COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF STATISTICS

BAYESIAN GEO-ADDITIVE MODEL TO ANALYSE THE SPATIAL PATTERN AND DETERMINANTS OF CHILDHOOD ANAEMIA IN ETHIOPIA

BY

KEDIR MOKONIN KENO

A THESIS SUBMITTED TO DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES, COLLEGE OF NATURAL SCIENCE, JIMMA UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS IN BIOSTATISTICS.

MARCH,2021

JIMMA, ETHIOPIA

BAYESIAN GEO-ADDITIVE MODEL TO ANALYSE THE SPATIAL PATTERN AND DETERMINANTS OF CHILDHOOD ANAEMIA IN ETHIOPIA

BY

KEDIR MOKONIN KENO

ADVISOR: - GEREMEW MULETA (Ph.D. Scholar)

CO-ADVISOR: - YASIN NEGASH (MSc)

MARCH, 2021

JIMMA, ETHIOPIA

Approval Sheet

This is to certify that the thesis titled " **Bayesian Geo-additive model to analyses the Spatial pattern and Determinants of Childhood Anaemia in Ethiopia**" submitted in partial fulfillment of the requirement for the degree of Master of Science in Biostatistics to the college of Natural science Jimma University, and is the record of original research carried out by **Kedir Mokonin Keno ID No RM0193/11**, under my supervision and no part of the thesis has been submitted for another degree or diploma. The assistance and the help received during this investigation have been duly acknowledged. Therefore, I recommended that would be accepted as fulfilling the thesis requirement.

Geremew Muleta (PhD Scholar)	.....	24-03-21
Main advisor	Signature	Date
Yasin Negash (MSc)	.....	23/03/2021.
Co-advisor	Signature	Date

As the members of the board of examiners of MSc. thesis open defense examination of **Kedir Mokonin Keno**, we certify that we have read and evaluated the thesis and examined the candidate. We recommend that the thesis has been accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

KIBREALEM SISAY	.....	23-03-21
Name of Chairman	Signature	Date
Reta Habtamu	.....	23-03-21
Name of Internal Examiner	Signature	Date
Dr. Habte Tadesse	.....	30/07/2013
Name of External Examiner	Signature	Date

Statement of the Author

As the author of this research study, I declare that the thesis is a result of my genuine work, the support of my supervisors and the helpful hands of other individuals. Thus, all those had who participated in the study and sources of materials used for writing this thesis have been duly acknowledged. I have submitted this thesis to Jimma University as partial fulfillment for the requirements of Degree of Master of Science in Biostatistics.

The library directorate of Jimma University can deposit a copy of the thesis in the university library so that students and researchers can refer to it. Moreover, I declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Kedir Mokonin Keno	……	…….
Name	Signature	Date

Acknowledgements

All praise to Allah (Subhanahu Wata-Alaa), the Most Gracious, Charitable, Beneficent and Billions of peace and blessings be upon Holy Prophet (Nabi Muhammad S.A.W). I bow in reverence to Almighty for giving me enough courage, patience, and success in this venture.

I would like to extend my profound gratitude to my advisor, Geremew Muleta (Ph.D. Scholar), for his unconditional support, guidance, and encouragement during this study. I would like to express my heartfelt thanks to my co-advisor, Yasin Negash (MSc) for his encouragement, supervision, and constructive comments that develop my research abilities and in improving the quality of the Thesis. It has been a great honour and pleasure for me to work with both of you.

I would like to extend my gratitude to Prof. Legesse Kassa Debusho., from UNISA University, South Africa, for being such a great person who always giving ideas and comments from different dimensions I never thought of. You allowed me to be able to work under minimal supervision and at the same time not deviate from the objective of this project. May the good Lord bless you abundantly and increase your territories in the academic field and life.

I would like to extend my sincere appreciation to Dire Dawa University for offering me the opportunity and financial support during my study at Jimma University. Furthermore, I would like to express my deepest thanks and appreciation to the MEASURE DHS team for allowing me to use their data to write my research report.

My immense gratitude goes to my **Mother** and my **Father** for their prayers, encouragement, and for believing in me even during the moments when I thought it was impossible. Finally, I would like to say a big thanks to my sisters and brothers for their support, kind words, and encouragement throughout the entire period.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

WHO	World Health Organization
EDHS	Ethiopian Demographic and Health Survey
MCMC	Markov Chain Monte Carlo
DIC	Deviance Information Criteria
STAR	Structured Additive Regression
GLM	Generalized Linear Models
GAM	Generalized Additive Models
PD	effective number of parameters
MRF	Markov Random Field
UNICEF	United Nations Children's Fund
GMRF	Gaussian Markov random field
IWLS	Iteratively Weighted Least Squares
Hgb	Hemoglobin Concentration
DHS	Demographic Health Survey
HIC	High Income Countries
SSA	Sub-Saharan Africa

Abstract

Background: Anaemia poses significant public health challenges to most developing countries, associated with serious health consequences and affecting about one-fourth of the world's population, mostly under five-year children.

Objectives: This study aimed to Analyze the Spatial Pattern and Determinants of childhood anaemia in Ethiopia using Bayesian Geo-additive approach.

Methods: Our study participants were all the children U5 who were confirmed to anaemia from the 2016 EDHS data source. The survey considered 10,641 children U5; of which 7,953 children with complete anaemia levels were included in this study. The outcome variable was defined as the presence or absence of anaemia based on the WHO cut-off points. In this study Moran's, I was used to investigate the presence of spatial autocorrelation. A geo-additive model which allowed joint analyses of nonlinear effects of some covariates, spatial effects, and other fixed covariates were used. Inference used a fully Bayesian approach via Markov Chain Monte Carlo techniques.

Results: Out of 7,953 children U5 years included in this study 4567 (57.4%) were anemic. Based on DIC model selection criteria Bayesian Geo-additive model was found to be appropriate. From the Model, household wealth index, types of toilet facilities, size of child at birth, education levels of mothers, and mother's anemia status are found to be the significant determinants of childhood anaemia. Child age and mother BMI were found to have a nonlinear relationship with childhood anaemia.

Conclusion: Our finding revealed that there was spatial variation in childhood anaemia across the region of Ethiopia with higher prevalence in the eastern and north-eastern parts of Ethiopia. Bayesian Geo-additive models that capture spatial effects fit the data well. Therefore, the concerned body may use the anemia prevalence map as a basis for interventions and resource allocations.

Keywords: Childhood anaemia; spatial effect; Bayesian Geo-additive models; MCMC

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the Study

Anaemia in childhood is defined by the World Health Organization (WHO) as a decline in the concentration of circulating red blood cells or in the hemoglobin (Hgb) concentration and a related impaired capacity to transport oxygen. It is also defined as a hemoglobin level below 11g/dl for children under 5 years. WHO classified anaemia as mild if the hemoglobin concentration level is between 10-11.9 g/dl, moderate if the level is between 7-9.9 g/dl, and severe if the hemoglobin level is below 7g/dl [1].

Anaemia is a global public health problem affecting both developing and developed countries with major consequences for human health and socio-economic development and affecting about one-fourth of the world's population, mostly under five-year children and women of reproductive ages[2, 3]. The presence of anaemia in children under five can negatively impact their mental development and future social performance. Children suffering from iron deficiency anaemia during their 2 years of life have slower cognitive development, poorer school performance, and poorer work capacity in later years[4]. It is also not only a major cause of pre and postpartum morbidities and mortalities for children in developing countries but it also affects the physical and cognitive development of children, so the health of children in particular and adults, in general, is very essential for their survival and the future development of the country [5].

The etiology of anaemia involves the interaction between multiple factors including nutritional deficiencies, genetic red blood cell disorders, and infectious diseases, particularly malaria and hookworm infections. Iron deficiency was also the most common cause of anaemia in High-Income Countries (HIC). Approximately 50% of anaemia cases are due to iron deficiency, though the proportion varies among population groups and in different areas, according to the prevailing local conditions [1]. In highly malaria-endemic countries, particularly in Sub-Saharan Africa (SSA), malaria is a significant contributing factor to childhood anaemia[6].

Globally, anaemia affects 1.62 billion people, which corresponds to 24.8% of the population. The highest prevalence is in under-five children which are 47.4%. WHO regional estimates generated for preschool-age children and pregnant and non-pregnant women indicate that the highest proportion of individuals affected are in Africa (47.5–67.6%), while the greatest number are in South-East Asia where 315 million individuals are affected and in the United States, the prevalence is approximately 3.6%. In Europe, the prevalence of anemia in countries such as Sweden and Germany is 8.6% and 7.8% respectively[7].

According to the most recent estimates of the WHO, the highest anaemia prevalence was 42.6% in 2011 occurred in children under the age of five years old which translated to just over 273 million children suffering from anaemia globally[8]. According to a study conducted in other parts of the world, the magnitude of anaemia in under-five children was 62% in India [9]. On the other hand, the prevalence of anemia was 40% in Bangladesh[10]. Furthermore, a study conducted in western China, showed that the prevalence of mild, moderate, and severe anemia to be 27.4%, 21.9%, and 3.2%, respectively [11]. Another study conducted in Brazil's children under-five indicated that the overall prevalence of anemia was 51.2 % [12].

The global estimate for anaemia prevalence in the year 2010 was 32.9%, with East sub-Saharan Africa having the highest-burden and the hardest-hit age group was children under age 5[13]. In a recent report in 2011, the WHO [8] estimated that the prevalence of anaemia in children was reaching 62.3% in Sub-Saharan Africa remains the most affected region. A study conducted in some parts of Africa shows that the prevalence of childhood anaemia ranged from 36.4% to 61.9% for Malawi, Uganda, and Tanzania [14]. In Cape Verde, West Africa, the prevalence of anemia was 51.8% [15]. A cross-sectional hospital-based study conducted in Kassala, Eastern Sudan indicated that 86% of all children were anemic[16] and the prevalence of anaemia was high in Uganda which was 58.8% [17].

In Ethiopia, the prevalence of anaemia in under-five children ranges from 34% to 68.5%[18–21]. The national data from Ethiopian Demographic and Health Survey (EDHS) in 2011 showed that the prevalence of anaemia among children under five was 44%, which was around more than four out of ten under-five children were anaemic. From these, about 21% of children were mildly

anaemic, 20% were moderately anaemic, and 3% were severely anaemic[22]. A study conducted by Habte *et al.,* [20] showed that the prevalence of anaemia among children between 6-59 months old was 50.3% with a peak at age of 6-11 months (68.5%). The prevalence of anemia is also reported about 37 % in Northern Ethiopia with a higher magnitude among children less than 6-11 months (53.2%)[23]. These findings suggest that the prevalence of anaemia among children less than five is higher for the younger age children.

According to the 2011 EDHS report, the prevalence of anaemia among children 6- 23 months group is 60.9% [22]. There are also relatively few studies regarding the prevalence of anemia among children 6- 23 months of age in Ethiopia. Woldie *et al.,* [18] reported an anaemia prevalence of 66.6% in northern Ethiopia. Another study conducted in the Sidama region of southern Ethiopia revealed that 24% of children were anaemic at 6 months and increased to 36% at 9 months[19]. Also, a study conducted in Ethiopia among an under-five year of age children identified that the magnitude of anemia was 27.1%,43.7%, and 32.4% in the eastern part of Ethiopia, southwest part of Ethiopia, and Gilgel Gibe Hydroelectric dam of South West Ethiopia [24–26]. Furthermore, a study conducted in Wollo also shows that the overall prevalence of anemia in under-five children was 41.1%[27]. Another study conducted in Debre Berhan Town, North Shewa, Ethiopia shows that the overall prevalence of anemia was 47.5%, of which 18.3% were mildly anemic, 25% moderately anemic, and 4.1% severely anemic[28].

Although substantial progress in control has been achieved over the past decade, anaemia remains one of the world's largest killers of children U5 in 2016. Anaemia prevalence among children U5 old in Ethiopia was 57%, rising unexpectedly from 44% in 2011. Infants and young children bear the highest burden of anaemia in Ethiopia, with a 72% prevalence among those under two years of age[29]. Even though the previous study has been emphasized the importance of understanding the prevalence of anaemia, limited statistical work has been carried out from data arising from demographic and health survey in Ethiopia with evidence confirming that prevalence of anaemia among children U5 in Ethiopia was a severe public health problem, by applying classical models under frequentist settings that assumed that the random components at the contextual level (a region in our case) are mutually independent [30–33]. EDHS data are based on a random sample of the region which, in turn, introduces a structured component. Such a component allows us to

borrow strength from neighbors to cope with the posterior uncertainty of the region effect and obtain estimates for areas that may have inadequate sample sizes. The geo-additive model was used to examine the potential bias incurred when ignoring the dependence between aggregated spatial areas, through simultaneously investigating the geographical variation and the risk factor of child anaemia. Therefore, this study applies a Bayesian Geo-additive model that accounts for possible non-linear effects of some factors on childhood anaemia at the disaggregated regional level that cannot be explained by the classical set of fixed linear socioeconomic and bio-demographic factors while simultaneously controlling for geographical variation.

1.2 Statement of the Problem

Research involving childhood growth and development has become increasingly important as several international agencies, governments and researchers agree that investing in the health of the young child today will lead to an improved society in the future. Children who lack basic health care needs are more susceptible to infections diseases and more prone to delayed mental and cognitive development. These lead to poor school performance when they are enrolled in school, leading to a reduction in intellectual achievement and subsequently reduced work capacity in their adulthood. In effect, economic productivity is decrease[34].

From the current Ethiopian demographic and health survey reports in 2016, anaemia prevalence among children under five years was 57%, rising unexpectedly from 44% in 2011. Infants and young children bear the highest burden of anaemia in Ethiopia with a 72% prevalence of anaemia among those under two years of age[29]. Although the biological immediate causes of anaemia are also documented and the government of Ethiopia applied tremendous efforts to decrease the prevalence of childhood anaemia but still it was a major public health problem according to WHO criteria [35].

Previous studies in Ethiopia have mainly focused on the contribution of individuals and household factors in explaining the childhood anaemia prevalence in the country using Classical models like Binary logistic regression and the Multilevel model[30–33]. Such studies, while neglecting the critical influence of community-level variables and regional variation that can be explored by

considering spatial dependence in the analysis. Neglecting of the spatial pattern (spatial dependence) in our data and unobserved heterogeneity among clusters may lead to considerably biased estimates for the remaining effects as well as false standard estimates. Also considering some nonlinear continuous covariate in linear form into models leads them no more significant even though they have more significant effects on anaemia status. Bayesian Geo-additive model, on the other hand, allows for joint modeling of fixed effects, nonlinear effects of the metrical covariate, and spatial effect simultaneously and provide valid and efficient inference.

Mapping of areas with high under five anaemia under investigation is crucial in economically constrained country like Ethiopia. The map enables efficient allocation of scarce resource. To achieve this one the structured additive regression model was used. The novelty of this study was the use of recently developed structured additive regression model which provides the valid and realistic statistical inference. Researcher motivated to focus on anaemia since it is the world's second-largest cause of disability and death after malaria for children U5 years and women of reproductive ages [36] and the lack of statistical investigations applied on childhood anaemia data by adopting the Bayesian Geo-additive model in looking at the effect of some covariates and geographical variation on the health of children with anaemia.

Therefore, this study focused on modeling the determinants of anaemia among under-five children in Ethiopia via the Bayesian Geo-additive model which captures spatial heterogeneity for unobserved influential factors and also accounts for nonlinear effects of a metrical covariate.

The study has attempted to answer the following basic research questions: -

- What is the pattern of anaemia among under-five years of age children in Ethiopia?
- Which factors are significantly associated with anaemia among U5 children in Ethiopia?
- What is the spatial distribution of anaemia in children under five years in Ethiopia?
- Which statistical model allows for joint modeling of fixed effects, nonlinear effects of the metrical covariate, and spatial effect?

1.3 Objectives of the study

1.3.1 General Objective

The General Objectives of this study is to Analyze the Spatial Pattern and Determinants of childhood anaemia in Ethiopia using Bayesian Geo-additive approach.

1.3.2 Specific objectives

- To identify the pattern of anaemia among under-five years of age children in Ethiopia.
- To determine factors significantly associated with anaemia among U5 children in Ethiopia.
- To determine the spatial distribution of anaemia among U5 year of age children in Ethiopia.
- To select the best flexible statistical model that allows for joint modeling of fixed effects, nonlinear effects of the metrical covariate, and spatial effect.

1.4 Significance of the Study

Knowledge of disease burden in populations especially in U5 children in Ethiopia was essential for health authorities who seek to use limited resources to the best effect by identifying priority health programs for prevention and care. Therefore, studying the spatial distribution and factors associated with anaemia in U5 year children would provide more insight and practical guidelines to the formulation of policies aimed at fighting the spread of childhood anaemia.

Mapping the spatial distribution of anaemia prevalence by the regions of the country was the main contribution of this study. Thus policymakers, program planning committee members, and other non-governmental organizations are well informed of the regions with the greatest prevalence of this disease and provide more insight and practical guidelines to the formulation of policies aimed at fighting the spread of this disease. This study was also providing the specific factors that lead to this prevalence of anaemia and help to reduce the prevalence of childhood anaemia by giving awareness for society on the factors that increase the risk of this disease. Lastly, this study would also use as a bridge for further studies for the health sector and another researcher.

CHAPTER TWO

2 Literature Review

This section provides a review done by other authors about anaemia. The review discusses different statistical approaches previously used in assessing the relationship between childhood anaemia and influential factors and also determinant factors.

2.1 Burden of anaemia among under-five children in the Africa and Globe

Anaemia is often associated with childhood malnutrition, impairment of red cell production, and increased red cell destruction, which could increase mortality risk [36]. In 2012, 6.6 million children under age 5 died. Most of these deaths occurred in low-income countries and specifically in the African region where children are 16 times more likely to die than children in the developed countries. More than half of these deaths are due to preventable infections and lack of access to simple affordable interventions[37].

Globally the prevalence of anaemia in children under age 5 is 47.4%, that is about 293 million children are anaemic according to WHO of 2008 and 2009 reports. The highest overall prevalence is in Africa, 67.6%, and South-East Asia 65.5%. In the Eastern Mediterranean, the prevalence is 46% and around 20% in the other WHO regions, the Americas, Europe, and Western Pacific[7].

The prevalence of anaemia among children under age 5 in countries such as Monaco and Australia were as low as 5.0% and 1.1% respectively, while the prevalence of anaemia in most African countries was high ranging from 74% in Tanzania to 43% in the Democratic Republic of Congo. In almost all countries in the Sub-Saharan African region, the prevalence of anaemia in children under 5 is above the severe prevalence threshold of 40%. The highest overall prevalence of anaemia in children under 5 age is recorded in the Western and Central African Region, around 75%[7].

In Ethiopia, according to the EDHS reports in 2011, more than four out of ten under-five children (44%) were anemic [22]. A different study showed that the prevalence of anaemia among children

6-59 months old was 50.3% in Ethiopia[20] and 37 % in Northern Ethiopia with a higher magnitude among children less than 6-11 months (53.2%)[23]. While in 2016, anaemia prevalence among children under five years old in Ethiopia was 57%, rising unexpectedly from 44% in 2011. Infants and young children bear the highest burden of anaemia in Ethiopia, with a 72% prevalence of anaemia among those under two years of age[29]. WHO classifies anaemia prevalence above 40% as a severe public health[35].

2.2 Determinants of anaemia in children under five

A large number of variables have been associated with childhood anaemia status as discussed in many kinds of literature. The variables are classified under child factors, maternal factors, and socio-demographic factors. The conceptual framework depicting the classification of these factors was shown below in Figure 3.1.

