Governorate.

Harrison KA., Rossiter CE. (1985), Maternal mortality. Br J Obstet Gynaecol, 83:449 53.

Hosmer DW. And Lemeshow S. (2000). Applied logistic regression (2nd ed.). New York: Wiley & Sons.

Hosmer, D. W., and Lemeshow, S. (1980). A goodness-of-fit test for the multiple logistic regression model. Communications in Statistics, **A10**, 1043–1069.

Hurtado A., Merino C., Delgado E. (1945). Influence of anoxemia on haematopoietic activities. Archives of Internal Medicine; 75:284–323.

Ibrahim Isa., Kemebradikumo P., Dennis A. (2012) .The Burden of Anaemia among Pregnant Women at Booking in the Niger Delta of Nigeria. Online Journal of Medicine and Medical Science Research ISSN 2277-0879; Volume 1 Issue 5, Pages 91-95.Available Online at http://www.onlineresearchjournals.org/JMMSR

INACG (2002) .Why iron is important and what to do about it: A new perspective. From: http://www.idpas.org/pdf/2071WhyIronisImportantandWhattoDO.pdf Date of access [cited 21/06/2010].

IPHN. (2007). National Strategy for anaemia prevention and control in Bangladesh. Ministry of Health and Family Welfare Government of the People's Republic of Bangladesh.

Judith A., Aparna B., H. Vinod. (2008) Prevalence of anaemia among pregnant women: a community-based study in Udupi district. Health and Population-Perspectives and Issues 31 (1): 31-40.

Kalaivani K. (2009), Prevalence and consequences of anaemia in pregnancy. Indian J Med Res.; 130(5):627-33.

Kayihan P., Nilgun D. (2008) Prevalence & risk factors of anaemia among women of reproductive age in Bursa, Turkey. Indian J Med Res 128, 282-286

Koller O. (1982). The clinical significance of hemodilution during pregnancy, Obstetrical and Gynecological Survey, 37:649–652.

McCullagh P. (1980). Regression models for ordinal data (with discussion). Journal of the Royal Statistical Society Series B, 42, 109-142.

McCullagh P. and Nelder J. A. (1989), Generalized linear models (2nd Ed.). London: Chapman and Hall.

McFadden D. (1974), Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.); Frontiers in Econometrics (pp. 104-142), New York: Academic Press.

Melku M., Addis Z., Alem M. and Enawgaw B. (2014) . Prevalence and predictors of maternal anemia during pregnancy in Gondar, Northwest Ethiopia. An institutional based cross sectional study. Available from: http://dx.doi.org/10.1155/2014/108593

Micronutrient Initiative (2006), Ethiopian country profile [document on the internet]. The institute; Available from: http://www.micronutrient.org.

Ndukwu G.U., Dienye P.O. (2012) .Prevalence and socio-demographic factors associated with anaemia in pregnancy in a primary health centre in Rivers State, Nigeria. Afr J Prm Health Care Fam Med, 4(1), Art. #328, 7 pages. http:// dx.doi.org/10.4102/phcfm. v4i1.328

Nordenberg D, Yip R, Binkin NJ. (1990). The effect of cigarette smoking on hemoglobin levels and anemia screening. Journal of the American Medical Association, 264:1556–1559.

Numbare S., Maindarker G., Darade R., Yenge S., Tolani M. and Patole K. (2012) Maternal risk factors associated with term low birth weight neonates: A matched-pair case control study. Indian Journal of Paediatrics' [online] 49(1): pp. 25-28. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11291789 [Accessed on 23rd February 2012].

Peter R, Kumar R, Sangwan L, Pandey S. (2012).Prevalence of anemia and its correlation to body mass index: study among unmarried girls. International Journal of Basic and Applied Medical Sciences ISSN: 2277-2103 (Online) An Online International Journal Available at http://www.cibtech.org/jms.htm 2012 Vol. 2 (3) September - December, pp. 58-62.