Age of Child: The current age of the child has been identified to be significantly associated with anaemic status of the child.

The study conducted by Roberts *et al*., [14] on investigating the demographic and socio-economic determinants as well as the spatial variation of anaemia in children aged 6 to 59 months in Kenya, Malawi, Tanzania, and Uganda using Bayesian hierarchical model, their study result showed that prevalence of childhood anaemia ranged from 36.4% to 61.9% across the four countries. The results of this study also revealed that the non-linear effect of a child's age in months has significant effects and there was an increase in effect from 6 to 10 months, after which the effect declined.

Sex of Child: According to the available literature on childhood anaemia, there exists some association between the sex of the child and his/her anaemic status and the prevalence was in males than females.

According to Leite *et al*., [12], the results of the National Survey of Brazil point to an association between sex and anaemia in indigenous children, male have been observed to be at a higher risk of anaemic than females. Similar research conducted in Tanzania on the socio-demographic

determinants of anaemia among children aged 6-59 months in mainland using Alternating Logistic Regression, Males aged 6 to 59 months had 1.26 times higher odds of anaemia than their female counterparts[38].

Birth Weight: The weight of a child at birth is significantly associated with an increased risk of infections, diseases as well as mortality. Children with low birth weight have increased exposure to infections and diseases and are also more likely to die before their 1st birthday[39].

A study conducted in Ghana by Ewusie *et al.,* [4], on the prevalence of anaemia among under 5 children using Ghana demographic and health survey identified that the prevalence was lower in children with larger birth weight than those with small birth weight. The birth weight used in this study was the categorical subgroups which are originally coded as "size of child at birth" in the 2016 EDHS. For ease of interpretation, the birth weight was recoded into 3 categories: Larger than Average, Average and less than average.

Child's Nutritional Intake: Poor nutrition in children leads to anaemia. Nutritional anaemia occurs due to insufficient intake of nutrients by cells. Among them, Iron deficiency anemia is the most common cause of anaemia and blood diseases in developing countries. It is estimated that 75% of anaemia is related to iron deficiency, followed by folate and vitamin B12 deficiencies[39].

A study conducted on the prevalence of anaemia in Ethiopia using multilevel logistic regression revealed that iron deficiency is one of the main factors for the prevalence of anaemia and the odds of the prevalence of anaemia among under-five children had statistically decreased by 26.76% for children who had been given iron pills/syrup[40]. Another study conducted in Cameroon using logistic regression revealed that children who did not give iron pills/syrup were 1.3 times more likely to be anaemic as compared to children with children who had been given iron pills/syru[41].

Had fever in previous weeks: According to Konstantyner *et al*.,[42] fever is a common symptom of acute and chronic inflammatory diseases, most infections, which have been associated with lower hemoglobin levels.

The study conducted by Ngwira and Kazembe [43] in Malawi on investigating the risk factors affecting the severity of childhood anaemia using multinomial cumulative logistic regression in the Bayesian approach, revealed that fever is associated with higher levels of childhood anemia, severe anaemia compared to having no fever. The risk of anaemia was 0.423 times more on a child who had fever two weeks before the study compared to having no fever.

There was a similar study conducted in Lesotho, South Africa by Gaston *et al*.,[44], on determinants of factors associated with anaemia among children under five years, revealed that child who has a fever in the last two weeks before study was 1.674 times more likely to be anaemic than the child who did not have a fever in the last two weeks before the survey.

Another study was conducted in Rwanda on assessing the determinants of childhood anaemia using structured additive quantile regression, showed that fever has a significant negative effect on childhood hemoglobin concentration in all quantiles of interest and therefore a child having a fever increases the likelihood of childhood anaemia[45].

Diarrhea before two weeks of the study: The presence of infections can result in loss of appetite and malabsorption of nutrients which increases the risk for anaemia and also increases metabolic rate after infection.

A study conducted in Ethiopia on factors associated with anaemia among children aged 6–23 months show that from anaemic child, 14.7% of them had diarrhea in the last two weeks. They used the Binary logistic regression analysis model and their results revealed that children with a history of diarrhea before two weeks of the study were 4.9 times more likely to be anemic than children without diarrhea[18].

Mother Educational Level: One variable that has been significant in almost all studies of the anaemia status of children was the educational level of the mother. Educated mothers are better informed on the knowledge and use of health facilities, appropriate infant and young child feeding practices, and better sanitation practice[46].

A study conducted in Kenya by Oscar Ngesa and Henry Mwambi [47] on the title "Prevalence and Risk Factors of Anaemia among Children Aged between 6 Months and 14 Years in Kenya". Their study revealed that children whose mothers had secondary, and higher levels of education, were less likely to be anaemia positive. The risk of anaemia was 1.5 times more in children whose mothers had no education as compared to children whose mothers had post-secondary education.

A cross-sectional study conducted in Northeast Ethiopia on a title "Factors Associated with Anemia among Children Aged 6-23 Months Attending Growth Monitoring at Tsitsika Health Center, Wag-Himra Zone" using Binary logistic regression analysis, revealed that mother level of education may positively influence practices related to the health care and feeding practice of their children. The risk was children of mothers with no formal education were 2.6 times more likely to be anemic than children of a mother with secondary and above education levels[18].

Mother Occupation: Maternal occupation was another significant risk factor of anaemia in children in Ethiopia.

A study conducted by Kindie [30]in Ethiopia to investigate the determinants of severity levels of anaemia among children aged 6–59 months using ordinal logistic regression analysis, based on cross-sectional data from EDHS 2011, revealed that the odds of being anaemic status was higher for children from non-employed mothers. The risk is about 1.13 times higher in children from non-employed mothers than children from employed mothers.

A similar study was conducted in Ethiopia on title "Spatial Distribution and factors associated with childhood anemia in Ethiopia".based cross-sectional study design using mixed effect logistic regression model, revealed that the likelihood of developing anaemia among children who had employed mothers was decreased by 13% as compared with children whose mothers were not employed currently[48].

Mother Body Mass index: The impact of the nutritional situation of the mothers, measured using the Body Mass Index (BMI**,** defined as the weight in kg divided by the square of height in meters) on the child's anaemia status is presumed to follow nonlinear effects. Mothers who exhibit a very low BMI**,** indicating their poor nourishment, are likely to have poorly nourished children. At the

same time, parents with a very high BMI might also have poorly nourished children as the obesity associated with their high BMI indicates poor quality of nutrition and might therefore indicate poor quality of nutrition for their children.

A study conducted by Roba *et al*.,[49] on the Prevalence of stunting and anaemia among children 6-23 months of age in two agro-ecological zones of Rural Ethiopia using Binary logistic regression, revealed that as the haemoglobin level of the mother increased, the risk of the child being anaemic decreased by 29%. Similarly, as mothers' BMI increased the likelihood of the children being anaemic decreased. This study provides strong evidence that there is a relationship between child anaemia and maternal nutritional status (measured by her Body Mass Index).

Other studies on Prevalence of anemia and associated factors among indigenous children in Brazil: using the First National Survey of Indigenous People's Health and Nutrition, using hierarchical multivariate analysis, revealed that the odds of being severely/moderately anaemic were higher for children whose mother's BMI was underweight compared to children whose mother's BMI was normal[12].

Mother Anaemia level: Maternal anaemia level is one of the determinant factors of child anaemia.

Erkihun and Tiruneh [48] used a Mixed effect logistic regression model to analyze the spatial distribution and factors associated with childhood anemia in Ethiopia, The findings show that there is a strong association between child anaemia status and mother anaemia status which reveals that children who live with anaemic mothers were 53% more likely to develop anaemia as compared with those children who lives with non anaemic mother.

A study conducted in Uganda on Socio-Demographic Determinants of Anaemia in Children using multilevel analysis, identified that children of anaemic women are associated with a higher prevalence than those of no anaemic women, the risk was children of anaemic mothers are associated with two folds increased risk of anaemia compared with those whose mothers are not anaemic[50].

Another study undertaken in Ethiopia by Gebrehaweria and Lemma[21], on Factors associated with anaemia among children 6–23 months of age in Ethiopia using a Multilevel Analysis showed that the odds of anaemia were higher for children from anaemic mothers than non-anaemic mothers. The risk were children of mothers who had anaemia had 1.53 times greater odds of being at higher levels of anaemia compared to the children of mothers who had no anaemia.

Household wealth index: Some of the socio-economic factors that have been studied as indicators of the socio-economic status of a household was the household income/wealth index.

Gebremedhin *et al*.,[23] studied the determinants of anaemia among children aged 6–59 Months Living in Kilte Awulaelo Woreda, Northern Ethiopia using Multivariate logistic regression analyses based on a cross-sectional study, showed that the higher prevalence of anaemia among the children living in the household with lower monthly income compared to those with higher income.

Stephen [51] conducted cross-sectional study on title "Spatially Adjusted Determinants of Malaria and Anaemia Morbidity among Children Under age 5 years in Ghana" using GDHS data, noted that lower wealth index status was allied with the increase in the danger of development of anaemia in children.

Source of drinking water: Access to clean water is found to be negatively correlated to all forms of nutrition deficiency[46, 52]. Access to clean drinking water decreases the risk of water-borne diseases such as cholera, diarrhea, and typhoid.

Rashid *et al*.,[53] conducted a cross-sectional study on Predictors and prevalence of anemia, among children aged 6 to 59 months in Shabelle zone, Somali region, eastern Ethiopia, revealed that the source of drinking water has significant effects on childhood anaemia, drinking of unprotected water were almost five times (AOR = 4.9, 95% CI= 2.204, 10.820) more likely to develop anemia compared to drinking protected water.

A study conducted in Bangladesh [54] aiming to analyze determinants of severity levels of childhood anaemia using a proportional odds model, based 2011 Bangladesh Demographic and

Health Survey (BDHS), show that the odds of being at higher anemia status were higher for children whose households used non-improved source of drinking water (OR = 1.328; 95% CL: 1.002–1.758).

Toilet facilities: The type of toilet used by a household is an indicator of household wealth and a determinant of environmental sanitation. This means that poor households, which are mostly located in rural areas for both countries, are less likely to have sanitary toilet facilities.

A study conducted by Roberts *et al*., [14] on investigating the demographic and socio-economic determinants as well as the spatial variation of anaemia in children aged 6 to 59 months in Kenya, Malawi, Tanzania, and Uganda using multivariable hierarchical Bayesian geo-additive model, showed that a significantly lower odds of anaemia was suggested for children living in households with improved toilet facilities (PIT latrine and flush toilet).

Several studies revealed that there were spatial variations of anaemia disease in under-five children in various countries around the world. Research conducted by Gayawan *et al*., [55] on Possible determinants and spatial patterns of anaemia among young children in Nigeria, exploring geographical variations of anaemic status avails policy-makers with tools to prioritize and roll out interventions in a more prudent manner. Findings from this study show a significant geographical variation in the anaemic status of children in Nigeria. Children from Jigawa, Kebbi, Kwara, and Yobe states were significantly associated with higher chances of being anaemic.

A study conducted in Nigeria titled "Spatial heterogeneity and determinants of childhood anaemia carried out by Chigozzie and Temesgen [56], presented solid evidence of geographical heterogeneity for childhood anaemia prevalence in Nigeria. This study suggested that both socioeconomic and environmental covariates may be essential risk factors for anaemia prevalence across geographical variation.

A study conducted by Erkihun and Tiruneh [48] on Spatial Distribution and factors associated with childhood anaemia in Ethiopia based cross-sectional study design using a generalized linear mixed-effect logistic regression model, revealed that there was spatial clustering of childhood anaemia in Ethiopia.

Research conducted by Stephen [51], further disclosed that the clusters of high moderate and severe anaemia prevalence were located in a spatial difference in all ten regions of the country in children under age 5 years in Ghana. It shows that anaemia "hotspots" are clustered in mostly rural districts of very high poverty and low level of educated mothers.

2.4 Spatial pattern of anaemia among under-five children

Spatial epidemiology was founded on the premise that individuals who lived in close proximities were generally exposed to similar factors which were likely to affect detected outcomes[65]. Bayesian geostatistical methods are increasingly utilized in spatial analysis, disease mapping, and consequently, decision-making. Their flexibility enables them to integrate spatial correlation and modeling of fixed and random variables[66]. In so doing, using Bayesian geostatistical analysis provides researchers with a tool for the identification of high prevalent areas where great variations exist in disease epidemics. This provides correct estimates of parameters tested, predicts risk at non sampled locations and estimates heterogeneity in areal data, which was used to identify high risk determinants behind the spread of a disease may assist in guiding health and policy planners in developing and allocating resources for anaemia prevention programs among under five children[67].

Spatial variations were observed in childhood anaemia at the African level [68]. The variation is more pronounced based on the geographical and socio-economic status of the regions. Study on Possible determinants and spatial patterns of anaemia among young children in Nigeria using Bayesian geo-additive modeling and come up with the results among Nigerian states, namely, Jigawa, Kano, Kwara, and Yobe, had a significantly higher likelihood of being anaemic, while Benue, Delta, Kogi, and Rivers states were significantly associated with non-anaemia (normal) [55].

According to Roberts *et al*.,[14] study entitled "Investigating the spatial variation and risk factors of childhood anaemia in four sub-Saharan African countries," the study revealed distinct spatial variation in childhood anaemia within and between Malawi, Uganda, and Tanzania. The spatial variation appeared predominantly due to unmeasured district-specific factors that do not transcend.

According to a study conducted by Alfred and Lawrence [43] on “Bayesian random-effects modeling with application to childhood anaemia in Malawi,” they found that the observed residual spatial pattern in childhood anaemia shows most districts in the north reducing child anaemia, and the districts that increased risk of anaemia were all close to water bodies. The observed spatial heterogeneity may be due to unobserved factors not captured by the covariates in the models, and it is a matter of conjecture to identify them. Therefore, this research focuses on a geo-additive model that allows the mapping of spatial effects to childhood anaemia in Ethiopia's case, while accounting for non-linear covariate effects under the assumption of addictiveness. Modeling of metrical continuous covariates non linearly revealed their subtle influences that could not be observed when modeled linearly and the incorporation of spatial effect in the linear models made some covariates not to be significant anymore.

CHAPTER THREE

3 DATA AND METHODS

3.1 Description of the study area

This study was conducted in Ethiopia. Ethiopia is strategically located in the northeastern part of Africa popularly known as “the Horn of Africa". It shares a boundary with the North and South Sudan on the west, Somalia and Djibouti on the East, Eritrea on the North and northwest, and Kenya on the South. Ethiopia is officially known as the Federal Democratic Republic of Ethiopia, is a landlocked country located in the Horn of Africa. It is the second-most populous nation in Africa, with over 105,350,020 populations [29] and the tenth-largest by area, occupying 1,100,000k^2m. Ethiopia has eleven geographic or administrative regions: nine regional states (Tigray, Afar, Amhara, Oromia, Somali, Benishangul-Gumuz, SNNPR, Gambela, and Harari) and two city administrations (Addis Ababa and Dire Dawa that are considered as a region) with a capital city of Addis Ababa.

3.2 Source of Data

The standard EDHS of 2016 dataset was used in this study, which is designed to encompass the national populace and includes a wide range of households. The primary objective of the EDHS is to provide up-to-date estimates of key demographic and health indicators. Here, a sample of children under five years of age with anaemia confirmed test results from the EDHS in 2016 would be included in this study. More specifically, the EDHS Conducted haemoglobin testing on eligible children age 6-59 months to provide information on the prevalence of anaemia in these groups.

Children aged 6 months to 59 months who stayed in the household and eligible to be measured for hemoglobin during the survey comprise the study population, whereas the target population of this study was under-five aged children in Ethiopia. The lower limit of 6 months was chosen since the cut-off for anaemia is not defined in children below 0.5 years of age but by 6 months hemoglobin has risen to normal values.

3.3 Study Variables

Response Variable

The outcome variable of this study was the anaemia status among under five years children in Ethiopia, classified based on the WHO definition of anaemia in children under five years[7].

$$y_i = \begin{cases} 1 & if\ haemoglobin\ level\ is < 11\ g/dl \quad (Anaemic) \\ 0 & if\ haemoglobin\ level > 11\ g/dl \quad (Non\ anaemic) \end{cases}$$

Explanatory Variables

From the literature support, the explanatory variables considered in this study were several demographics, socio-economic, and environmental factors (Fig.3.1).

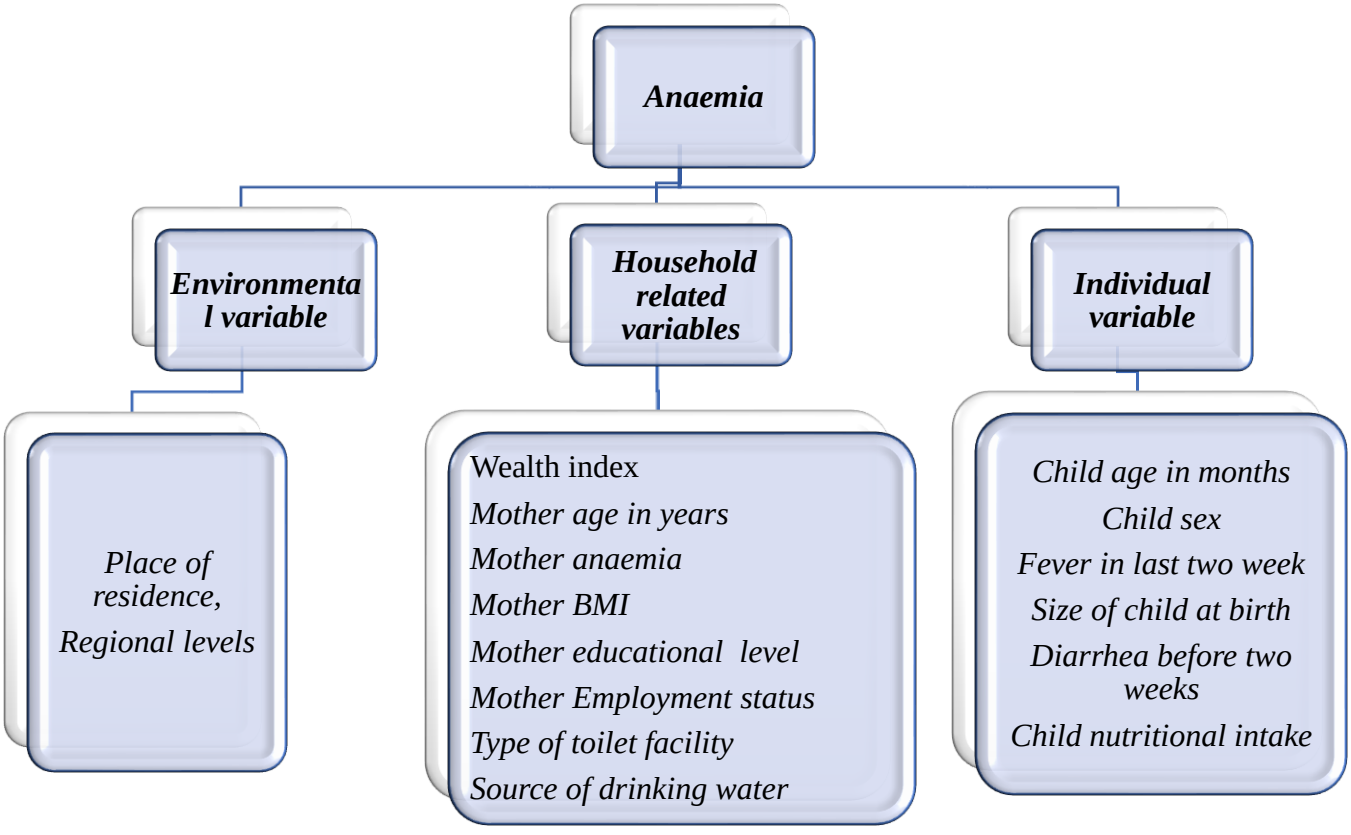


Figure 3.1: Conceptual framework for anaemia and its related factors

3.4 Methods of Data Analysis

3.4.1 Advantages of Bayesian Methods

From the complex nature of hierarchical models to the knowledge required to simulate from the posterior distribution, Bayesian methods can appear daunting. However, Bayesian methods also offer many advantages over classical inference. This is especially true for spatial models.

First, spatial data are by nature autocorrelated. To account for autocorrelation, it is necessary to use some sort of hierarchical structure. Bayesian models are inherently hierarchical and thus accounting for autocorrelation within the Bayesian framework is straightforward. This includes the use of the CAR models as MRF prior distributions for Bayesian inference. The complicated structure of hierarchical models often results in the posterior being intractable, but MCMC sampling techniques overcome this problem[69].

Second, Bayesian hierarchical models are very flexible, accommodating for the uncertainty in estimated random and fixed effects, a priori knowledge through the specification of priors and hyper-priors [70]. Despite accounting for heterogeneity in spatial data, non-Bayesian methods can lead to underestimation of uncertainty in the model parameters.

Last, Bayesian analysis provides better estimates in several ways. Bayesian estimation provides a distribution rather than a more informative point estimate. Bayesian estimation avoids ‘over-fitting’ of the model by integrating over the model parameters and estimating parameters in mixture models, especially when the number of mixture components is unknown, is more convenient and accurate in a Bayesian framework [71]. The specification of a prior distribution allows for the formal inclusion of information from previous studies or expert opinion. The posterior distribution, in turn, allows for easy estimation of a (posterior) probability that is more intuitive and interpretable than the frequentist p-value.

3.4.2 Advantages of Bayesian Geo-Additive Model

There are many potential advantages of this approach over classical approaches like regression models with fixed or random provinces effects; or standard 2-level multilevel modeling with unstructured spatial effects[72]. In the classical models, it is assumed that the random components at the contextual level (a region in our case) are mutually independent. In practice, these approaches specify correlated random effects[73], which are contrary to that assumption.

Further, Borgoni and Billari [74] point out that the independence assumption has an inherent problem of inconsistency. They argue that if the location of the event matters, it makes sense to assume that areas close to each other are more similar than areas that are far apart.

Also, treating groups (in our case region) as independent is unrealistic and leads to poor estimates of the standard errors. As Rabe-Heskesth and Everitt [75] stipulate, Standard errors for between-region factors are likely to be underestimated because we are treating observations from the same region as independent and thus increasing the apparent sample size. On the contrary, standard errors for within-province factors are likely to be overestimated [76].

On the other hand, Demographic and Health Survey data are based on a random sample of the region which, in turn, introduces a structured component. Such a component allows us to borrow strength from neighbors to cope with the posterior uncertainty of the region effect and obtain estimates for areas that may have inadequate sample sizes or are not represented in the sample.

In an attempt to highlight the advantages of our approach in a spatial context and examine the potential bias incurred when ignoring the dependence between aggregated spatial areas, we fitted three models with and without the structured and random components in this study. Controlling for important risk factors such as geographical location (spatial autocorrelation) arising from environment impact on health gives statistically robust estimates of prevalence.

3.4.3 Overview of Geo-additive Model for other areas of Application

Structured additive regression (STAR) models provide a flexible framework for modeling possible nonlinear effects of covariates and spatial effects: They contain the well-established frameworks of generalized linear models (GLM) and generalized additive models (GAM) as special cases but also allow a wider class of effects, e.g., for geographical or Spatio-temporal data, allowing for the specification of complex and realistic models [57].

Geo-additive model, which combines the idea of geo-statistics and additive models uses mostly recorded observations in which there is provision for the assessment of geographical information of the location as well as nonlinear effects of metrical covariates, have been shown, over the years and by various researchers, to be very useful in some other areas.

Kamman and Wand [59] have shown that linear mixed models could be used for geo-additive model fitting and inference. Extension of geo-additive models in the direction of generalized responses are contained in Fahrmeir and Echavarria [60] deal with exponential family models like for count response. They used a Bayesian mixed model framework, with fitting via MCMC and provide applications.

The extension of Geo-additive models to survival data has seen considerably researched since 2003. Hennerfeind *et al.*, [61] developed geo-additive survival models for both geographical point data and count data. They take a Bayesian P-spline approach and use Gaussian and Markov random fields for the spatial components.

Geo-additive models with missing data covariate are studied by French, Wand, and Ibrahim [62] extended that work to Geo-additive models that allow for specification of the covariate distribution and the missing data mechanism.

3.4.4 The Concept of Spatial Dependence

The main concern of spatial statistics is to account for observation correlational effects arising from the geographic configuration of data[77]. The geographical configurations of anaemia

prevalence were assessed to investigate the presence of spatial autocorrelation in the distribution of the data. The essence of spatial analysis is that "space matters", i.e., what happens in one region is related to what happens in neighboring regions. This has been made more precise in what Tobler [78] refers to as the First Law of Geography: "Everything is related to everything else, but closer things more so". One way to approach this is via the notion of spatial autocorrelation[79]. Therefore, the spatial distribution of childhood anaemia in Ethiopia was determined by the geographic relationships among them.

Griffith and Layne [80] also assert that observations are correlated strictly due to their relative locational positions resulting in a spillover of information from one location to another. Hence, spatial autocorrelation is defined as the relationship among a single quantitative variable that results from the geographical patterning of the areas in which the values occur. It is a measure of similarity of objects within an area, the degree to which a spatial phenomenon is related to itself in space.

According to Anselin and Bera[79], spatial autocorrelation can be loosely defined as the coincidence of value similarity with location similarity. In other words, high or low values for a random variable tend to cluster in space (positive spatial autocorrelation) or locations tend to be surrounded by neighbors with very dissimilar values (negative spatial autocorrelation). If it is positive, the anaemia prevalence at a given site tends to be similar to the prevalence of a nearby site. Conversely, negative autocorrelation among the site prevalence indicates that dissimilar prevalence is in nearby or adjacent locations. This study investigated whether there is this systematic spatial variation in the distribution of anaemia prevalence in under-five children.

3.4.5 Global Measures of Spatial Autocorrelation

Tests for global spatial autocorrelation examine whether the data as a whole exhibit spatial autocorrelation (against Ho: no spatial autocorrelation) as well as the strength and direction (positive or negative) of any spatial autocorrelation [81]. The Moran's I [82] and Geary's c [83] are the most commonly used global measures of spatial autocorrelation. They indicate the nature and extent of spatial autocorrelation present in the anaemia prevalence data.

Defining Spatial Weights Matrix

To assess the nature and degree of spatial autocorrelation, it is necessary to represent the spatial arrangement of observations to get a sense of how close or distant they are apart from each other. To express the degree of proximity between observations in space we may attribute a value of one if the observations are nearby (neighbors) and zero otherwise. There are different other options for defining these weights, they may be based on neighborhood which has common boundary [84]. In these cases, pairs of observations might be defined as neighborhoods as a measure of the degree of proximity. In this research, we were adopting the binary adjacency weights such that $\omega_{ij} = 1$ if sites i and j are neighbor's and zero otherwise.

Tests of Spatial Autocorrelation

The two most commonly used measures for spatial autocorrelation are Moran's *I* and Geary's *C* statistics. These tests indicate the degree of spatial association as reflected in the data set as a whole [81]. Test for spatial autocorrelation is designed to quantify the extent of clustering and to allow for statistical inference. The null hypothesis (under the normality and independence assumptions) is given by: H_0: No spatial autocorrelation (H_0 **:** $\boldsymbol{\rho}$ =0). Under the alternative hypothesis (H_1**:** $\boldsymbol{\rho} \neq 0$**)** of spatial autocorrelation (spatial dependence), the interest focuses on instances where large values are systematically surrounded by other large values, or where small values are surrounded by small values.

Global Moran's *I*

For binary weights, Moran [82] introduced the following coefficient of autocorrelation:

$$I = \frac{N \sum_i \sum_j \omega_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i \sum_j \omega_{ij} \sum_i (y_i - \bar{y})^2} \qquad (3.1)$$

Where N is the number of spatial units indexed by i and j; y is the variable of interest; $\bar{y}$ is the mean of y; and ω_{ij} is an element of a matrix of spatial weights; y_i and y_j are the values of the dependent variable at locations i and j respectively; The observed value of *I* can be compared to

its distribution under the null hypothesis of no spatial autocorrelation or no clustering i.e., when the values of y_i are independent of the values $y_j (i \neq j)$ at neighboring locations. Under the normal and randomization assumptions, the resulting z-values can be compared to a table of standard normal to assess significance. The null hypothesis (no spatial autocorrelation), will be rejected if the calculated value of $|Z| > z_{\alpha/2}$ and the z-statistic is given by:

$$Z(I) = \frac{I - E(I)}{S_{E(I)}}$$

Where

$$E(I)_N = \frac{-1}{N-1} = E(I)_R \qquad (3.2)$$

The variance of Moran's *I* and Geary's *C* will vary under the assumption's normality and randomization. Under the normality assumption, the variance of Moran's *I* ($Var(I)_N$) is given as

$Var(I)_N = E(I^2) - \left(E(I^2)\right)^2 = \frac{N^2(N-1)S_1 - N(N-1)S_2 - 2{S_0}^2}{(N+1)(N-1){S_0}^2}$, whereas under randomization $Var(I)_R$ is given by

$$Var(I)_R = \frac{N\left(S_1(N^2-3N+3) - NS_2 + 3{S_0}^2\right)}{(N-1)(N-2)(N-3){S_0}^2} - \frac{K\left(S_1(N^2-N) - 2NS_2 + 6{S_0}^2\right)}{(N-1)(N-2)(N-3){S_0}^2} - \left(\frac{1}{N-1}\right)^2$$

Where $S_0 = \sum_{i \neq j}^{n} \omega_{ij}, S_1 = \frac{1}{2}\sum_i \sum_j \left(\omega_{ij} + \omega_{ji}\right)^2, S_2 = \sum_i \left(\sum_j \omega_{ij} + \omega_{ji}\right)^2, K = \frac{N\sum_{i=1}^{N}(y_i - \bar{y})^4}{\sum_{i=1}^{N}((y_i - \bar{y})^2)^2}$

Interpretation: a positive global Moran's that differs significantly from the expected value under the null hypothesis indicates positive spatial autocorrelation and implying the clustering of similar values (i.e., high values are found closer together, and low values are found closer together) on the dependent variable among neighboring observations.

A negative global Moran's that differs significantly from the expected value under the null hypothesis indicates negative spatial autocorrelation and implies the clustering of dissimilar values

(means high values are found far away from other high values, and low values are found far away from other low values) on the dependent variable among neighboring observations [81].

Global Geary's *C*

The global Moran's *I* define value (dis)similarity as deviations from the mean, whereas the global Geary's *C* defines value (dis)similarity as the squared difference in values between neighboring observations. For binary weights, Geary[83] introduced the following coefficient.

$$c = \frac{(N-1)\sum_i \sum_j w_{ij}\,(y_i - y_j)^2}{2\sum_i \sum_j w_{ij} \sum_i (y_i - \bar{y})^2} \tag{3.3}$$

The z-statistics of Geary's C is given by

$$Z(C) = \frac{C - E(C)}{S_{E(C)}}$$

Where

$$E(C)_N = 1 = E(C)_R \tag{3.4}$$

A variance of Geary's C under normality assumption $(Var(C)_N)$ is given as

$(Var(C)_N) = \frac{((2S_1+S_2)N(N-1)-4{S_0}^2)}{2(N+1)S_0}$, and under randomization is given by

$$(Var(C)_R) = \frac{S_1(N-1)\left(N^2-3N+3-K(N-1)\right)}{S_0 N(N-2)(N-3)} + \frac{\left(N^2-3-K(N-1)^2\right)}{N(N-2)(N-3)} - \frac{(N-1)S_2\left(N^2+3N-6-K\left(N^2-N+2\right)\right)}{4N(N-2)(N-3){S_0}^2}$$

Where all the notations are as in (3.2). Therefore, the null hypothesis of no spatial autocorrelation $(H_0 = \rho = 0)$ will be rejected if the calculated values of $|Z(C)| > z_{\alpha/2}$

Interpretation: A value of Geary's *C* that is significantly larger than one indicates negative spatial autocorrelation, while a value that is significantly smaller than one indicates positive spatial autocorrelation [81]. Due to the squared term in the numerator in (3.3), Geary's *C* gives greater

weight to extreme values than Moran's *I*. As a consequence, the global Moran's *I* is generally preferred in practice[85].

Moran Scatter Plot

The Moran Scatter plot enables us to visualize the linear correlation and the Moran's *I* coefficient will be the slope of the regression curve. In additions to this, inspection of global and local spatial instability is carried out by the means of the Moran scatter plot. The four different quadrants of the scatter plot correspond to the four types of local spatial association between a region and its neighbors: the first quadrant, (HH) a region with a high value surrounded by regions with high values (top on the right), the second, (LH) a region a with low value surrounded by regions with high values (top on the left),the third (LL) a region with a low value surrounded by regions with low values (bottom on the left) and the last (HL) a region with a high value surrounded by regions with low values (bottom on the right) as shown in the following Figure 4.3 below. The first and the third quadrants refer to positive spatial autocorrelation indicating spatial clustering of similar values whereas the second and the forth quadrants represent negative spatial autocorrelation indicating spatial clustering of dissimilar values[81].

Generalized linear models

A common way to build regression models extending the classical linear model for Gaussian responses to more general situations such as binary responses is generalized linear models, originally introduced by Nelder and Wedderburn [86]. They are used for modeling non-Gaussian response variables. In these models the influence of covariates on a response variable y is assumed to satisfy the following two assumptions:

Distributional assumption: - Conditional on covariates x, the responses *y* is independent and the distribution of y_i belongs to an exponential family, i.e., its density can be written as

$$p(y_i|x_i) = exp\left\{\frac{[y_i\theta_i - b(\theta_i)]}{\phi}\omega_i + c(y_i, \theta_i, \omega_i)\right\}, \qquad i=1,\ldots,n \tag{3.5}$$

under the univariate response properties. Where,θ_i is the natural parameter of the exponential family,ϕ is the dispersion parameter common to all observations and ω_i represent a weight for the observations. Furthermore, *b*(.) and *c*(.) are functions depending on the specific exponential family.

Structural assumption: -The (conditional) expectation $E(y|x) = \mu$ is linked to the strictly linear predictor

$$\eta_i = x_i'\beta. \quad \text{Via} \quad \mu_i = h(\eta_i) \; or \; \eta_i = g(\mu_i), \tag{3.6}$$

where the design vector x_i usually includes the *grand mean*, h *is* a smooth, bijective response function, $g(.)$ *is* the inverse of h called the link function and β is a vector of unknown regression coefficients. Both assumptions are connected by the fact that the mean of *y* is also determined by the distributional assumption and can be shown to be given as

$$\mu_i = E(y_i|x_i) = b'(\theta_i)$$

Also, $var(y_i|x_i)$ is the variance of y_i in general which it dependent on the linear predictor with $\frac{\phi v(\mu_i)}{\omega_i} = b''(\theta_i)$ being the variance function of the underlying exponential family.

$$\sigma^2(\mu_i) = var(y_i|x_i) = b''(\theta_i)/\omega_i$$

For binary responses $y\epsilon(0,1)$, the expectation is given by the probability $\pi = p(y = 1)$, which requires appropriate response functions to ensure $\pi\epsilon[0,1]$.Obviously, any cumulative distribution function satisfies this condition and different model formulations are obtained for different choices of the distribution function. In any case, the scale parameter is again fixed at $\phi = 1$. When choosing the natural link function

$$g(\pi) = \log\frac{\pi}{1-\pi} = \eta,$$

the logit model is obtained, which corresponds to the logistic distribution function as response function:

$$h(\eta) = \frac{exp(\eta)}{1 - exp(\eta)} = \pi$$

The logistic distribution function is symmetric and has somewhat heavier tails than the standard normal distribution function used in probit models. The logit model is most commonly used when analyzing binary data, especially in medical applications.

3.4.6 Bayesian Generalized Linear Model

Bayesian methods are an increasingly popular choice for sufficiently flexible to accommodate the complex forms of relationships between the response variable and the predictors. In contrast to other statistical methods, Bayesian models assume the vector of unknown parameters to be estimated is random, rather than fixed, and vary according to some 'prior' distribution [87]. In this study, first, we fitted categorical covariates that have a linear effect on the response variable and compare them with the result from the semiparametric model and geo-additive model. The effect of the covariates on the response is modeled by a linear predictor as:

$$\eta_i = \omega_i{}'\gamma \qquad\qquad i = 1,2, \dots \dots, p \qquad\qquad (3.7)$$

where:

$\omega_i = \left(\omega_{i1}, \omega_{i2}, \dots \dots, \omega_{ip}\right)'$ is a vector of categorical covariates

$\gamma = \left(\gamma_0, \gamma_1, \dots \dots \dots, \gamma_p\right)'$ is a vector of regression coefficients for the categorical covariates.

3.4.7 Bayesian Semiparametric Model

Generalized Additive Models are methods and techniques developed and popularized by Hastie and Tibshirani [88]. In our study, however, the data contain detailed information on continuous covariates like body mass index of the mother, mother age, and child age in a month. The practical experience has shown that this continuous covariate often have nonlinear effects as suggested in [14, 89–91]. Therefore, in this study, we examine the generalized additive model as an alternative to the common linear model in the context of analyzing childhood anaemia status in Ethiopia,

which can simultaneously incorporate the usual linear effects as well as nonlinear effects of continuous covariates within a semi-parametric Bayesian approach. Semiparametric model [88] assume that, given $x_i = (x_{i1}, \ldots\ldots, x_{ik})$ and ω_i,the distribution of y_i belongs to an exponential family, with mean $\mu_i = E(y_i|x_{i,}\omega_i)$ linked to an additive semiparametric predictor η_i by

$$\mu_i = h(\eta_i)$$

$$\eta_i = f_1(x_{i1}) + f_2(x_{i2}) \ldots\ldots\ldots\ldots\ldots + f_k(x_{ik}) + \omega_i'\gamma \qquad (3.8)$$

Here h is a known link and $f_1, f_2, \ldots\ldots, f_k$ are possibly unknown nonlinear smooth functions of continuous covariates.

3.4.8 Bayesian Geo-additive Models

Structured Additive Regression Models

Structured additive regression (STAR) models provide a unified framework for extending classical models to a more flexible approach. This approach allows for the inclusion of the different types of covariates such as the spatial random effects and nonlinear effects in the linear predictor. STAR models are based on the framework of Bayesian generalized linear models [92, 93], which cover several well-known model classes as special cases include: Geo-additive Model, Generalized Additive, Generalized Additive Mixed Model, Varying Coefficient Models, and Geographical Weighted Regression

Bayesian structured additive regression (STAR) has been proposed in Fahrmeir, Kneib, and Lang [94] as a comprehensive class of semiparametric regression models with discrete responses and different types of covariates and corresponding effects. STAR models allow to combination of these different model classes and several extensions in a unifying framework that also facilitates the development of generally applicable inferential schemes. Bayesian Geo-additive models were subclasses of STAR models that consist of nonparametric effects of continuous covariates, spatial effects, and cluster-specific random effects in different combinations [59]. The assumption of a parametric linear predictor for assessing the influence of covariate effects on responses seems to

be rigid and restrictive in our practical application situation and also in many real statistically complex situations.