Peterson, B. L., and Harrell, F. E. (1990). Partial Proportional Odds Models for Ordinal Response Variables. *Applied Statistics*, *39*(*3*), 205-217.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, 9, 705–724.

Priyanka S., Dr. Seema M., Dr. Ranjana N. (2013) Prevalence of anemia and socio-demographic factors associated with anemia among pregnant women attending antenatal Hospital in Jaipur City, India. IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS) e-ISSN: 2278-3008, p-ISSN: 2319-7676. Volume 6, Issue 3, PP 01-05. www.iosrjournals.org

Rawlings O., Pentula G., Dickey A. (1998) Applied regression analysis: Department of Statistics, North Carolina State University.

Sanku D., Sankar G., Madhuchhanda G. (2010) Prevalence of anaemia in women of reproductive age in Meghalaya: A logistic regression analysis. Turk J Med Sci, 40(5): 783-789.

Scholl T.O., Hediger M.L., Fischer RL. & Shearer J.W. (1992), Anemia vs. iron deficiency: increased risk of preterm delivery in a prospective study. American Journal of Clinical Nutrition 55, 985–988.

Sharman A. (2000). ORC Macro, editor. Anemia testing in population based surveys. Calverton, Maryland USA. Statistical Association, 74, 169–174.

Stoltzfus RJ. (2001) Iron-deficiency anemia: reexamining the nature and magnitude of the public health problem. Summary: implications for research and programs. J Nutr; 131(2S 2):697S-700S; discussion 700S.

Stoltzfus RJ. (2003).Iron deficiency: global prevalence and consequences. Food Nutr Bull; 24(4 Suppl):S99-103.

Stoltzfus RJ. (2011). Iron interventions for women and children in low-income countries. J Nutr; 141: 756S-762S; doi:10.3945/jn.110.128793

Suega K, Dharmayuda TG, Sutarga IM, Bakta IM (2002). Iron-deficiency anemia in pregnant women in Bali, Indonesia: a profile of risk factors and epidemiology; Southeast Asian J Trop Med Public Health; 33(3):604-7.

Tadege B. (2009). Determinants of anaemia in pregnant women with emphasis on intestinal helminthic infection at bushulo health center southern Ethiopia.

Thomson K., Riis B., Kerbbe S., Christiansen C. (1986). Testosterone regulates the haemoglobin concentration in male puberty. Acta Paediatrica Scandanvian 75 793-796.

Uche-Nwachi EO, Odekunle A and Singh R. (2010) Anaemia in pregnancy: associations with parity, abortions and child spacing in primary healthcare clinic attendees in Trinidad and Tobago.Afr Health Sci, 10(1): 66–70.

UNFPA (2012). Mortality In-depth Analysis of the EDHS 2000-2011 December. Trends in Maternal Health in Ethiopia, December 2012 Addis Ababa.

Walker, S.H. and Duncan, D.B. (1967). Estimation of the probability of an event as a function of several independent variables, *Biometrika,* vol. 54, no. 1/2, pp. 167-179.

Washington S., Karlaftis M., Mannering F. (2003), Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC, Boca Raton, FL.

WHO. (1993). Division of Family Health; Prevention and management of severe anaemia in the pregnancy, Reported of Technical Working Group, WHO, Geneva.

WHO. (1992). The prevalence of anaemia in women: a tabulation of available information, World Health Organization, Geneva.

WHO. (2001). Iron deficiency anaemia, assessment, prevention and control, A Guide for Programme Managers, World Health Organization. Geneva.

WHO. (2008). Centers for Disease Control and Prevention, Worldwide Prevalence of Anemia: World Health Organization Global Database of Anemia.

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

Wondu T, M Bijlsma. (2012). "The hidden hunger": understanding the burden of anaemia and its determinants among pregnant and non-pregnant women in Ethiopia. African journal of food, agriculture, nutrition and development: Volume 12 No. 7, 6913-6930

Wooldridge J.M. (2002), Econometric analysis of cross section and panel data. MIT Press.

Yuan X., Hong Y., Shaonong D., Bianba Z., Xiaoyan Z., Duolao W.(2009).Haemoglobin levels and anaemia evaluation during pregnancy in the highlands of Tibet: a hospital-based study.BMC Public Health, 9:336 .doi:10.1186/1471-2458-9-336

Appendix

Table A1: Chi- square test of association between explanatory variables with response variable

Explanatory variables	Pearson chi-square (P-value)
Age of women (AGE)	0.000
Region of women(REG)	0.000
Place of residence(PR)	0.000
educational level(WEL)	0.000
Wealth index(WI)	0.000
Parity(TCEB)	0.000
Pregnancy Status(PS)	0.000
Contraceptive methods(CM)	0.000
Body Mass Index(BMI)	0.000
Marital status(MS)	0.000
Habit of Cigarette(HC)	0.13
Occupation type(WOT)	0.000
All these variables significantly associated at 15%	

Table A2: POM maximum likelihood estimates, goodness-of-fit statistics and score test of proportionality using SAS 9.0

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 4	1	-4.8327	0.1469	1082.3912	<.0001
Intercept 3	1	-2.8996	0.1239	547.5645	<.0001
Intercept 2	1	-1.3538	0.1206	126.0546	<.0001
Afar	1	0.7476	0.1179	40.1944	<.0001
Amhara	1	-0.1117	0.1194	0.8751	0.3496
Oromiya	1	0.0105	0.1165	0.0080	0.9285
Somali	1	1.2628	0.1204	110.0766	<.0001
Benishangul-Gumuz	1	-0.0568	0.1269	0.2006	0.6543
SNNP	1	-0.6089	0.1260	23.3549	<.0001
Gambela	1	0.0732	0.1282	0.3256	0.5683
Harari	1	0.3369	0.1234	7.4532	0.0063
Tigray	1	-0.5159	0.1263	16.6945	<.0001
Dire Dawa	1	1.0606	0.1133	87.6098	<.0001
Urban	1	-0.4879	0.0659	54.7691	<.0001
Primary	1	-0.2235	0.0543	16.9563	<.0001
Secondary and higher	1	-0.2025	0.1230	2.7118	0.0996
1-2 children	1	0.2006	0.0743	7.2768	0.0070
3-5 children	1	0.1775	0.0776	5.2246	0.0223
6 and above	1	0.1846	0.0808	5.2198	0.0223
Yes(pregnant)	1	0.4480	0.0702	40.6850	<.0001
Folkloric and traditional	1	-0.00875	0.2343	0.0014	0.9702
Modern methods	1	-0.5634	0.0707	63.4583	<.0001
Between 18.5 and 24.9	1	-0.1782	0.0470	14.3456	0.0002
25.0 and above	1	-0.4720	0.0979	23.2404	<.0001
Married	1	0.1579	0.0716	4.8605	0.0275
Widowed	1	0.3136	0.1226	6.5429	0.0105
Divorced	1	0.0670	0.1085	0.3819	0.5366

Goodness- of-fit statistics

Criteria	Value	Df	Value/df	Pr>Chisq
Deviance	3854.7504	6666	0.5783	1.0000
Pearson	6441.2260	6666	0.9663	0.9752

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
241.6795	48	<.0001

Table A3: Test of parallel lines (using SPSS 20) for POM

Test of Parallel Lines[a]				
Model	-2 Log Likelihood	Chi-Square	Df	Sig.
Null Hypothesis	6197.586			
General	5988.245	209.341	48	.000
The null hypothesis states that the location parameters (slope coefficients) are the same across response categories. Link function: Logit.				