Besides, when practical experience has shown that metrical covariates often have nonlinear effects and data may have spatial dependence by nature, we are facing one of the following problems: -

In the application, for the continuous covariates in the data set, the assumption of a strictly linear effect on the predictor may not be appropriate. Another difficulty is that we have a spatial covariate in our models. Hence, it is necessary to seek a more flexible approach for estimating the metrical covariates by relaxing the parametric linear assumptions. This, in turn, allows assessing of the true functional form of the metrical effects of the data and this approach is referred to as the nonparametric regression model. To specify a nonparametric regression model, an appropriate function that contains the unknown regression function needs to be chosen. This choice is usually motivated by smoothness properties, which the regression function can be assumed to possess. To overcome these difficulties, the Geo-additive Bayesian model was considered which caters for all the shortcomings of most of the regression analysis as well as retains its flexibility to accommodate nonlinear and spatial effects. We replace the strictly linear predictor in 3.2 with a geo-additive predictor.

Observation model

Suppose that regression data consists of observations $(y_i, x_i, \omega_i, s_i), i = 1,2, \ldots\ldots n$ on a response $y_{i.}$ The response variables in this application used the logit model in the case of childhood anaemia. In this application ω_i include a vector of fixed (categorical) effects, which are coded in effect code such as such child sex, educational level of the mother, …, etc., x_i represent the metrical(continuous) covariates include the child age in months and mother's BMI and the spatial covariate, s_i, which including the region in which the most of child's anaemia would be considered. Bayesian Geo-additive models of Kamman and Wand [59] assume that, given $x_i = (x_{i1}, .. x_{ip,})$, $\omega_i = (\omega_{i1}, \ldots, \omega_{ip})$ and $s_i = 1, \ldots, S$ labeling the region of in the country, the distribution of y_i belongs to an exponential family, with mean $\mu_i = E(y_i|x_i, \omega_i, s_i)$ linked to an

additive predictor η_i by an appropriate response function h. We assume a semiparametric regression model with geo-additive predictors η_i by

$$\mu_i = h(\eta_i), \eta^{geo} = f_1(x_{i1}) + f_2(x_{i2}) + , \ldots . + f_p(x_{ip,}) + f_{spat}(s_i) + \omega_i{}'\gamma. \quad (3.9)$$

Here h is a known response function, and $f_1, f_2, \ldots \ldots f_p$ are possibly nonlinear functions of metrical covariates and f_{spat} is the effect of the spatial covariate $s_i \ \epsilon$ 1,….., S labeling the region in the country. Regression models with predictors as in 3.9 are referred to as geo-additive models. These random effects are incorporated in the model to capture extra variation. Thus, to capture unobserved influential factors that vary across the regions, the model accounts for the structured random effects and unstructured random effects account for unobserved heterogeneity within each region. In a further step, we may split up the spatial effect f_{spat} into a spatially correlated (structured) and uncorrelated spatial (unstructured) effect.

$$f_{spat}(s_i) = f_{str}(s_i) + f_{unstr}(s_i) \quad (3.10)$$

One rationale is that a spatial effect is usually a surrogate of many unobserved influences, some of them may obey a strong spatial structure and others may be present only locally. Also, the two components are assumed to have independent prior distributions [95]. By estimating a structured and an unstructured effect we attempt to separate these effects.

3.4.9 Specification of Prior distribution

The Bayesian approach provides a cohesive framework for mixing complex data and external knowledge[96]. Within this framework, the models are assigned appropriate mixing probability distribution. This probability is determined by the prior distribution, assigned before the data are observed. In Bayesian inference, the unknown smooth functions f_j, the fixed effects parameters γ as well as the variance parameter σ^2 are considered as random variables and therefore, have to be supplemented by appropriate priors' distribution.

Suppose that $(f(1), f(2) \ldots \ldots f(n))'$the vector of corresponding function evaluations at observed values of x. Then, the general form of the prior for $f$is

$$f|\tau^2 \propto exp\left(-\frac{1}{2\tau^2}f'Kf\right) \quad (3.11)$$

Where *K* is a penalty matrix that penalizes too abrupt jumps between neighboring parameters. In most cases *K* will be rank deficient, therefore the prior for *f* would be improper. This implies that $f|\tau^2$ follows a partially improper Gaussian prior $f|\tau^2 \sim N(0, \tau^2 K^-)$ where K^- is a generalized inverse of a band-diagonal precision or penalty matrix *K*.

In the frequentist approach, the smoothing parameter is the equivalent with the variance parameter τ^2 which controls the tradeoff between flexibility and smoothness. To estimate the smoothness parameter f, a highly dispersed but proper hyperprior is assigned to τ^2. The proper prior for τ^2 is required to obtain a proper posterior for f [97]. The variance parameters must have distributions on the positive real line. The gamma, inverse gamma, or uniform families are often the noninformative distributions in this range. The common are choices are highly dispersed inverse gamma distribution with hyperparameters a and b, i.e.

$$p(\tau^2) \sim IG(a, b)$$

A particular prior depends on the type of the covariates and prior beliefs about the smoothness of f. Furthermore, a prior for a function f is defined by specifying a smoothness prior, and the hyperparameters a and b of the inverse gamma prior for τ^2.A possible choice for a and b is very small $a = b$, for example, a = b = 0.0001, leading to almost diffuse priors for the variance parameters. An alternative proposed, for example, in Besag *et al*. [98] is a = 1 and small value for b, such as b = 0:005. The choice of such a highly dispersed but proper prior avoids problems arising with improper priors [99].

Priors for Fixed Effects

The noninformative prior is the type of prior distribution that is assumed not to make a strong preference over the data [100]. According to Lawson[100], the prior choice can be usually made based on some general understanding of the range and behavior of the variable. For the parameter

vector γ of fixed effects the usual approach is to assign diffusepriors (uninformative prior) to the parameters of the fixed effects, that is:

$$\gamma_i \propto const, \qquad j = 1,2, \ldots . r.$$

Another choice would be to work with a multivariate Gaussian distribution $\gamma \sim N(\gamma_0, \Sigma_0)$ In this application, we used diffuse priors for the fixed effects.

Priors for Metrical (Continuous) Effects

Several alternatives are available to specify the priors of the unknown (smooth) functions f_j,j= 1,2 ….. p. These are basis function approaches with adaptive knot selection and approaches based on smoothness priors. Also, several alternatives have been recently proposed for specifying a smoothness prior for the effect f of metrical covariate x. These are random walk priors [101], Bayesian smoothing splines [97], and Bayesian P-splines[102].

Random Walk Models

In models where parametric modeling is not sufficient, a more flexible approach is adopted. Ideally, this approach is used to handle covariates differently, such as allowing for nonlinear effects for continuous covariates which the data may contain. These continuous covariates are modeled with a semiparametric and generalized additive approach. Such models are used to describe smooth curves in time or surface in space [103]. Similar to spatial area effects, metrical covariates are assigned specific priors to allow smoothing. Several alternative specifications are available for smoothness prior functions of metrical covariates, the commonly used priors for smooth functions are first or second-order random walk models, but we focus on the second-order random walk model.

Let us consider the case of a metrical covariate *x* with *equally-spaced observations* $x_i, i = 1,2 \ldots .. m$, $m \leq n$. Then $x_{(1)} < x_{(2)} < \cdots .. < x_{(m)}$ defines the ordered sequence of distinct covariate values. Here m denotes the number of different observations for x in the data set. A common approach in dynamic or state-space models is to estimate one parameter $f(t)$ for each distinct

$x(t)$; i: e Define, $f(t) = f\left(x_{(t)}\right)$ and let $f = (f(1), \ldots., f(t), \ldots\ldots., f(m))'$denote the vector of function evaluation. Fahrmeir and Lang [103] show that a first and second-order random walk smoothness prior to Normal errors would be specified as:

Then a first-order random walk prior for $f$is defined by

$$f(t) = f(t-1) + u(t) \qquad (3.12)$$

A second-order random walk is given by

$$f(t) = 2f(t-1) - f(t-2) + u(t) \qquad (3.13)$$

$$u(t) \sim N(0; \tau^2)$$

with diffuse priors $f(1) \propto const$ and $f(2) \propto const$, for initial values, respectively. A first-order random walk penalizes too abrupt jumps $f(t) - f(t-1)$ between successive states. While a second-order random walk penalizes large deviations from the linear trend $2f(t-1) - f(t-2)$. Also, the variance τ^2 controls the degree of smoothness f. Thus, the conditional prior distribution of $f(t)$ given its immediate past $f(t-1)$ is given by:

$$f_t | f_{t-1}, \tau^2 \sim N(f_{t-1}, \tau^2) \qquad (3.14)$$

Moreover, Random walk priors may be equivalently defined in a more symmetric form by specifying the conditional distributions of function $f(t)$ given its left and right neighbors. That means we can write the prior in (3.12 and 3.13) in a general form as

$$f|\tau^2 \propto exp\left(-\frac{1}{\tau^2} f'Kf\right) \qquad (3.15)$$

The penalty matrix is of the form $K = D\,D'$ where D is a first or second-order difference matrix. Here the design matrix K is the penalty matrix that penalizes too abrupt jumps between neighboring

parameters. More often, K is not full rank and this implies that $f|\tau^2$ follows a partially improper Gaussian prior

$$f|\tau^2 \sim N(0, \tau^2 K^-)$$

where K^- is a generalized inverse of the penalty matrix K.

For the case of nonequally spaced observations, random walk or autoregressive priors have to be modified to account for non-equal distances $\delta_t = x(t) - x(t-1)$ between observations. Random walks of the first order are now specified by

$$f(t) = f(t-1) + u(t) \qquad (3.16)$$

$$u(t) \sim N(0; \delta_t \tau^2)$$

i.e., by adjusting from τ^2 to $\delta_t(\tau^2)$.

Random walks second order are

$$f(t) = \left(1 + \frac{\delta_t}{\delta_{t-1}}\right) f(t-1) - \left(\frac{\delta_t}{\delta_{t-1}}\right) f(t-2) + u(t) \qquad (3.17)$$

$$u \sim N(0; w_t \tau^2)$$

where w_t, is an appropriate weight. Several possibilities are conceivable for weights. The simplest one is $w_t = \delta_t$ for the first-order random walk.

Spatial Covariates

Conditional Autoregressive Model

The conditional autoregressive model has been widely used in the field of epidemiology and other studies of diseases and was developed by Besag [69]. These models are also known as the Markov random field (MRF) model and are in the class of the Gaussian Markov random field (GMRF) models. In the spatial modeling for administrative regional areal data, such as disease mapping,

the MRF models are commonly employed. The virtual common form of MRF incorporates the structure of spatial dependence, based on the idea that areas that share a border or boundary are regarded as neighbors. The neighboring areas are bound to have too many similarities to those far apart. Thus, smoothing of the health outcome risk for an areal unit depends on its neighbor's risk. In this research, we focus on the MRF model for the incorporation of the spatially structured random term.

Consider first that the spatial index $s\epsilon\{1,2,\ldots.S\}$ represents a location or site in connected geographical regions. It is assumed that neighboring sites that share boundaries are more homogenous than any other arbitrary sites. Therefore, for a valid prior definition, a set of neighbors must be defined for each site s. Hence sites s and t are neighbors if they share a common boundary. Depending on the application, the spatial effect may be further split into a spatially correlated (structured) and an uncorrelated (unstructured) effect, i.e., $f_{spat} = f_{str} + f_{unstr}$. A rationale is that a spatial effect is usually a surrogate of many unobserved influential factors, some of them may obey a strong spatial structure while others may exist only locally. Besag *et.al* [69] proposed a Markov random field prior for the correlated spatial effects f_{str}. The spatial smoothness prior of function evaluations f_{str}(s) is

$$f_{str,s}|f_{str,t,}t \neq s, \tau^2 \sim N\left(\sum_{t\in\delta_s}\frac{f_{str,t}}{N_s}, \frac{\tau^2{}_{str}}{N_s}\right), \tag{3.18}$$

where N_s are the number of adjacent sites and $t \in \delta_s$ denotes, that site f_s is a neighbor of the site $f_{t,}$. Thus the (conditional) mean of f_s is an unweighted average of function evaluations of neighboring sites. Note that spatial data conditioning is undirected since there is no natural ordering of different sites f_s as in the case for metrical covariates.

In a general form, (3.18) can be given by

$$f_{str,s}|f_{str,t,}t \neq s, \tau^2 \sim N\left(\sum_{t\in\delta_s}\frac{w_{st}}{w_{s+}}f_{str,t}, \frac{\tau^2{}_{str}}{w_{s+}}\right), \tag{3.19}$$

where w_{st} are known equal weights and w_{s+} denotes the marginal sum of w_{st} over the missing subscript. Such a prior is called a Gaussian intrinsic autoregression. The design matrix X_{str} is a

$n \times S$ incidence matrix whose entry in the i^{th} row and s^{th} the column is equal to one if observation i has been observed at location s and zero otherwise.

Unit- or cluster-specific heterogeneity

In many situations, we observe the problem of heterogeneity among clusters of observations caused by unobserved covariates. Neglecting unobserved heterogeneity may lead to considerably biased estimates for the remaining effects as well as false standard error estimates. Suppose now $x\epsilon\{1, \dots, K\}$ is a cluster variable indicating the cluster a particular observation belongs too. A common approach to overcome the difficulties of unobserved heterogeneity is to introduce additional Gaussian i.i.d. effects $f(x) = \beta_x$ with

$$\beta_x \sim N(0, \tau^2{}_{unstr}), \qquad x = 1, \dots\dots, K. \tag{3.20}$$

The design matrix $\boldsymbol{X}$ is again a $n \times K$-dimensional $0/1$ incidence matrix that represents the grouping structure of the data, while the penalty matrix is simply the identity matrix, i.e., $\boldsymbol{K} = \boldsymbol{I}$. From a classical perspective, (3.20) defines i.i.d. *random effects*. However, from a Bayesian point of view, all unknown parameters are assumed to be random, and hence the notation" random effects" in this context is misleading. Hence, one may also think of (3.20) as an approach for modeling an unsmooth function.

Formally, the priors for f_{str} and f_{unstr} can both be brought into the form (3.15). For f_{str}, the elements of K given by $K_{ss} = w_{s+}$. and

$$K_{st} = \begin{cases} w_s = -1 \; where \; t \in \delta_s \\ 0, \qquad\quad other \; wise \end{cases}$$

For f_{unstr} we may set $K = I$.

Furthermore, the inverse Gamma priors are assumed for $\tau^2{}_{unstr}$ $[IG(a_{unstr}, b_{unstr})]$ and $\tau^2{}_{str}$ $[IG(a_{str}, b_{str})]$.

3.4.10 Models Specification

The following set of models were examined to investigate the linear, spatial, and nonlinear effects of metric covariates on childhood anaemia. The first model is a Bayesian Generalized Linear Model which incorporates fixed effects of categorical covariates. This model is given by

Model 1: $\eta_i = \omega_i{}'\gamma$ (3.21)

The second model is Bayesian Semiparametric Model given by

Model 2: $\eta_i = f_1(x_{i1}) + f_2(x_{i2}) \ldots\ldots\ldots\ldots + f_k(x_{ik}) + \omega_i{}'\gamma$ (3.22)

this model is similar to Model 1, it accounts for fixed effects of categorical covariates, and assumes nonlinear effects of child age and mother BMI which are continuous covariates of an individual. The final (Bayesian Geo-additive Models) model is a structured additive model which captures spatial heterogeneity for unobserved influential factors and also accounts for nonlinear effects of child age and mother BMI and the effects of categorical covariates. The model is given by

Model 3: $\eta^{geo} = \omega_i{}'\gamma + f_1(x_{i1}) + f_2(x_{i2}) + f_{str}(s_i) + f_{unstr}(s_i)$ (3.23)

Here spatially structured random effects account for unobserved covariates across the region or spatial location in general and the unstructured heterogeneity caters for unobserved influential covariates that are inherent within the regions.

In all the models' formulation in this section, we assumed an independent diffuse prior for the fixed effects $\gamma \propto const$, the smooth functions (f_1 and f_2) of continuous covariates child age and mother BMI were both assigned second-order random walk priors discussed in Section 3.13.

Second-order random walk priors which permit enough flexibility while avoiding overfitting the data was suggested by Wecker and Ansley [104] and our prior was supported by Gebrenegus and Kandala, considered spatial modeling of under-five mortality in Nigeria, based on data from Nigeria Demographic and Health Survey (NDHS) [50], adopted independent diffuse prior for the

fixed effects and second-order random walk priors for the smooth functions (f_1 and f_2) of continuous covariates.

Furthermore, the spatially structured effects $f_{str}(s_i)$ were assigned Markov random fields (MRFs) prior (Equation (3.18)) and the spatially unstructured $f_{unstr}(s_i)$ were assigned the i.i.d Gaussian prior (Equation (3.20)). Rue and Held [105] suggested that if the spatial data are in form of discrete, for cluster-specific heterogeneity, the Gaussian random field priors, and structured spatial effect, we assume spatial correlations defined implicitly by assuming a Markov random field prior for a suitable neighbourhood structure derived from the spatial orientation of the data. The most common case would be to treat regions as neighbours if they share a common boundary.

For variance parameter τ^2 we assigned an inverse gamma $\tau^2 \sim IG(a, b)$ (with $a, = b = 0.001$) to obtain a data-driven amount of smoothness and since the variance parameters must have distributions on the positive real line.

3.4.11 Inference

3.4.11.1 Fully Bayesian inference based on MCMC techniques

Statistical inference is done using Markov chain Monte Carlo techniques in a fully Bayesian setting. Fully Bayesian inference is based on the entire posterior distribution meanings all the unknown parameters are assumed to be random variables and are assigned priors and further hyperparameters are assigned hyperpriors. In FB, the unknown variance parameters ${\tau^2}_j$ are also considered as random variables supplemented with suitable hyperprior assumptions. The highly dispersed (but proper) inverse Gamma priors $p({\tau^2}_j) \sim IG(a_j, b_j)$ are assigned to the variances. The corresponding probability density function is given by

$$ {\tau^2}_j \propto ({\tau^2}_j)^{-a_j-1} exp\left(-\frac{b_j}{{\tau^2}_j}\right) $$

Using proper priors for ${\tau^2}_j$ (with $a_j > 0$ and $b_j > 0$) ensures propriety of the joint posterior despite the partial impropriety of the priors for the γ_j. A common choice for the hyperparameters are small

values for a_j and b_j, e.g., $a_j = b_j$ =0.001. In some situations, the estimated nonlinear functions f_j may depend considerably on the particular choice of hyperparameters a_j and b_j. This may be the case for a very low signal-to-noise ratio and/or a small sample size. Bayesian inference is based on the posterior of the model given by

$$p(\beta_1, \ldots., \beta_p, {\tau_1}^2, \ldots, {\tau_p}^2, \gamma|y) \propto L(y, \beta_1, \ldots., \beta_p, \gamma) \prod_j^p \{p(\beta_j|{\tau_j}^2)p({\tau_j}^2)\} \quad (3.24)$$

Where $L(.)$ denotes the likelihood which, under the assumption of conditional independence, is the product of individual likelihood contributions.