Table A4: GOM for determinants of anaemia levels on womens of reproductive age

	Anaemia level						
		Coef.	P>z	Coef.	P>z	Coef.	P>z
Region	Afar	.775	0.000	.858	0.000	-.265	0.653
	Amhara	-.053	0.657	-.418	0.111	-1.363	0.034
	Oromiya	.061	0.598	-.079	0.751	-.698	0.240
	Somali	1.184	0.000	1.573	0.000	1.038	0.064
	Benishangul-Gumuz	-.004	0.977	-.084	0.750	-1.428	0.041
	SNNP	-.562	0.000	-.552	0.038	-1.511	0.020
	Gambela	.122	0.343	-.019	0.943	-1.328	0.067
	Harari	.357	0.004	.404	0.121	-.496	0.441
	Tigray	-.474	0.000	-.488	0.068	-1.268	0.051
	Dire Dawa	1.005	0.000	1.473	0.000	.467	0.408
Residence	Urban	-.450	0.000	-.619	0.000	-.908	0.005
Education level	Primary	-.216	0.000	-.244	0.010	-.669	0.014
	Secondary and higher	-.196	0.111	-.352	0.175	-.709	0.367
Parity	1-2 children	.214	0.004	-.003	0.984	.731	0.034
	3-5 children	.191	0.015	-.041	0.747	.562	0.110
	6 and above	.210	0.010	-.080	0.544	.240	0.515
Pregnancy	Yes(pregnant)	.312	0.000	1.038	0.000	.607	0.011
Contraceptive methods	Folkloric and traditional	.010	0.965	-.348	0.499	-13.19	0.986
	Modern methods	-.567	0.000	-.726	0.000	-.646	0.088
BMI	Between 18.5 and 24.9	-.172	0.000	-.223	0.005	-.185	0.338
	25.0 and above	-.442	0.000	-.571	0.001	-1.467	0.017
Marital status	Married	.137	0.058	.452	0.000	-.236	0.472
	Widowed	.311	0.012	.483	0.032	-.690	0.297
	Divorced	.066	0.546	.089	0.685	-.470	0.408
	Cons	-1.393	0.000	-3.055	0.000	-3.803	0.000
		None anemic		Mild		Moderate	

Table A5: Three binary logit models and there Hosmer and Lemeshow goodness-of-fit test

	Anaemia level						
		Ana1		Ana2		Ana3	
		Coef.	P>z	Coef.	P>z	Coef.	P>z
Region	Afar	0.775	<.0001	0.7952	0.0012	-0.1534	0.7949
	Amhara	-0.048	0.6889	-0.4523	0.0887	-1.3281	0.0407
	Oromiya	0.066	0.5743	-0.1537	0.5455	-0.7014	0.2415
	Somali	1.188	<.0001	1.5339	<.0001	1.1222	0.0461
	Benishangul-Gumuz	0.002	0.9905	-0.1579	0.5573	-1.4006	0.0468
	SNNP	-0.559	<.0001	-0.6395	0.0185	-1.4438	0.0292
	Gambela	0.123	0.3365	-0.1074	0.6977	-1.3715	0.0633
	Harari	0.360	0.0035	0.4451	0.0899	-0.2866	0.6533
	Tigray	-0.469	0.0002	-0.5642	0.0393	-1.2453	0.0588
	Dire Dawa	1.005	<.0001	1.5220	<.0001	0.6212	0.2688
Residence	Urban	-0.447	<.0001	-0.6779	<.0001	-1.0319	0.0018
Education level	Primary	-0.214	<.0001	-0.2231	0.0252	-0.6371	0.0148
	Secondary and higher	-0.188	0.1258	-0.3434	0.1899	-0.7277	0.3432
Parity	1-2 children	0.215	0.0042	0.0219	0.8688	0.7283	0.0329
	3-5 children	0.190	0.0158	-0.0307	0.8200	0.5435	0.1213
	6 and above	0.210	0.0102	-0.1181	0.4003	0.1805	0.6248
Pregnancy	Yes(pregnant)	0.312	<.0001	1.0543	<.0001	0.5810	0.0158
Contraceptive methods	Folkloric and traditional	0.016	0.9444	-0.2910	0.5748	-11.921	0.9785
	Modern methods	-0.568	<.0001	-0.7470	<.0001	-0.6365	0.0988
BMI	Between 18.5 and 24.9	-0.172	0.0003	-0.2489	0.0027	-0.1874	0.3268
	25.0 and above	-0.444	<.0001	-0.6014	0.0008	-1.3542	0.0270
Marital status	Married	0.138	0.0570	0.4826	0.0003	-0.1776	0.5873
	Widowed	0.314	0.0110	0.4747	0.0405	-0.6462	0.3281
	Divorced	0.072	0.5049	0.0914	0.6816	-0.5809	0.3147
	Cons	-1.400	<.0001	-3.0134	<.0001	-3.8808	<.0001
		H-L test p-value=0.2899		H-L test p-value=0.6758		H-L test p-value=0.0795	

Plots of DFBETA(S) with each explanatory variable

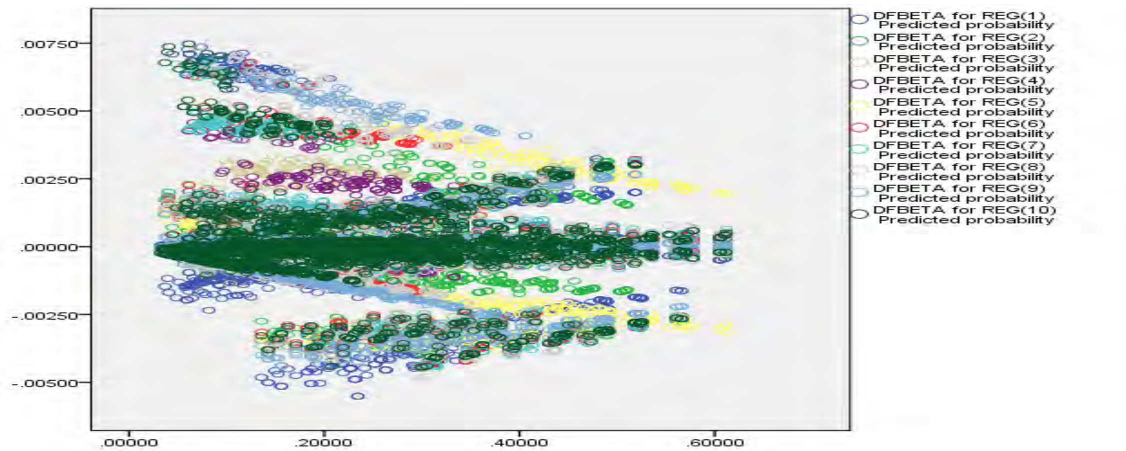

Figure A1: Plots of DFBETA(S) for Regions vs predicted probability

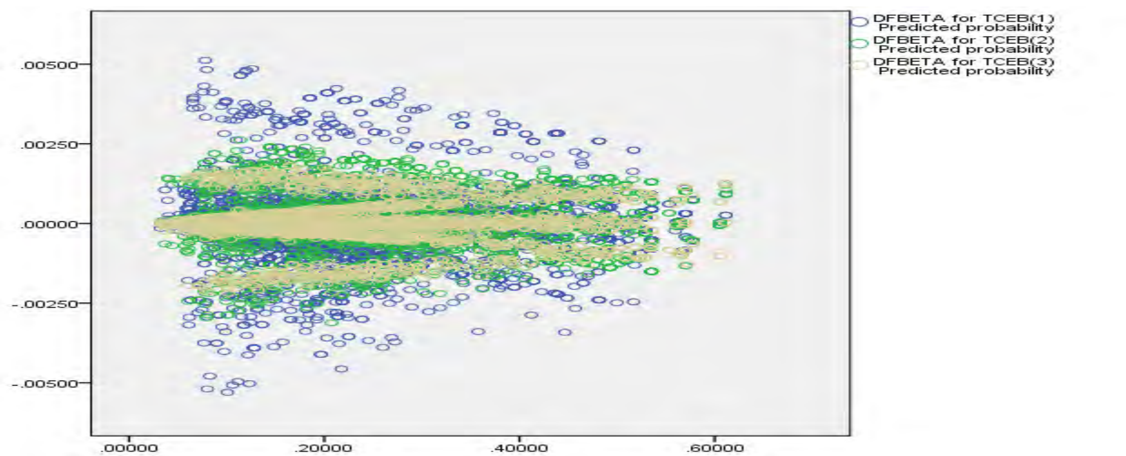

Figure A2: Plots of DFBETA(S) for TCEB vs predicted probability

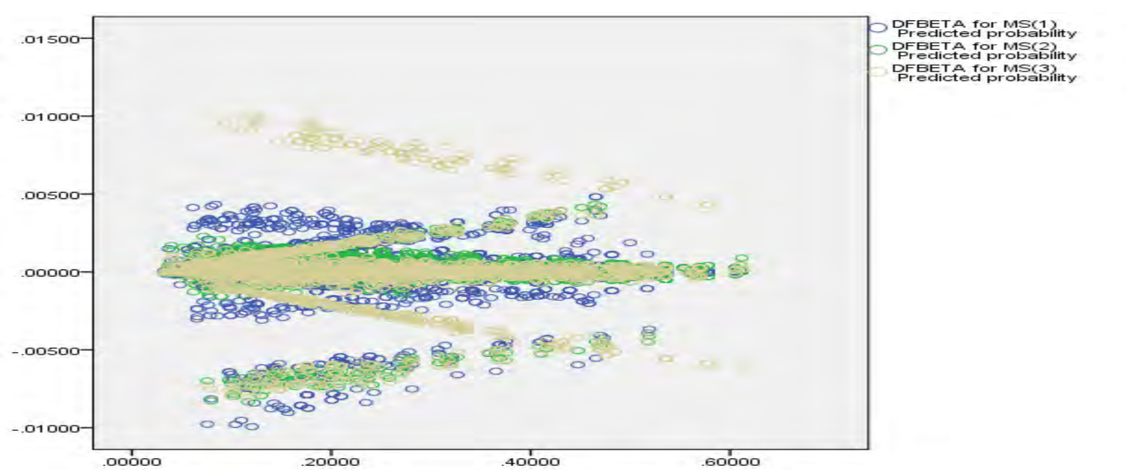