3.4.11.2 Markov Chain Monte Carlo (MCMC) Methods

In many practical situations (and in particular for most structured additive regression models) the posterior distribution is numerically intractable. A technique that overcomes this problem is the Markov Chain Monte Carlo (MCMC) simulation method that allows drawing of random samples from the posterior. From these random samples, characteristics of the posterior such as posterior means, standard deviations, or quantiles can be estimated by their empirical analogs. Instead of drawing samples directly from the posterior, MCMC devices a way to construct a Markov chain with the posterior as stationary distribution. Hence, the iterations of the transition kernel of this Markov chain converge to the posterior yielding a sample of dependent random numbers. Usually, the first part of the sample (the burn-in phase) is discarded since the algorithm needs some time to converge [102]. Also, some thinning is typically applied to the Markov chain to reduce autocorrelations.

Bayesian inference via MCMC is based on updating full conditionals of single parameters or blocks of parameters, given the rest and the data. For Gaussian models, Gibb's sampling with so-called multi-move steps can be applied. For non-Gaussian responses, Gibb's sampling is no longer feasible and Metropolis-Hastings's algorithms are needed. More details can be found in [101]. let α denote the vector of all unknown parameters in the model. Then, the logit model is given:

$$P(\alpha|y)$$

$$\propto \prod_{i=1}^{n} L_i\,(y_i;\eta_i) \prod_{j}^{p} \{(\beta_j|{\tau_j}^2)p({\tau_j}^2)\}\, p(f_{str}|{\tau_{str}}^2)p(f_{unstr}|{\tau_{unstr}}^2) \prod_{j=1}^{r} p(\gamma_j)\;\; p(\sigma^2). \qquad (3.25)$$

where $\beta_j;\ j = 1\ldots\ldots\ldots\ p;$ are the vectors of regression coefficients corresponding to the functionsf_j.While the full conditionals for the variance components $\tau^2;\ j = 1,\ldots\ldots p,\ str,\ unstr$ and σ^2 are inverse gamma distributions. More details can be found in [106–108].

The estimation of models is based on different sampling schemes depending on the distribution of the response. For non-gaussian responses here, we now turn the attention to general responses from an exponential family. In this case, the full conditionals are no longer Gaussian. For fixed effects and i.i.d. random effects we use a slightly modified version of the Metropolis-Hastings algorithm which correctly applied for non-Gaussian data and if the posterior distribution doesn't follow some known distribution (no conjugate distribution) based on iteratively weighted least squares(IWLS) proposal suggested by Brezger and Lang, [109]. In addition, Fahrmeir and Lang propose a MH-algorithm for updating unknown regression parameters based on conditional prior proposals[101].

The basic idea behind IWLS proposals is to combine Fisher scoring or IWLS for estimating regression parameters in generalized linear models, and the Metropolis-Hastings algorithm[93]. More precisely, the goal is to approximate the full conditionals of regression parameters β_j and γ by a Gaussian distribution, obtained by accomplishing one Fisher scoring step in every iteration of the sampler. Suppose we want to update the regression coefficients β_j of the function f_j with current state ${\beta^c}_j$ of the chain. Then, according to IWLS, a new value ${\beta^p}_j$ is proposed by drawing a random number from the multivariate Gaussian proposal distribution $q\left({\beta^c}_j, {\beta^p}_j\right)$.

The acceptance rates are significantly higher for the sampling scheme of IWLS-proposals based on the current posterior mode ${m^c}_j$ rather than the current ${\beta^c}_j$. This is particularly useful for updating spatial effects based on Markov random fields where, in many cases, a sampling scheme based on the current state of the chain yields quite low acceptance rates.

3.4.11.3 Model Comparison and Selection

Deviance Information Criteria (DIC)

The classical approach to model comparison involves a trade-off between how well the model fits the data and the level of complexity. Spiegelhalter *et al*. [110] devised a selection criterion that was based on Bayesian measures of model complexity and how good the fit of a model is for the data. The deviance information criterion (DIC) proposed by Spiegelhalter is a commonly used tool for model comparison and assessment. It is a generalization of the AIC. The DIC has become a popular model comparison criterion in a fully Bayesian (FB) context [110]. A complexity measure, pD is suggested by using an information theoretic argument to get more effective number of parameters in a model. As the difference between the posterior mean of the deviance and the deviance at the posterior estimates of the parameters of interest.

pD is assumed to be an approximate trace of the product of Fisher's information and the posterior covariance matrix. It could be obtained through a Markov Chain Monte Carlo analysis. In the case of normal models, pD corresponds to the trace of 'hat' matrix projection observations onto fitted values. In an exponential family model, $\overline{D}$ which calls for posterior mean deviance, can be taken as a measure of fit. Assume that $f(y)$ is a fully specified standardizing term, then

$$pD = \overline{D}(\bar{\theta}) - D(\bar{\theta}) \qquad (3.26)$$

$D(\theta) = -2\log p(y|\theta) + 2\log f(y)$, is Bayesian deviance.

A Deviance Information Criteria (DIC), which could be used for model comparison, is computed by adding the fit $\overline{D}$ to a complexity pD. DIC is defined as a "Plugin" estimate of fit plus twice the effective number of parameters, as follows:

$$DIC = D(\bar{\theta}) + 2pD = \overline{D} + pD, \qquad (3.27)$$

where the posterior mean of the deviance $\overline{D}(\theta)$ is penalized by the effective number of model parameters pD. Therefore, to select the best model among several fitted models in this study, DIC

was used. The advantage of DIC over other criteria, for Bayesian model selection, is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation. Assessing goodness of fit involves investigating how close the values are predicted by the model with that of observed values. The model goodness of fit was assessed based on the DIC that states that the smaller the value of the DIC the better is the model fit[110].

3.4.11.4 Model Diagnostic

Model diagnoses were performed based on MCMC post-estimation diagnosis, to examine the convergence of MCMC. Among several ways of a test convergence, the most popular and straightforward convergence assessment methods have been used for this study. The following methods were more likely considered for this study.

Autocorrelation plot: High correlation between the parameters of a chain tends to give slow convergence, whereas high autocorrelation within a single parameter chain leads to slow mixing and possibly individual nonconvergence to the limiting distribution because the chain tends to explore less space infinite time. In analyzing Markov chain autocorrelation, it is helpful to identify lags in the series to calculate the long-run trends in correlation, and in particular, whether they decrease with increasing lags[111].

Data management and relevant software of the study

To analyses the spatial pattern and determinants of childhood anaemia software such as SPSS 24. version: - were for Data cleaning and management, (renaming, recoding, and range checking of variables). R version 3.6.3 and ArcGIS 10.3: -were used for practical modeling and analysis.

CHAPTER FOUR

4 Results and Discussion

4.1 Descriptive statistics

4.1.1 Child Anaemia prevalence rate

From the total of 7953 children under age 5 years with anaemia test results were included in this study,3386 (42.6%) were non-anemic and 4567 (57.4%) are anaemic children, which presented in table 4.1 below.

Table 4.1: Frequency distribution of childhood anaemia status in Ethiopia

Anaemia status	Frequency	Percent
Not Anaemic	3386	42.6%
Anaemic	4567	57.4%
Total	7953	100%

4.1.2 Trends of childhood anaemia (2005 -2016)

The prevalence of anaemia among Ethiopian children declined from 54% to 44% from 2005 to 2011 but increased to 57.4% in 2016.

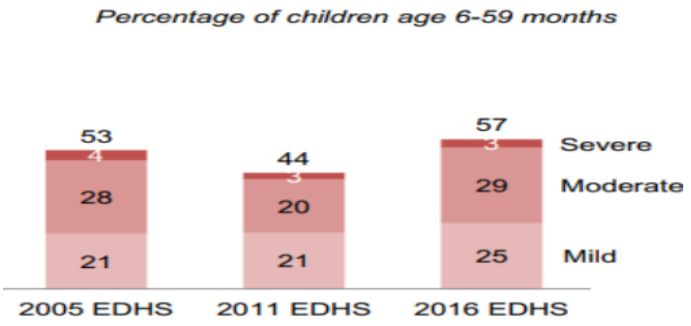


Figure 4.1 Trends in childhood anaemia

The percentage of children U5 years of age suffering from anaemia in the poorest family was high (62.8%) as compared with the wealth index status in the middle and rich categories, (53.7% and 52.0% respectively). Thus, lower wealth index status was allied with the increase in the danger of the development of anaemia in children. This association between the wealth index status of the family and anaemia among children under 5 was found to be statistically significant ($p<0.001$).

Results from Table 4.2 also revealed that proportions of children U5 years suffering from anaemia were 58.2%,57.2%, and 49.1% among those whose mothers had no education, educated up to the primary, and secondary, and above respectively. It was found that the lower the level of education of the mother, the higher the probability of the child suffering from anaemia. This association in the mothers' educational status and anaemia was found to be significant statistically ($p< 0.001$).

From the study population,393(49.8%) children with anaemia were from urban areas and 4174 (58.3%) were from the rural parts. A significant association was found between anaemia and place of residence ($p<0.001$). The Percentage of having anaemia is lower among urban children as compared to their rural counterparts. The region was also found to be significantly associated ($p<.001$), with the percentage of children who had anaemia ranging from 42.7% to 83.3%. The Greater prevalence in the Somali region (83.3%) and in the Afar region (74.7%) followed by Dire Dawa city (72.4%) and Harari region (66.7%) while the lowest percentage of children with anaemia in the Addis Ababa city (50%) followed by Benishangul (44.2%) and the Amhara region (42.7%). Hence, there appears to be some variation in the prevalence of anaemia among the region of Ethiopia (see Table 4.2).

As shown in table 4.2, a slightly more prevalence of anaemia was observed in female children (57.9%) than male children (57.0%). Regarding on size of child at birth, the highest prevalence of anaemia was observed among children whose size at their birth is smaller than average (61.6%) as opposed to the lowest prevalence of anaemia was recorded from children whose size at their birth is the larger than averages and average level (57.3% and 55.1%) respectively.

As observed in Table 4.2, the prevalence of anaemia among children U5 years also varies with their mother age groups. For instance, a higher proportion of anaemia was observed for children

whose mothers are under 15-24 year of age (63.1%), and the lowest proportion was found for children whose mother age group between 25 - 34 age (57.7%) followed by mother between 35-49 age (52.3%). Hence, as the age of the mother increased the prevalence of anemia among children U5 years was decreased.

From results found in Table 4.9, source of drinking water had also a significant association with anemia status among under five years children. The proportion of anemia status was also high for children who did not use improved water which is around (54.7%) and (58.4%) for those who used improved water respectively. The prevalence of anemia among children U5 years was 57.1% for children who lived with 0-3 U5 children in the household and 69.9% for children who lived with 4-6 U5 children in the household.

Table 4.2: The description of the socio-economic, demographic, and environmental factors of childhood anaemic status in the regional states of Ethiopia.

		Anaemia status		
Variables Names	Category	Not Anaemic (%)	Anaemic (%)	Total
Region	Tigray	232(45.7)	276(54.3)	508
	Afar	20(25.3)	59(74.7)	79
	Amhara	890(57.3)	664(42.7)	1554
	Oromia	1205(34.8)	2257(65.2)	3462
	Somali	55(16.7)	274(83.3)	329
	Benishangul	48(55.8)	38(44.2)	86
	SNNPR	844(48.8)	886(51.2)	1730
	Gambela	8(44.4)	10(55.6)	18
	Harari	5(33.3)	10(66.7)	15
	Addis Ababa	72(50)	72(50)	144
	Dire Dawa	8(27.6)	21(72.4)	29
Residence	Urban	396(50.2)	393(49.8)	789
	Rural	2989(41.7)	4174(58.3)	7163
Gender of Child	Male	1769(43.0)	2346(57.0)	4115
	Female	1616(42.1)	2221(57.9)	3837

Mother occupation	No	2426(41.7)	3393(58.3)	5819
	Yes	959(45.0)	1174(55.0)	2133
Mother Education level	No education	2226(41.8)	3105(58.2)	5331
	Primary	920(42.8)	1230(57.2)	2150
	Sec. and above	239(50.9)	231(49.1)	470
Currently breastfeeding	No	1220(48.6)	1292(51.4)	2512
	Yes	2166(39.8)	3275(60.2)	5441
Taking iron pills, sprinkles/syrup	No	3058(42.4)	4151(57.6)	7209
	Yes	327(44.0)	416(56.0)	743
Fever in last two weeks	No	2970(43.9)	3794(56.1)	6764
	Yes	415(34.9)	773(65.1)	1188
Had diarrhea recently	No	2995(43.2)	3941(56.8)	6936
	Yes	390(38.3)	627(61.7)	1017
cough in last two weeks	No	2716(43.2)	3572(56.8)	6288
	Yes	670(40.2)	995(59.8)	1665
No. of child U5in HH	0-3	3320(42.9)	4416(57.1)	7736
	4-6	65(30.1)	151(69.9)	216
Birth size of child	S. than average	777(38.4)	1244(61.6%)	2021
	Average	1511(44.9)	1851(55.1)	3362
	L. than average	1097(42.7)	1471(57.3)	2568
Wealth Index	Poor	1377(37.2)	2328(62.8)	3705
	Middle	797(46.3)	926(53.7)	1723
	Rich	1212(48.0)	1313(52.0)	2525
Mother age	15-24	590(36.9)	1009(63.1)	1599
	25-34	1841(42.3)	2512(57.7)	4353
	35-49	955(47.7)	1046(52.3)	2001
House Hold size	1-4	836(42.9)	1114(57.1)	1950
	5-8	2140(43.4)	2789(56.6)	4929
	9 and above	409(38.1)	664(61.9)	1073
Husband occupation	Not Working	239(36.6)	414(63.4)	653

	Working	3146(43.1)	4153(56.9)	7299
Mother Anaemia level	Not anaemic	2633(47.5)	2909(52.5)	5542
	Anaemic	752(31.2)	1658(68.8)	2410
Types of toilet facility	Not improved	1180(38.7)	1870(61.3%)	3050
	Improved	2205(45.0)	2697(55.0)	4902
Sour. of drinking water	Not Improved	2464(41.6)	3454(58.4)	5918
	Improved	922(45.3)	1113(54.7)	2035
Husb. education level	No education	1563(41.4)	2211(58.6)	3774
	Primary	1389(42.8)	1856(57.2)	3245
	Sec. and above	408(47.0)	460(53.0)	868
	Unknown	25(38.5)	40(61.5)	65

4.2 Spatial Distribution of Childhood Anaemia in Ethiopia

Of all the regions in the country, the Somalia regions recorded the highest prevalence of anemia in children (see Figure 4.2, Table 4.2), almost 83.3% of children in these regions were anaemic. The second highest prevalence was recorded in the Afar region,74.6%, followed by the Dire Dawa region, 72.40%. The lowest prevalence was recorded in the Amhara region, 42.7%. The result shows that the prevalence of anaemia varies from region to region. Given this high prevalence between most of the regions, further investigation was needed to check that region that is correlated with that of neighboring regions contributing to childhood anaemia.

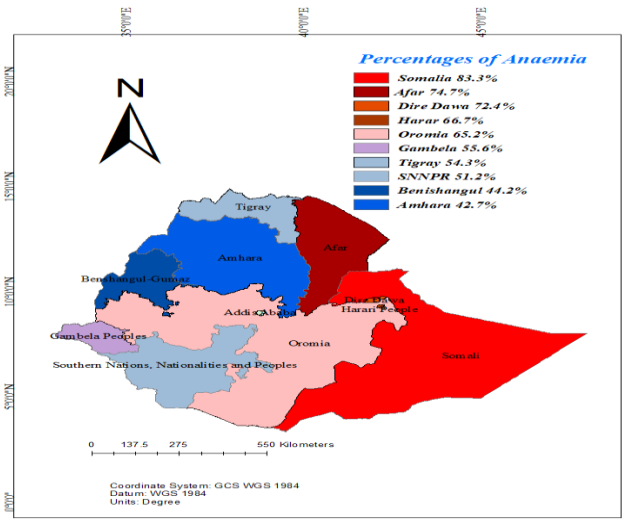


Figure 4.2: Spatial Distribution of Childhood Anaemia in Ethiopia

4.3 Testing for Spatial Autocorrelation

In this study, the distribution of childhood anemia across the country among children under 5 age was also investigated. To get a general insight into the spatial clustering of anaemia, a global spatial statistic was estimated using Moran's *I* statistic and Geary C test statistic. This was done after establishing the number of anaemia cases in each of the clusters using the ArcGIS 10.3 version and the results are found in table 4.3 below.

Table 4.3: Results of Global Moran's I and Geary's C statistics

Assumption	Coefficient	Observed	Expected	Std dev	p-value
Normality	Moran, I test	0.478	-0.100	2.960	0.002
Normality	Geary C test	0.471	1.000	1.644	0.050
Randomization	Moran, I test	0.478	-0.100	2.915	0.002
Randomization	Geary C test	0.471	1.000	1.789	0.037

The test result showed the presence of significant global positive spatial autocorrelation for the prevalence of anaemia (*I* = 0.478, P-value <0.0015, and Geary c =0.4706, P-value < 005011). Based on the global Moran's *I* statistic, Geary C test statistic result and P-values of the reported

Moran's *I* and Geary's *C* coefficients, we can reject the null hypothesis of no spatial auto-correlation which indicating the existence of significant positive spatial autocorrelation (clustering). Both results suggest that there is spatial dependence in our data which needs to be further explained by including spatial dependency in our model.

In order to visualize global spatial autocorrelation, we use also Moran's scatter plot under the assumptions of normality (Figure 4.3). The figure also shows that anaemia prevalence is spatially correlated with neighboring values.

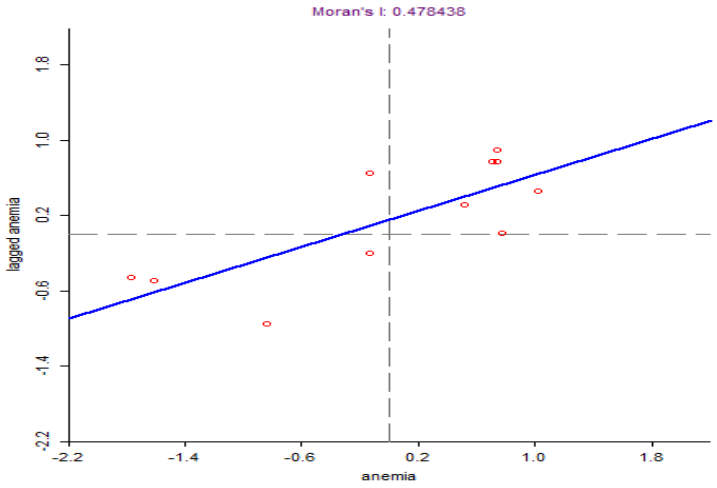


Figure 4.3: Global Moran's *I* Scatter Plot for Anaemia prevalence

4.4 Diagnosing Nonlinearity

Scatterplot (partial-residual plots)

Before we consider continuous independent variables in our model in linear form, we had checked their linearity on response variable using scatterplot especially partial-residual plots. Component-plus-residual plots, also called partial-residual plots, are a simple graphical device that can be effective in detecting departures from linearity and detecting the need to transform a predictor. The plots in Figure 4.9 are for child age, Maternal age, and MBI, using child anaemia status as the response. Both for child age and mother body mass index plots look nonlinear. The partial relationship of child anaemia to child age and BMI is simply tending to increase at the lower ages

and BMI and turning back down at the higher ages and BMI. Whereas for mother age it looks linear so we fit this variable in fixed effects (see Appendix A).