Figure A3: Plots of DFBETA(S) for marital status by predicted probability

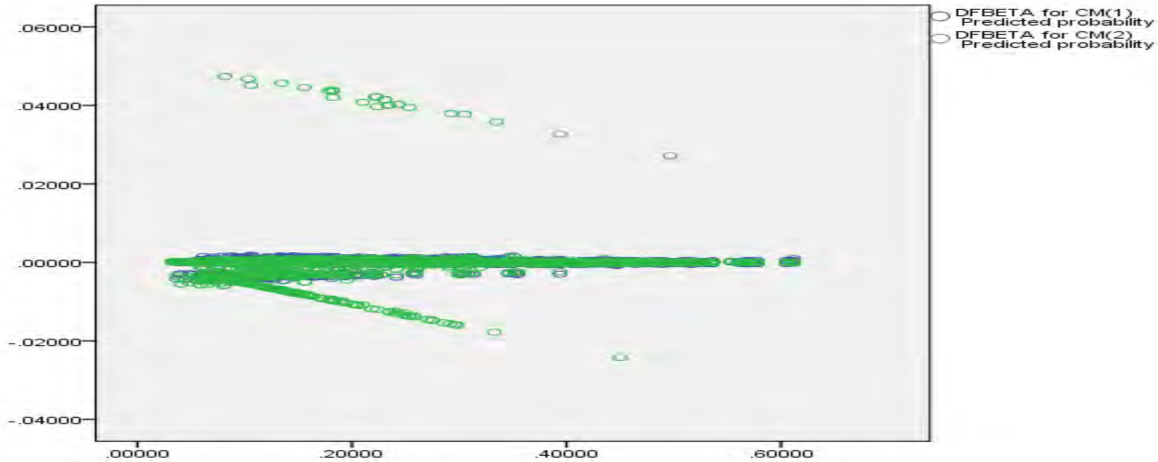

Figure A4: Plots of DFBETA(S) for contraceptive methods vs predicted probability

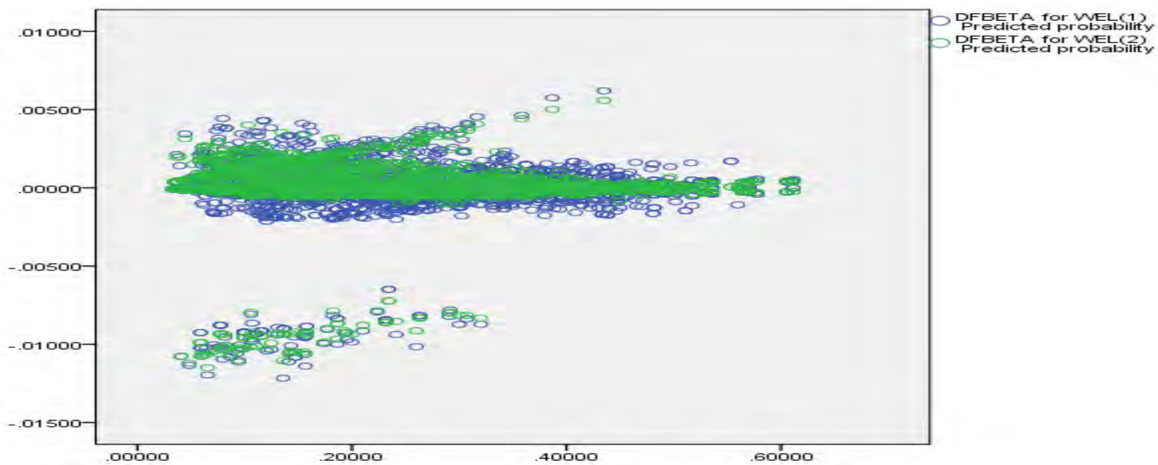

Figure A5: Plots of DFBETA(S) for Education levels vs predicted probability

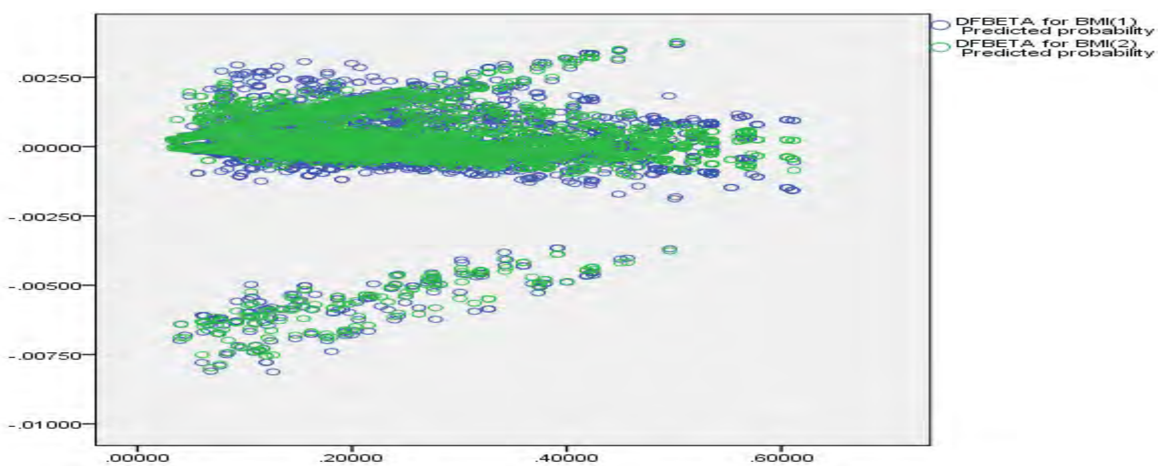

Figure A6: Plots of DFBETA(S) for Body mass index by predicted probability

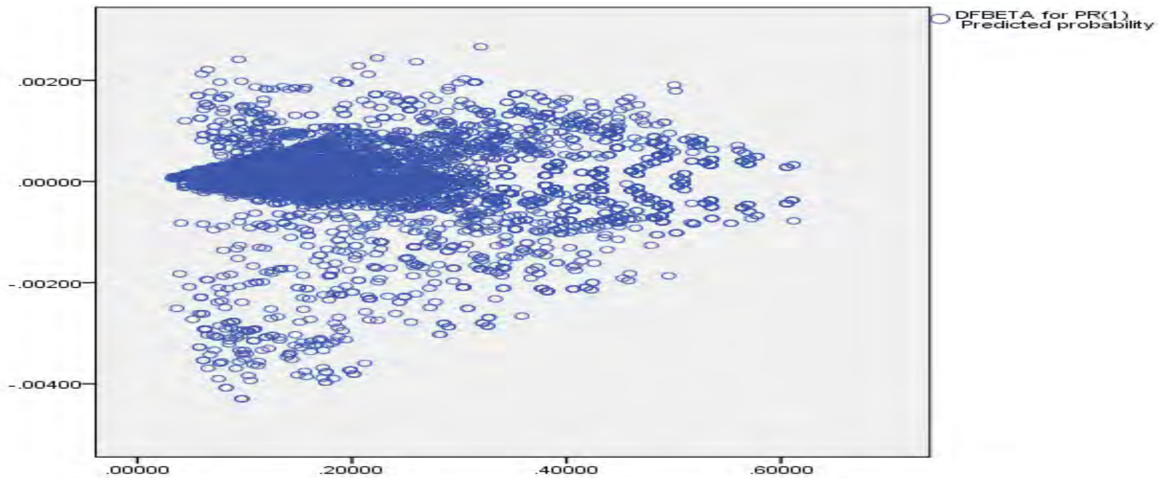

Figure A7: Plots of DFBETA(S) for Place of residence vs predicted probability

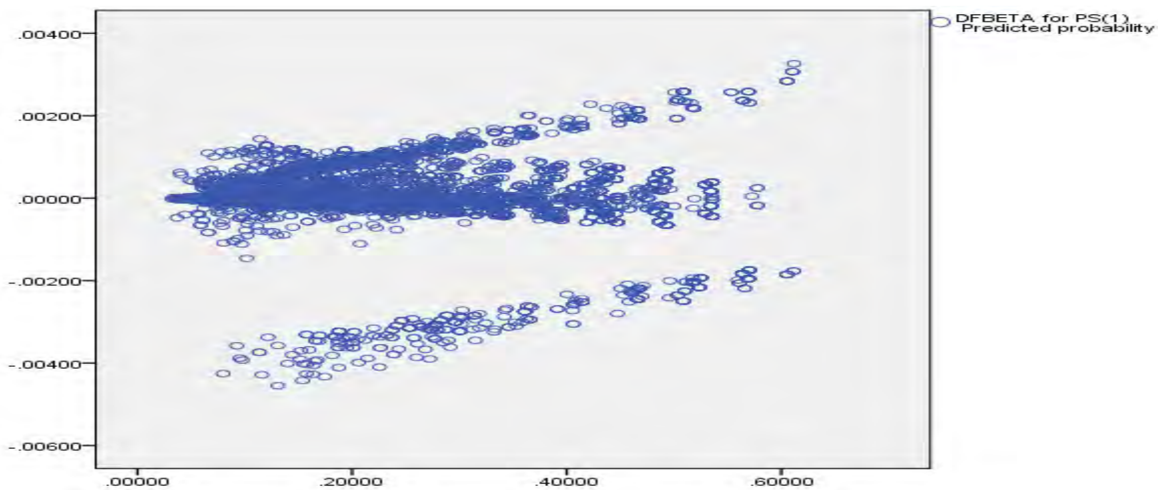

Figure A8: Plots of DFBETA(S) for Pregnancy status vs predicted probability

DECLARATION

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

Declared by:

Name: **Birhane Zelalem**

Signature: ___________________

Date: _______________________

Confirmed by Advisor:

Name: **Dr. Shibru Temesgen**

Signature: ___________________

Date: _______________________