4.5 Model-Based Data Analysis

4.5.1 Model fit criterion for model comparison

The table (table 4.4) provides the model fitted and the DIC values which used to select the best fitting model, the smaller the DIC value the better the fit. The results reveal that Model 3 which include both structured and unstructured spatial effects is the more preferred model, with the DIC value given by (DIC = 8563.32), where Model 1, Model 2, and Model 3 are Bayesian Generalized linear Model**,** Bayesian Semiparametric Model, and Bayesian Geo-additive Models respectively.

Table 4.4: Summary of the model fit criterion for model comparison for all the fitted models.

Models	$\bar{D}$	**pD**	**DIC**
Model1	8815.65	19.67	8854.97
Model2	8783.91	33.68	8851.26
Model3	8477.71	42.81	8563.32

4.5.2 Results of Bayesian Geo-additive Analysis

Starting with very simple models, we increase complexity to show what can be gained by more sophisticated approaches. A flexible approach was adopted for such models that allows capturing of different types of covariates. One of interest was the incorporation of spatial random effects which allow for correlated and uncorrelated heterogeneity, significant effects of nonlinear and categorical covariates. Furthermore, these models should be best in terms of DIC too.

4.5.3 Fixed effects

The results based on Model 3 are more concerned to be presented and interpreted. The posterior odds ratio (POR) estimates and their corresponding 95% credible intervals (CI) are presented in Table 4.5. Results of this model revealed that that, most of the covariates such as mother anaemia

level, House Hold wealth index, type of toilet facility they used, mother education level, size of a child at birth and whether the child had a fever for the last two weeks has a significant effect on anaemic disease of childhood.

The study revealed that the household wealth index has significant effects on childhood anaemia prevalence. The odds of being anaemic were steadily dropped as household wealth increased. Children from medium-income households were found to have 27.2% times lower odds of being anaemic compared to those from the poorest households (POR = 0.728, CI: 0.622, 0.859), while holding other variables in the model constant. The odds further dropped as the children in the richest households had 29.46% lower odds of being anaemic (POR = 0.7054, CI: 0.609, 0.826), while holding other variables in the model constant. This can be shown that a child born to a rich family is less likely to be anaemic compared to a child born to a poor family.

The study also revealed that a significantly higher odds of being anaemic was observed for children living in households with not improved toilet facilities. Children living in households with improved toilet facilities had lower odds of being anaemic compared to those living in households with not improved toilet facilities (POR = 0.88; CI: 0.766, 0.997) while holding other variables in the model constant. Furthermore, children whose sizes were larger than averages at birth were found to have 18% times lower odds of being anaemic compared to those their size at birth was smaller than average (POR= 0.82; CI: 0.715, 0.937) while holding other variables in the model constant. This means that the risk of anaemia remains high for children whose size at birth was smaller than average than others.

The study also revealed that fever was also found to be a significant predictor of childhood anemia. A child who had a fever in the last two weeks before the survey was found to have 25% times higher odds of being anaemic compared to those who did not have a fever in the two weeks before the survey (POR=1.25; CI: 1.077, 2.937) while holding other variables in the model constant. It can be deduced that a child who had a fever in the two weeks before the survey is more likely to be anaemic than a child who did not have a fever in the two weeks before the survey.

The study also revealed that the education levels of mothers were the most significantly associated with of childhood anemia. It was noted from the results that childhood anaemia was negatively affected among children born to an illiterate mother compared to that of children that are born to a literate mother. The odds of being anaemic was 19.5% times lower in children whose mothers had secondary and above education as compared to children whose mothers had no education (POR= 0.805; CI: 0.649, 0.996), while holding other variables in the model constant.

The study also revealed that a mother's anaemia status has also a significant effect on her child's anaemia status. Children from anaemic mothers were found to have 58% times higher odds of being anaemic compared to those from non anaemic mothers (POR= 1.58; CI: 1.41, 1.763) while holding other variables in the model constant. This can be shown that a child born to an anaemic mother is more likely to be anemic than a child born to a non-anaemic mother.

Table 4.5: Posterior mean estimates of Bayesian geo-additive models

Fixed effects	Category	Estimate	Std. Error	Median	95%CI
(Intercept)		0.4221	0.1758	0.4251	(0.0622, 0.7715)
Mother Anaemia	Not Anaemic(ref)	0.00			
	Anaemic	0.4558	0.0596	0.4571	(0.3434, 0.5672)
HH Wealth Index	Poor(ref)	0.00			
	Middle	-0.317	0.0817	-0.3171	(-0.4752, -0.152)
	Rich	-0.3490	0.0811	-0.3498	(-0.4966, -0.191)
Toilet facility	Not Improved (ref)	0.00			
	Improved	-0.1327	0.0676	-0.1308	(-0.2663, -0.0034)
Mother Education level	No education(ref)	0.00			
	Primary	0.0060	0.0667	0.0064	(-0.1221, 0.144)
	Second. and above	-0.2175	0.1154	-0.2151	(-0.433, -0.0035)
Child Birth Size	S. than average(ref)	0.00			
	Average	-0.1248	0.0668	-0.2084	(-0.2545, 0.007)
	L. than averages	-0.2005	0.0711	-0.1966	(-0.3356, -0.065)
Had fever in the last 2 weeks	No(ref)	0.00			
	Yes	0.2239	0.0770	0.2235	(0.0745,0.3771)

Smooth terms variances:				
	Estimate	**Std. Error**	**Median**	**95%CI**
Effects Nonlinear of Metrical covariates				
sx(Mother BMI)	0.0007	0.0005	0.0005	(0.0002.0.0020)
sx(C. age of child in month)	0.0010	0.0008	0.0007	(0.0002,0.0035)
Spatial structured and unstructured Random effects variances				
sx (regions): mrf	0.3911	0.3623	0.3348	(0.0014,1.2762)
sx(regions):re	0.1322	0.1713	0.0712	(0.001, 0.6067)

4.7.4 Nonlinear Effects of child age and mother BMI on childhood anaemia

From our analysis, some effects have an unknown nonlinear form (such as child's age and mother's BMI). These two variables have a nonlinear effect on the response variables.

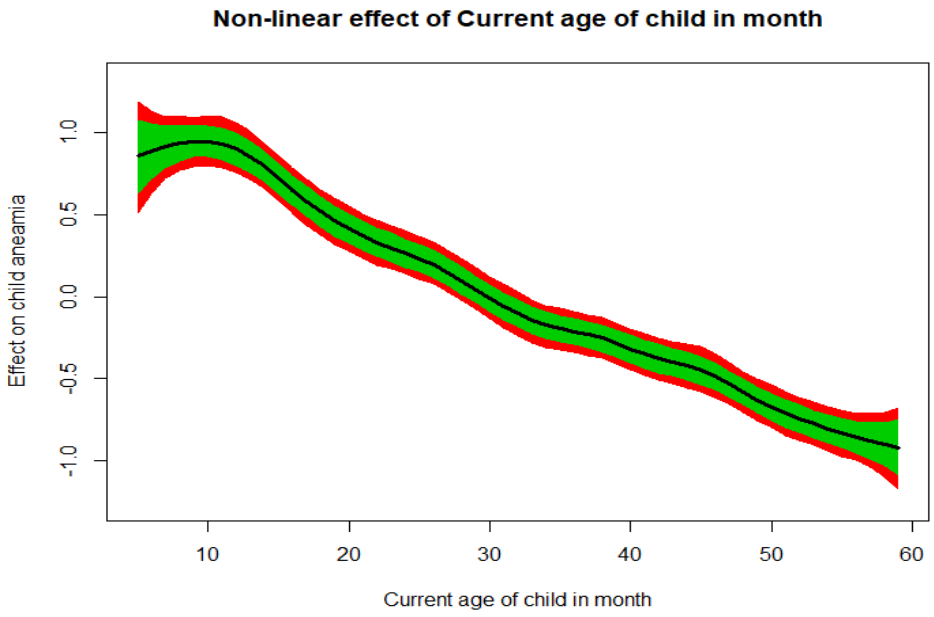


Figure 4.4 Non-linear effects of child age in months on childhood anaemia.

Note: red band (95% CI) and green band (80% CI).

Abbreviation: CI, credible interval

Each non-linear graph consists of a center line representing the posterior mean estimate bounded by 95% credible intervals (outer lines) and 80% credible intervals (inner lines). As shown in figure 4.4 above, child age has significant nonlinear effects on childhood anaemia, as child age increase, its effects on child anaemia decrease, meaning that the risk of anaemia was found to be highest among children of younger age. The chance of having anaemia is much higher in children aged about 6 months to 13 months and then decreases after this month.

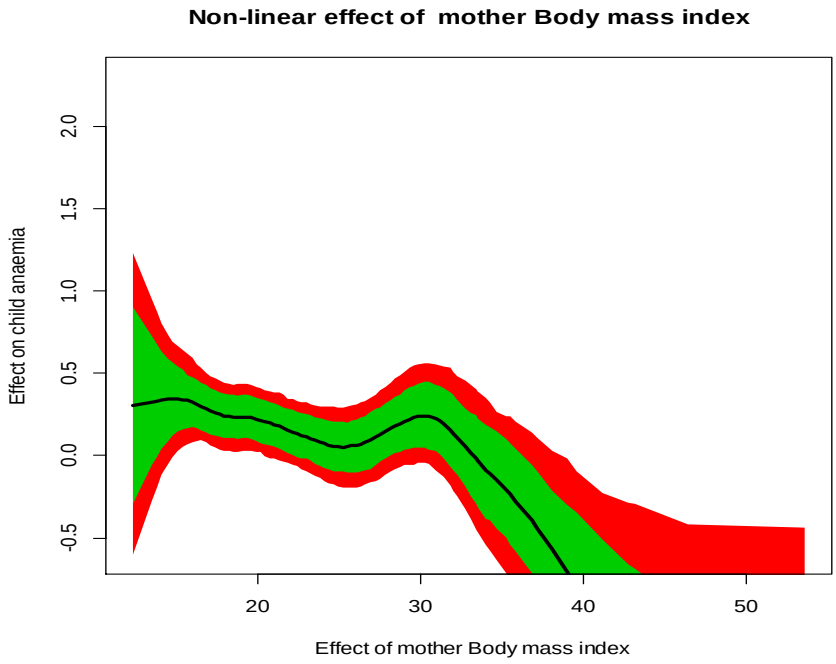


Figure 4.5 Non-linear effects of Mother Body mass index on childhood anaemia.

Note: red band (95% CI and green band (80% CI).

Abbreviation: CI, Credible Interval

As shown in figure 4.5 above, the effect of maternal body mass index (which measures the nutritional status of mothers) on childhood anaemia, which produced a similar trend line on the Hb concentration value of their children. It reveals that mother body mass index below 19 (underweight mother) produced increased childhood anaemia, then stabilizes in between 19 - 30, and decline of anaemia in children of maternal body mass index above 30 (overweight or obese mothers)

4.6.5 Geographical mapping of childhood anaemia (Spatial effects)

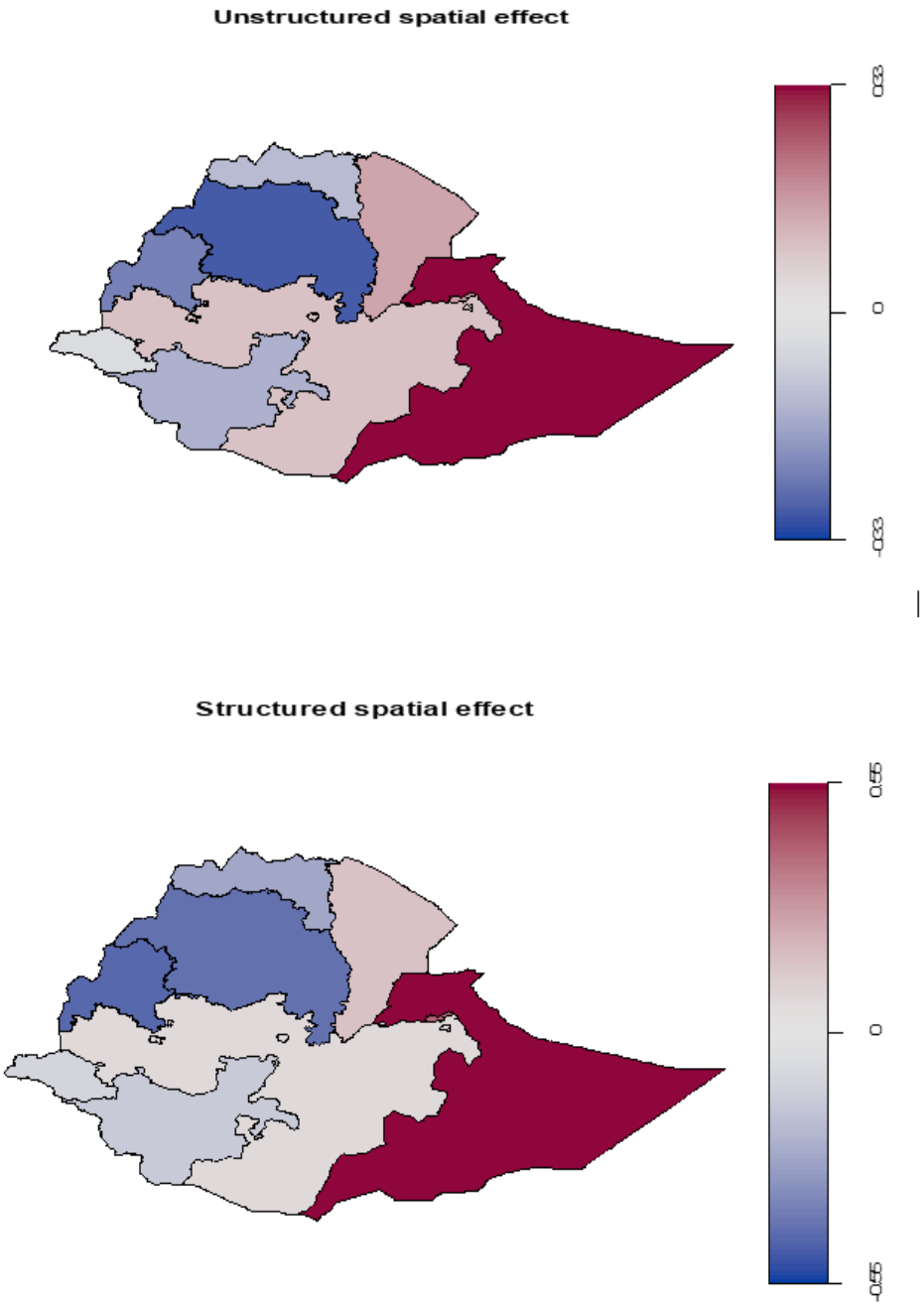


Figure 4.6 Estimated posterior means of the unstructured spatial effect (top) and structured spatial effect (bottom) on the log-odds of anaemia.

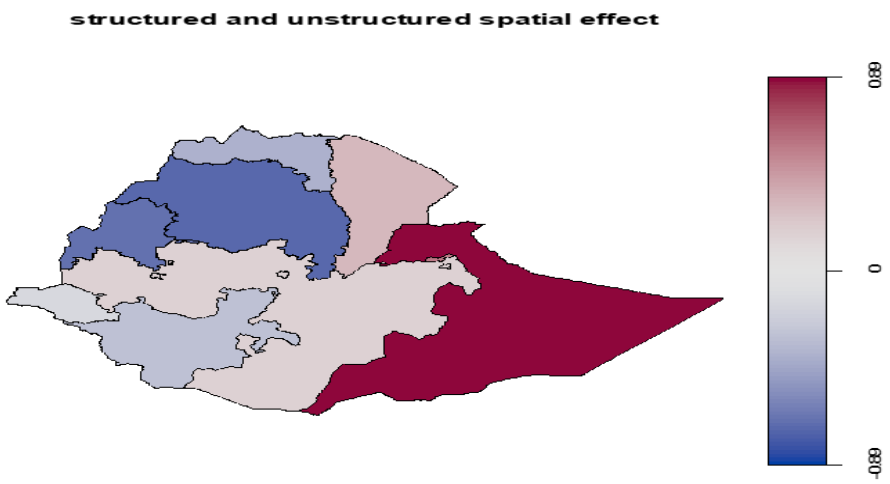


Figure 4.7: Estimated posterior means of both structured and the unstructured spatial effect on the log-odds of anaemia

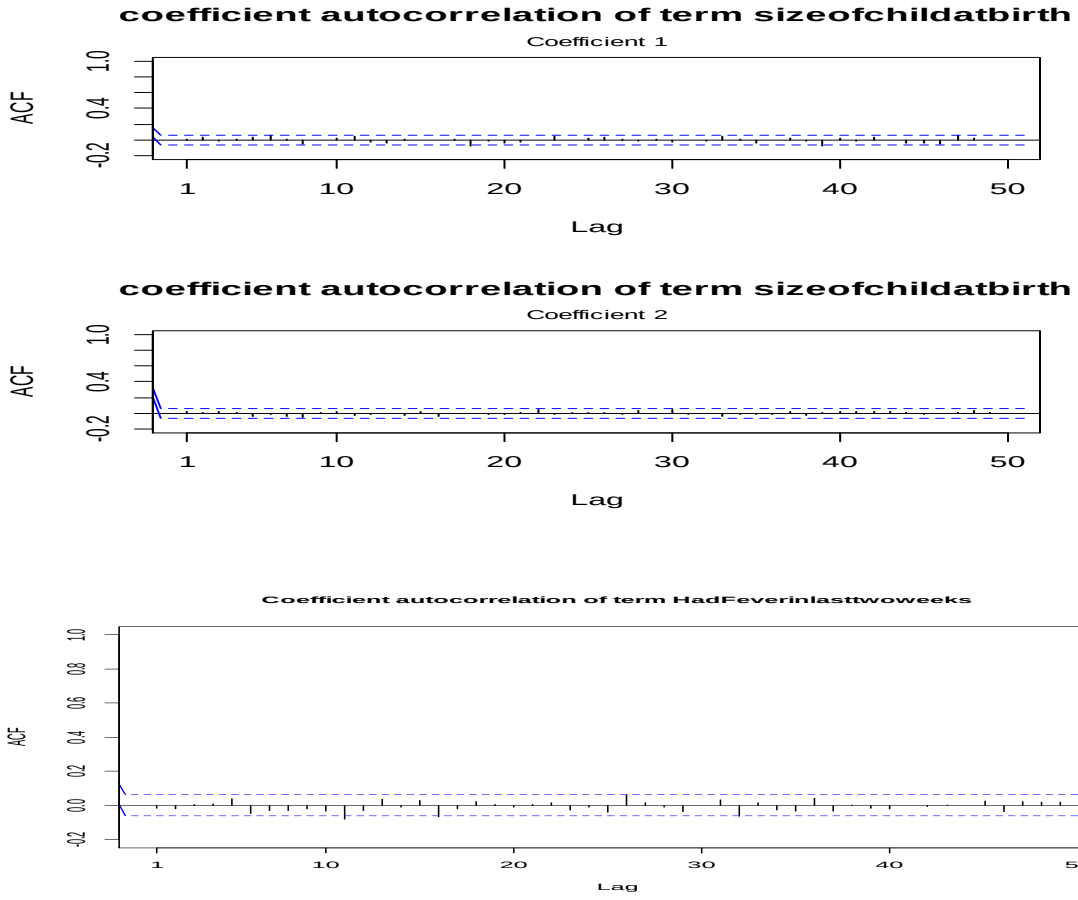
From the estimated posterior means of the structured and unstructured spatial effects on the log-odds of anaemia, the structured spatially correlated effect (0.3911) exceeds the unstructured effect (0.1322) which indicates that there is a spatial dependency in our data. From the map (fig 4.6 and fig 4.7), the blue region has a negative spatial effect and are therefore the north- western and northern parts of the country were associated with a lower odds of childhood anaemia and the red region have a positive spatial effect and are therefore the eastern and north-eastern parts of the country were associated with a higher odds of childhood anaemia.

The structured spatial effect, which ranged from -0.55 to 0.55, was high in comparison to the unstructured spatial effect, which ranged from -0.33 to 0.33. Furthermore, the effects of spatially correlated factors contributing to childhood anaemia in all the regions were statistically significant. Therefore, failure to take into consideration the posterior uncertainty in the spatial location would invariably lead to an overestimation of the precision in predicting childhood anaemia risks in the unsampled area. The spatial effects could therefore be interpreted as representing the cumulative effect of unidentified or unmeasured additional covariates that may reflect impacts of environmental and socio-cultural factors.

4.6.6Assessment of Model Convergence

There are several methods to check for convergence. From this, we used the Autocorrelation Plot.

Auto-correlation plot produces lag-autocorrelations for the monitored parameters within each chain. In Markov chain auto-correlation analysis, it is necessary to identify lags in the series to calculate the long-run trends in correlation, and in particular whether they decrease with increasing lags. In Figure 4.8, the auto-correlations for all lags closer to zero as lag increases. So, the figure has evidence of convergence. Not all auto correlation plots are presented here; the rest plots can be found in appendices (see Appendix A).



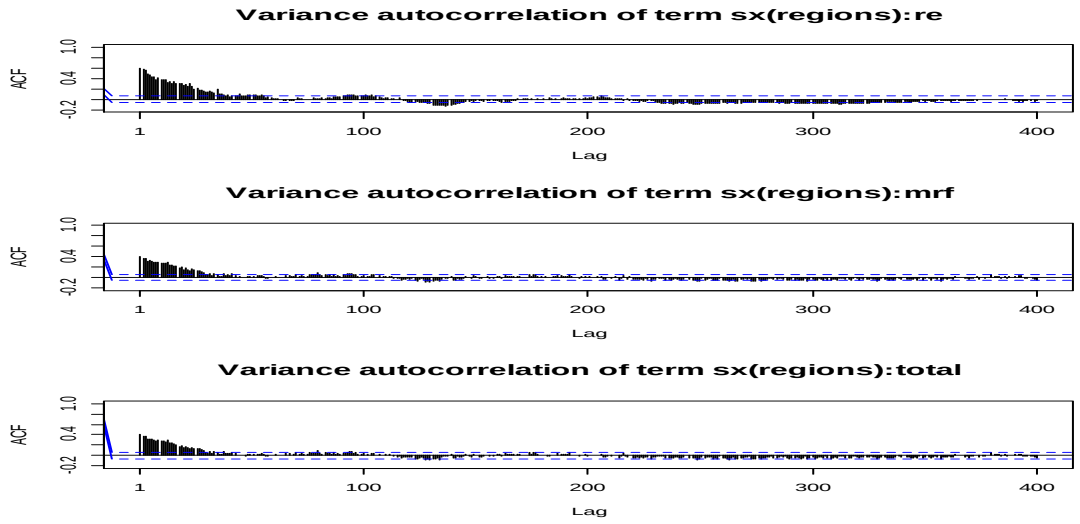


Figure 4.8: Convergence of autocorrelation plots for coefficients of child who had fever in the last two weeks, size of child at birth and regions

4.6.7 Discussion of the Results

The objective of this research was also to apply suitable statistical models that are used in assessing influential factors and geographical variation of childhood anaemia. One of the advantages of this research was to incorporates both the spatial variability and the nonlinear relationships between covariates and response variables by using the recent model (STAR) which provides valid and realistic statistical inference.

Bayesian Geo-additive models have highlighted specific socioeconomic and demographic risk factors associated with anaemia among children from Ethiopia. In particular, the results have shown that household wealth index, types of toilet facilities, size of child at birth, the child who had a fever in the two weeks before the survey, education levels of mothers and mother's anaemia status are strongly associated with anaemia among children under five in Ethiopia.

However, the relationship which was suggested in model1 and model2 analysis between anaemia and factors such as Mother's age, Mother's employment status, Husband/partner's education level, Number of children U5 years in the household, household size, mother occupation is not significant anymore after accounting for the effects of spatial effects in model 3. The spatial

component in model 3 according to Osei and Duker [112] helped to avoid underestimation model parameter standard errors which could result the insignificance of the covariates.

The descriptive results of this study indicate that the prevalence of childhood anaemia was found to be 57.4% with varied among regions of Ethiopia with the highest proportion of childhood anaemia was observed at Somali (83.3%) followed by Afar (74.7%) while the lowest percentage of the prevalence of childhood anaemia in the Benishangul region (44.2%) and the Amhara region (42.7%). The results of our study were relatively consistent with the 2016 Edhs report of the national prevalence of childhood anaemia in Ethiopia (57%) and the result was slightly higher than 2011 and 2005 EDHS reports, which were 44% and 54%, respectively [29]. The results of this study also higher than studies done in Kilte Awulaelo Woreda, Northern Ethiopia about 37.3% of children were anaemic[113].

The current study revealed that maternal anaemia status is an important risk factor of anaemia among children. Children of anaemic mothers are more likely to have anaemia. The possible reason might be related to the shared socio-economic status of the family which may affect both children and their mothers, and leading to anemia. This result was consistent with other findings of a study conducted in Uganda that children from anaemic mothers are associated with two folds increased risk of anaemia compared with those whose mothers are not anaemic [50]. Similar results were found in a study undertaken in Ethiopia on those children of mothers who had anaemia had 1.53 times greater odds of being at higher levels of anaemia compared to the children from mothers who had no anaemia [114].

The results of this study also found that type of toilet facility was an important factor for anemia levels of under five-year children. The risk was lower odds of being anaemic was suggested for children living in households with improved toilet facilities This might be because that poor sanitation is a known risk factor of the intestinal parasite hookworm which causes anaemia in infected children. This finding agreed with studies conducted in Kenya, Malawi, Tanzania and Uganda [47] showed that a significantly lower odds of anaemia was suggested for children living in households with improved toilet facilities (PIT latrine and flush toilet).

The results of this study also found that wealth index was significantly associated with anaemia in under five-year children. Children from the richer household had a significantly lower risk of anaemia compared to those from the poorer household. This is due to wealthier families are more likely to meet the expense of improved health services and afford suitable housing amenities which may prevent the increase of anaemia. Gebremedhin *et al*. [113] used multivariate logistic regression analyses to identify factors related to anaemia based on a cross-sectional study, higher prevalence of anaemia among the children living in a household with lower monthly income compared to those with higher income which is consistent with the results of this study. The results of this study are also consistent with what Stephen [51] noted that lower wealth index status was allied with the increase in the danger of development of anaemia among children under age 5 years in Ghana.

Our findings reveal that hat educational level of the mother was one of the most core factors associated with childhood anaemia. The results of this study found that the higher a mother's education, the lesser chance of their child being infected with anaemia. The reason was educated mothers are better informed on the knowledge and use of health facilities, appropriate infant and young child feeding practices, and better sanitation practice. This result is in line with a study conducted in some parts of Ethiopia [18, 48, 113], where the risk of having anaemia for children from mothers with no formal education much higher than those children from mothers with secondary and above education level. Also, this study is consistent with studies conducted in Kenya [47] revealed that the risk of anaemia was 1.5 times more in children whose mothers had no education as compared to children whose mothers had post-secondary education among children U5 in Kenya.

Furthermore, the results of this study also suggested that the size of a child at birth is significantly associated with the risk of anaemia in under-five children. This might be due to children with low birth weight have an increased exposure to infections and diseases. This is consistent with the result of a study conducted in Ghana [4], which identified that the prevalence was lower in children with a larger birth weight than those with a small birth weight.

Although every fever is not merely due to malaria, fever has been used as a proxy for malaria. Malaria is argued to be strongly related to anaemia and destroy red blood cells and reduce haemoglobin levels leading to anaemia[115]. The findings of this study also revealed that fever was found to be a significant predictor of childhood anaemia. The results of this study found that a child who had a fever in the two weeks before the survey is more likely to be anaemic than those who didn't have a fever. This is due to fever is a common symptom of acute and chronic diseases that have been associated with lower hemoglobin levels as well as anemia. This result was consistent with another finding of a study conducted by Ngwira and Kazembe [43] in Malawi, which revealed that fever is associated with higher levels of childhood anemia, severe anaemia compared to having no fever. Also, our finding was consistent with another study conducted in Rwanda that revealed that a child having a fever increases the likelihood of childhood anaemia [45].

With regards to age, younger children (6 months to 13 months) are more likely to be anaemic compared with those children aged between 14- and 59-month years old. Reasons for being anaemia is more prevalent within the ages of 6 to 13 months due to the reduction of haemoglobin that was available during birth and their antibodies could still be weak below the months of 14 years. The other factors which might expose younger children to anaemia are the lack of proper dietary food and lack of child health knowledge. The results are also in line with a study that suggested that the peak prevalence of anaemia occurs around 6 to 10 months of age after that prevalence of anaemia decline[14]. Similar results were found in studies looking at risk factors associated with other anaemia-related illnesses which suggest that odds of prevalence of childhood anaemia were increased at their young age[40].

The finding of this study also revealed that the body mass index of the mother is the main factor associated with childhood anaemia. The results indicate that a mother's body mass index of below around 19 increases the risk of childhood anaemia. This is due to mother who exhibits a very low BMI, indicating their poor nourishment, are likely to have poorly nourished children and mother's nutritional status which is assessed using BMI affects her ability to successfully carry, deliver, and care for her children. The result of this study was consistent with the finding of Roba *et al*., [49] conducted on prevalence of stunting and anaemia among children 6-23 months of age in two

agroecological zones of rural Ethiopia, show that as mothers' BMI increased the likelihood of the children being anaemic decreased.

Bayesian Geo-additive model allows the inclusion of generic types of covariates, such as nonlinear covariates and which incorporate both the spatially structured and spatially unstructured random effects. In our study, we have checked the spatial dependency in our data using Global Moran's I and Geary C statistic value before analysis of data using Bayesian Geo-additive models and it was shown that a significant positive spatial autocorrelation (regional variation) of childhood anemia in Ethiopia. The finding of this study showed that the eastern and north-eastern parts of the country mainly Somali and Afar regions are associated with a higher risk of anaemia whereas low anaemia rate areas were noted in the north western and northern parts of Ethiopia. This is because of living standards, socioeconomic status, cultural norms, and feeding habits among regions. Also, the regional nutritional disparities may explain the spatial heterogeneity of childhood anaemia in Ethiopia.

CHAPTER FIVE

5 Conclusion and Recommendation

5.1 Conclusion

The findings showed that more than half of the study population were found to be anaemic with 57.4%. The trend was increased to 57.4% from 44% in 2011 and 54% in 2005.

Based on the DIC model comparison, the most appropriate statistical model among other different models is the Geo-additive model. Bayesian Geo-additive models are good with their flexibility to allow the inclusion of generic types of covariates, such as nonlinear covariates and which incorporate both the spatially structured and spatially unstructured random effects were better fitting models of our data. Especially in targeted interventions, maps produced from our model can be of great importance to policy makers and the government for interventions and resource allocations. The inclusion of nonlinear effects of metrical covariate further improved the results.

The risk factors found to be significantly associated with childhood anemia were, household wealth index, types of toilet facilities, size of child at birth, a child who had a fever in the two weeks before the survey, education levels of mothers, and mother's anemia status. Odds of being anemic were higher for children from the poorest households, those living in households with not improved toilet facilities, whose mothers had no education and those their size at birth were smaller than average. The non-linear function showed observable relationships with anaemia status and continuous covariates. It is found that children are at high risk of anaemia during the young age of their life and when their mother's body mass index is in underweight mother.

The study also shows that there is a variation in the spatial distribution of childhood anaemia in the Ethiopia region and the prevalence was high in the eastern and north-eastern parts of the country, while low prevalence areas of childhood anaemia were noted in the north western and northern parts of Ethiopia.

5.2 Recommendation

Despite the limitation discussed below, we feel that this study fills the gap in knowledge of geographical variation, effects of metrical covariates, and fixed effects on childhood anaemia. Therefore, the following recommendations arise from the findings of this study. These are: -

- The concerned body, **government, and non-governments** may use anaemia prevalence map as a basis for interventions and resource allocations. Moreover, it can serve as a tool for getting funds to carry out more research in areas of high risk and over a large geographical area.
- The survey might be conducted across the district for further detailed analysis of the health facility factors associated with childhood anaemia.
- **Policy makers** should focus on activities that can improve household income to ensure sufficient food production.
- **The education sector** should focus on effective public education programs on child health, the dangers of anaemia, and appropriate feeding practices to target pregnant women.
- **The health sector** should focus on the improvement of clinical and health care infrastructure and services such as controlling infectious diseases which could also help in addressing the issue of childhood anaemia.
- **Future research** may consider other prior distributions for the spatial random effects, such as the two-dimensional P spline. Also, the Bayesian approach for modeling the different nature of anaemia response like ordinal response would be of interest for future research.

5.3 Limitation of the study

The following are some of the limitations of the study:

- Some limitations are related to the use of secondary data. Apart from the factors that were found to be associated with childhood anaemia in this study, there are a variety of other factors that may increase the risk of anaemia in children, but they were not included in this analysis due to lack of information on them, such variables include: like current status/

history of infectious disease particularly those which have a potential role in the risk of anaemia such as malaria and HIV**.**

- Another limitation of this research was the administration area level which ends at the regional level. It limits us not to know the hotspot area at district level from that region and compare the area levels such as district and regional levels, thus to focus on those areas in a more reliable manner and also advice policymakers to focus their interventions on such relevant areas.

Reference

1. WHO (2011) Haemoglobin on centrations for the diagnosis of anaemia and assessment of severity. Department of Nutrition for Health and Development (NHD) Geneva

2. Balarajan Y, Ramakrishnan U, Özaltin E, et al (2011) Anaemia in low-income and middle-income countries. 23–35

3. Osungbade, K. O., & Oladunjoye AO (2012) Anaemia in Developing Countries: Burden and Prospects of Prevention and Control:

4. Ewusie J, Ahiadeke C, Beyene, Hamid J (2014) Prevalence of anemia among under-5 children in the Ghanaian population: Estimates from the Ghana demographic and health survey. BMC Public Health;14(1): 1–9

5. Sung-Yong K, M. N R (2001) Health Knowledge and Consumer Use of Nutritional Labels: Journal of Nutritional Sciences

6. Dreyfuss ML, Stoltzfus RJ (2002) Guideline for the use of iron supplements to prevent and treat iron deficiency anaemia. Washington, DC, International Life sciences Institute Press (ILSI)

7. McLean E, Cogswell M, Egli I, et al (2009) Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005. Public Health Nutrition 12:

8. Stevens G, Finucane M, De-Regil L, et al (2013) Global, regional,and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995-2011.

9. UGC-CSSEIP C, Mysore U, Mysore I (2014) Prevalence of under-nutrition and anemia among under five rural children of south karnataka , india. Nitte Univ J Heal Sci. 4(4):24–7.

10. Mk U, Mh S, Mz H, Mm A, Mf B MU (2010) Prevalence of anaemia in children of 6 months to 59 months in Narayanganj , Bangladesh. J Dhaka Med Coll;19(2):126–30.

11. Gao W, Yan H, Dang S, Pei L (2013) Severity of Anemia among Children under 36 Months Old in Rural Western China. PLoS One;8(4):2–7

12. Leite M, Cardoso A, Ea C JC, et al (2013) Prevalence of anemia and associated factors among indigenous children in Brazil : results from the First National Survey of Indigenous People ' s Health and Nutrition. Nutr J. ;69(12):1–11

13. Kassebaum N, Jasrasaria R, Naghavi M, et al (2015) RED CELLS , IRON , AND ERYTHROPOIESIS A systematic analysis of global anemia burden from 1990 to 2010. 15–25

14. Roberts DJ, Matthews G, Snow RW, et al (2020) Investigating the spatial variation and risk factors of childhood anaemia in four sub-Saharan African countries. 1–10

15. Semedo R, Santos M, Baião M, et al (2014) Prevalence of Anaemia and associated factors among children below five years of age in Cape Verde, West Africa. Heal Popul NUTR;32(4):646–57.

16. Mahmoud H, Muddathir A, Osman S, et al (2014) Iron deficiency anemia among children under three years in Kassala, Eastern Sudan. J PUBLIC Heal;9(1):33–7.

17. Kuziga F, Adoke Y, Wanyenze RK (2017) Prevalence and factors associated with anaemia among children aged 6 to 59 months in Namutumba district, Uganda: A cross- sectional study. BMC Pediatrics 17:1–9. https://doi.org/10.1186/s12887-017-0782-3

18. Woldie H, Kebede Y, Tariku A (2015) Factors Associated with Anemia among Children Aged 6-23 Months Attending Growth Monitoring at Tsitsika Health Center, Wag-Himra Zone, Northeast Ethiopia. Journal of Nutrition and Metabolism 2015:. https://doi.org/10.1155/2015/928632

19. Aubuchon EN, SL G, Berhanu G, et al (2011) Hemoglobin, Growth, and Attention of Infants in Southern Ethiopia. Child Dev ,. 38–51

20. Dereje H, Asrat K, Magafu M, et al (2013) .Maternal Risk Factors for Childhood Anaemia in Ethiopia. Health. African Journal of Reproductive 17(3)

21. Gebrehaweria Gebremeskel M, Lemma Tirore L (2020) Factors Associated with Anemia Among Children 6–23 Months of Age in Ethiopia: A Multilevel Analysis of Data from the 2016 Ethiopia Demographic and Health Survey . Pediatric Health, Medicine and Therapeutics Volume 11:347–357. https://doi.org/10.2147/phmt.s258114

22. Central Statistical Agency (CSA) [Ethiopia] and ICF Macro, Ethiopia Demographic and Health Survey 2011. 2011, Central Statistical Agency and ICF Macro. Calverton, Maryland, USA: Addis Ababa, Ethiopia,.

23. Gebreegziabiher G, Etana B, Niggusie D (2015) Determinants of Anemia among Children Aged 6-59 Months Living in Kilte Awulaelo Woreda, Northern Ethiopia. Anemia 2014:. https://doi.org/10.1155/2014/245870

24. Mesfin F, Berhane Y, Worku A (2015) Anemia among Primary School Children in Eastern Ethiopia. 1–10. https://doi.org/10.1371/journal.pone.0123615

25. Desalegn A, Mossie A, Gedefaw L (2014) Nutritional Iron Deficiency Anemia Magnitude and Its Predictors among School Age Children, Southwest Ethiopia: A Community BasedCross-Sectional Study. PLoS ONE 9(12) e114059.doi:10.1371/journal.pone.0114059

26. Sudhakar M, Abdo N, Deribe K, Biadgilign S (2010) Malaria and Under Nutrition : A Community Based Study Among Under-Five Children at Risk of Malaria , SouthWest Ethiopia. PLoS One. 5(5):1–6.

27. Gebreweld A, Ali N, Ali R, Fisha T (2019) Prevalence of anemia and its associated factors among children under five years of age attending at Guguftu health center, South Wollo, Northeast Ethiopia. PLoS ONE 14:. https://doi.org/10.1371/journal.pone.0218961

28. Molla A, Egata G, Mesfin F, et al (2020) Prevalence of Anemia and Associated Factors among Infants and Young Children Aged 6-23 Months in Debre Berhan Town, North Shewa, Ethiopia. Journal of Nutrition and Metabolism 2020:. https://doi.org/10.1155/2020/2956129

29. Central Statistical Agency (CSA) [Ethiopia] And ICF (2017) 2016 Ethiopia Demographic and Health Survey Key Findings. Addis Ababa, Ethiopia, and Rockville, Maryland, USA. CSA and ICF

30. Muchie KF (2016) Determinants of severity levels of anemia among children aged 6-59 months in Ethiopia: Further analysis of the 2011 Ethiopian demographic and health survey. BMC Nutrition 2:1–8. https://doi.org/10.1186/s40795-016-0093-3

31. Kawo KN, Asfaw ZG, Yohannes N (2018) Multilevel Analysis of Determinants of Anemia Prevalence among Children Aged 6 – 59 Months in Ethiopia : Classical and Bayesian Approaches. 2018:

32. Asresie MB, Fekadu GA, Dagnew GW (2020) Determinants of Anemia among Children Aged 6 – 59 Months in Ethiopia : Further Analysis of the 2016 Ethiopian Demographic Health Survey. 2020:

33. Gebreegziabiher G, Etana B, Niggusie D (2015) Determinants of Anemia among Children Aged 6-59 Months Living in Kilte Awulaelo Woreda, Northern Ethiopia. Anemia 2014:. https://doi.org/10.1155/2014/245870

34. Nikoi EGA (2011) Child Nutritional Well-being in Ghana: An Analysis of Associated Individual, Household, and Contextual Health Indicators and Socioeconomic and Biophysical Environmental Variables

35. McLean E, Cogswell M, Egli I, et al (2009) Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005

36. Brabin B, Premji J (2001) Iron-Deficiency: Reexamining the Nature and Magnitude of the

Public Health Problem,An Analysis of Anaemia and Child Mortality. Journal of Nutrition and Metabolism

37. Danzhen You, Jin Rou New TW (2012) Estimates Developed by the UN Inter-agency Group for Child Mortality estimation

38. Stephano C (2011) Socio-Demographic determinants of anemia among children aged 6-59 months in mainland Tanzania. Statistical Science

39. Yip R, Ramakrishnan U (2002) Experiences and challenges in developing countries. Journal of Nutrition and Metabolism 132

40. Kawo KN, Asfaw ZG, Yohannes N (2018) Multilevel Analysis of Determinants of Anemia Prevalence among Children Multilevel Analysis of Determinants of Anemia Prevalence among Children Aged 6 – 59 Months in Ethiopia : Classical and Bayesian Approaches. https://doi.org/10.1155/2018/3087354

41. Blaise J (2014) Factors Affecting child anaemia in Cameroon

42. Konstantyner T, Oliveira T, Aguiar C, Taddei J (2012) Risk factors for anaemia among Brazillian infants from the 2006 national demographic health survey. BMC Nutrition

43. Ngwira A, Kazembe L (2016) Analysis of severity of childhood anemia in Malawi: a Bayesian ordered categories model. Open Access Medical Statistics 9. https://doi.org/10.2147/oams.s95159

44. Gaston RT, Ramroop S, Habyarimana F (2018) Determinants of factors associated with anaemia amaong children under five years in Lesotho. 3893–3902

45. Habyarimana F, Zewotir T, Ramroop S (2017) Structured additive quantile regression for assessing the determinants of childhood anemia in Rwanda. International Journal of Environmental Research and Public Health 14:1–15. https://doi.org/10.3390/ijerph14060652

46. Babatunde RO, Olagunju FI, Fakayode SB, Sola-Ojo FE (2011) Prevalence and determinants of anaemia among under-five children of farming households in kwara state, nigeria. Journal of Agricultural Science 3(3):173

47. Ngesa O, Mwambi H (2014) Prevalence and risk factors of anaemia among children aged between 6 months and 14 years in Kenya. PLoS ONE 9:1–10. https://doi.org/10.1371/journal.pone.0113756

48. Erkihun T, Tiruneh A (2018) Spatial Distribution and factors associated with childhood anemia in Ethiopia. Journal of Nutrition and Metabolism. https://doi.org/https://orcid.org 0000-0003-0677-7315.

49. Roba K., O'Connor T., Belachew T, O'Brien N. (2015) Prevalence of stunting and anaemia among children 6-23 months of age in two agro- ecological zones of Rural Ethiopia. Journal of Nutritional Sciences

50. Kandala, G.Ghilagaber(eds.) (2014) Advanced Techniques for Modelling Maternal and Child Health in Africa, The Springer Series on Demographic Methods and Population Analysis. https://doi.org/10.1007/978-94-007-678-2 9

51. Blam SN (2018) Spatially Adjusted Determinants of Malaria and Anaemia Morbidity among Children Under age 5 years in Ghana

52. Smith LC, Ruel MT, Ndiaye A (2005) Why is child malnutrition lower in urban than in rural areas? evidence from 36 developing countries. World Development. 33(8):

53. Rashid AG, Nikmazlan BM, Tefera B, et al (2017) PREDICTOR AND PREVALENCE OF ANEMIA,AMONG CHILDREN AGED 6 TO 59 MONTHS IN SHEBELLE ZONE, SOMALI REGION, EASTERN ETHIOPIA. African Journal of Reproductive

54. Afroja S, Kabir MR, Islam MA (2019) Analysis of determinants of severity levels of childhood anemia in Bangladesh using a proportional odds model. Clinical Epidemiology and Global Health 1–6. https://doi.org/10.1016/j.cegh.2019.06.009

55. Gayawan E, Arogundade ED, Adebayo SB (2014) Possible determinants and spatial patterns of anaemia among young children in nigeria: A bayesian semi-parametric modelling. International Health 6:35–45. https://doi.org/10.1093/inthealth/iht034

56. Chigozzie L, Ugwu J, Temesgen Z (2015) Spatial heterogeneity and determinants of childhood anaemia in Nigeria. 1–35

57. Kneib T, Lang S (2012) An R interface to BayesX. 2012. In: Statistische Inferenz

58. Adebayo S., Gayawan E, Ujuju C, Ankomah A (2013) Modelling Geographical Variations and Determinants of use of Modern Family Planning Methods Among Women of Reproductive age in Nigeria. Journal of Biosoc.sciences 45.:57–77

59. Kammann EE, Wand MP (2003) Geoadditive models. Journal of the Royal Statistical Society Series C: Applied Statistics 52:1–18. https://doi.org/10.1111/1467-9876.00385

60. Fahrmeir, L. and Echavarria LO (2006) Structured Additive Regression for Overdispersed and Zero-inflated Count Data. Applied Stochastic Models in Business and Industry, 22, 351-369

61. Hennerfeind A, Brezger A, Fahmeir L (2006) Geoadditive Survival Models. Journal of the American Statistical Association 10:65-75.

62. Chen Q., Ibrahim J. (2006) Semiparametric Models for Missing Covariate and Response in Data in Regression Models. Biometrics. 6:77-84.

63. Nkurunziza H, Gebhardt A, Pilz J (2011) Geoadditive Modelling of Malaria in Burundi. Biomed Central Journal

64. Wand H, Whitaker C, Ramjee G (2011) Geoadditive models to assess spatial variation of HIV infections among women in Local communities of Durban, South Africa. International Journal of Health Geographics 10:1–9. https://doi.org/10.1186/1476-072X-10-28

65. Musenge E, Vounatsou P, Collinson M, et al (2013) The contribution of spatial analysis to understanding HIV/TB mortality in children: a structural equation modelling approach.Global Health Action. 6:1–11

66. Best N, Richardson S, Thomson A (2005) A comparison of Bayesian spatial models for disease mapping. Statistical Methods in Medical Research. 14:35–59

67. Clements A., Lwambo N., Blair L, et al (2006) Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme. Spatial and Spatio-temporal Epidemiology

68. Adeyemi RA, Zewotir T, Ramroop S (2019) Joint spatial mapping of childhood anemia and malnutrition in sub-Saharan Africa : a cross-sectional study of small-scale geographical disparities. 19:2692–2712

69. Besag J, York J, Mollie A (1991) Bayesian image restoration with application in spatial ´statistics. Annals of the Institute of Statistical Mathematics 43 (1):1–20. https://doi.org/1007/BF00116466

70. Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. (Statistical Methodology). Journal of the Royal Statistical Society Series B: Applied Statistics 36 (2):192–236

71. Brezger A, Kneib T, Lang S (2002) BayesX-software for Bayesian inference based on Markov chain Monte Carlo simulation techniques. http://www.stat.uni-muenchen.de/la. Journal of Statistical Software

72. Goldstein H (1999) Multilevel statistical models. London: Institute of Education, Multilevel Models Project. http://www.soziologie.uni-halle.de/langer/multilevel/books/goldstein.pdf

73. Langford IH, Leyland AH, Rabash J, Goldstein H (1999) Multilevel modeling of the geographical distributions of diseases. J. ournal of Royal Statistical Society, Series A (Applied Statistics) 48:253–268

74. Borgoni R, Billardi FC (2003) Bayesian spatial analysis of demographic survey data: An application to contraceptive use at first sexual intercourse. Demographic Research, 83:61–92

75. Rabe-Heskesth S, Everitt B (2000) A handbook of statistical analysis using Stata (2nd ed.).Boca Raton: Chapman and Hall/CRC.

76. Bolstad WM, Manda SO (2001) Investigating child mortality in Malawi using family and community random effects: A Bayesian analysis. Journal of the American Statistical Association 96(453):12–19

77. Griffith A.D (2005) Comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States .International Journal of Health Geography.

78. Tobler R (1970) A computer movie simulating urban growth in the Detroit region.Economic geography supplements. 46:234–240

79. Anselin L, Bera K (1998) 'Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics', in: A. Ullah and D. Giles (eds), Handbook of Applied Economic Statistic,s eds. Aman Ullah and David E.A. Giles. New York: Marcel Dekker

80. Griffith A., Layne LJ (1999) Casebook for spatial statistical data analysis. A compilation of analysis of different thematic data sets. Oxford University Press, New York

81. Anselin L (1992) "SpaceStat Tutorial: A Workbook for Using SpaceStat in the analysis of Spatial Data." Typescript. University of Illinois at Urbana-Champaign. 8–67

82. Moran P (1948) The interpretation of Statistical maps. Journal of Royal Statistical Society, Series B, 10:243

83. Geary RC (1954) The contiguity ratio and statistical mapping. The incorporated statistician,. 5(3):115-146.

84. LeSage J (1998) “Bayesian Estimation of Spatial Autoregressive Models.” International Regional Science Review. 115–126

85. Cliff A, Ord J. (1981) Spatial processes - models and applications. (London: Pion).

86. Nelder JA, Wedderburn RW (1972) Generalized linear models. Journal of the Royal Statistical Society Series. 135:370–384

87. Carlin B, Xia H (1999) Assessing environmental justice using Bayesian hierarchical models: two case studies. Journal of Exposure Analysis & Environmental Epidemiology, 9:66-78.

88. Hastie, T. and Tibshirani R (1990) Generalized Additive Models. Chapman and Hall, London. Kammann, E. E. and Wand, M. P. 2003.

89. Khatab K (2007) Analysis of Childhood Diseases and Malnutrition in Developing Countries of Africa

90. Habyarimana F, Zewotir T, Ramroop S (2018) Spatial Distribution and Analysis of Risk Factors Associated with Anemia Among Women of Reproductive Age : Case of 2014 Rwanda. 425–437. https://doi.org/10.2174/18749445018110104

91. Khatab K, Adegboye O, Mohammed TI (2016) Social and demographic factors associated with morbidities in young children in Egypt: A Bayesian Geo-Additive Semi-Parametric multinomial model. PLoS ONE 11:. https://doi.org/10.1371/journal.pone.0159173

92. McCullagh P, Nedler J. (1989) Generalised Linear Models, Second Edition. Chapman and Hall/CRC Monographs on Statistics and Applied Probability

93. Fahrmeir L, Tutz G (2001) Multivariate Statistical Modelling Based on Generalized Linear Models. Springer, New York.

94. Fahrmeir L, Kneib T, Lang S (2004) Penalized structured additive regression: A Bayesian

perspective. Statistica Sinica,. 14:31–61

95. Besag J, York J, Molli´e A (1991) Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics. 43(1):1–20

96. Sudipto B, Gelfand Alan E, Carlin Bradley P (2004) Hierarchical Modeling and Analysis for Spatial Data. Chapman & Hall/CRC Press, London & New York NY.

97. Hastie T, Tibshirani R, Kammann EE, Wand MP (2003) Generalized Additive Models. Chapman and Hall, London. Statistische Inferenz

98. Besag J, Kooperberg C (1995) On conditional and intrinsic autoregressions.Biometrika. 82:73–746

99. Hobert JP, Casella G (1996) The efiect of improper priors on Gibbs sampling in hierarchical linear mixed models. Journal of the American Statistical Association 91:1461–1473

100. Lawson AB (2013) Bayesian disease mapping: hierarchical modeling in spatial epidemiology. CRC press, London & New York NY.

101. Fahrmeir L, Lang S (2001) Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors,2001a. Applied Statistics (JRSS C), 50:201–220

102. Lang S, Brezger A, Kneib T (2001) BayesX: Analysing Bayesian structured additive regression models. Journal of Statistical Software, 14, Heft 11

103. Fahrmeir L, Lang S (2001) Bayesian inference for generalized additive mixed models based on Markov random field priors. Journal of the Royal Statistical Society Series C: Applied Statistics 50:201–220. https://doi.org/10.1111/1467-9876.00229

104. Wecker W., Ansley C. (1983) The signal extraction approach to nonlinear regression and spline smoothing. J. Am. Stat. Assoc. 78, 81–89.

105. Rue H and, Held L (2005) Gaussian Markov Random Fields. Chapman & Hall / CRC

106. Fahrmeir L, Lang S (2001) Bayesian semiparametric regression analysis of multicategorical time-spacedata. (2001b).Annals of the Institute of Statistical Mathematics. 53:10–30

107. Lang, S. and Brezger A (2004) Bayesian P-splines. Journal of Computational and Graph ical Statistics 13, 183-212.

108. Kandala N., Lang S, Klasen S, Fahrmeir L (2001) Semiparametric Analysis of the Socio Demographic Determinants of Undernutrition in Two African Countries. Research in Official Statistics. Vol. 4 No.:81–217

109. Brezger A, Lang S (2005) Generalized structured additive regression based on Bayesian P-splines.Computational Statistics and Data Analysis. 50:967–991

110. Spiegelhalter D, Best N, Carlin B, LA V der (2002) Bayesian Measures of Models Complexity. Journal of the Royal Stat

111. Best N, Richardson S, Thomson A (2005) A comparison of Bayesian spatial models for disease mapping. Statistical methods in medical research. 14(1):35–59

112. Osei F., Duker A. (2012) Bayesian structured additive regression modeling of epidemic cholera data:application to cholera.Med Res Methodology. 12:118

113. Gebreegziabiher G, Etana B, Niggusie D (2014) determinant of anemia in children,northern Ethiopia. Hindawi Publishing Corporation 10:55–63

114. Gebrehaweria G, Lemma TT (2020) Factors Associated with Anemia Among Children 6–23 Months of Age in Ethiopia: A Multilevel Analysis

115. Menendez C, Fleming AF (2000) “Malaria-related anaemia.” Journal of Nutrition and Metabolism

Appendices

Table 4.6: Summary of Generalized linear models for binary logistic models

Fixed effects	Categories	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)		2.035	0.1955	10.410	< 2e-16
Mother Anaemia	Not Anaemic(ref)	0.000			
	Anaemic	0.6667	0.0561	11.877	< 2e-16
Wealth index	Poor(ref)	0.000			
	Middle	-0.4348	0.0789	-5.507	3.64e-08
	Rich	-0.349	0.0754	-4.628	3.70e-06
Mother age	15-24 (ref)	0.00			
	25-34	-0.0616	0.0728	-0.846	0.3974
	35-49	-0.285	0.0886	-3.218	0.0012
Toilet facility	Not Improv (ref)	0.0000			
	Improved	-0.2084	0.0625	-3.332	0.0008
Mother education level	No educ(ref)	0.00			
	Primary	-0.1089	0.0652	-1.670	0.0949
	Sec. and above	-0.3114	0.1013	-3.073	0.0021
Source of drinking water	Not improv.(ref)	0.000			
	Improved	0.1514	0.0637	2.378	0.0174
No of child U5 in the HH	0-3(ref)	0.00			
	4-6	0.534	0.1770	3.016	0.0026
Child size at birth	S. than aver.(ref)	0.00			
	Average	-0.1576	0.0643	-2.451	0.0142
	L. than averages	-0.1666	0.0691	-2.410	0.0159
Household size	1-4(ref)	0.000			
	5-8	0.1869	0.0675	2.767	0.0056
	9 and above	0.3375	0.0998	3.382	0.0007

Fever in last two weeks	No(ref)	0.000			
	Yes	0.2048	0.0755	2.711	0.0067
C. Breast Feeding status	No(ref)	0.000			
	Yes	-0.1264	0.0566	-2.233	0.0255
Mother Occupation	No(ref)	0.000			
	Yes	-0.1786	0.0583	-3.060	0.002

Table 4.7: Posterior mean estimates for the Bayesian Generalized linear model.

Fixed effects		**Estimate**	**Std. Error**	**Median**	**95%CI**
(Intercept)		2.0526	0.1959	2.0555	(1.6648, 2.4388)
Mother Anaemia	Not Anaemic(ref)	0.000			
	Anaemic	0.6680	0.0558	0.6683	(0.5606, 0.7808)
HH Wealth index	Poor(ref)	0.000			
	Middle	-0.4312	0.0790	-0.4321	(-0.5766 -,0.2720)
	Rich	-0.3480	0.0760	-0.3490	(-0.4950, -0.1905)
Mother age	15-24 (ref)	0.00			
	25-34	-0.0608	0.0696	-0.0620	(-0.1978, 0.0760)
	35-49	-0.2810	0.0875	-0.2810	(-0.4509, -0.1066)
Toilet facility	Not Improv. (ref)	0.0000			
	Improved	-0.2139	0.0632	-0.2136	(-0.3357, -0.0904)
Mother education	No educ(ref)	0.00			
	Primary	-0.1073	0.0644	-0.1064	(-0.2301 ,0.0198)
	Sec. and above	-0.3115	0.0981	-0.3112	(-0.5153, -0.1222)
Drinking water	Not improv.(ref)	0.000			
	Improved	0.1549	0.0643	0.1536	(0.0332, 0.2853)
No. children U5 HH	0-3(ref)	0.00			
	4-6	0.5417	0.1751	0.5440	(0.1978, 0.8666)
Child birth size	S. than aver(ref)	0.00			
	Average	-0.1600	0.0622	-0.1589	(-0.2845, -0.0302)
	L. than average	-0.1676	0.0670	-0.1694	(-0.2991, -0.0324)

Household size	1-4(ref)	0.000			
	5-8	0.1860	0.0680	0.1870	(0.0514, 0.3191)
	9 and above	0.3345	0.0958	0.3329	(0.1541, 0.5181)
Fever in last 2 weeks	No(ref)	0.000			
	Yes	0.2057	0.0741	0.2093	(0.0642,0.3566)
C. breast feeding sta.	No(ref)	0.000			
	Yes	-0.1284	0.0571	-0.1267	(-0.2422 -0.0185)
Mother Occupation	No(ref)	0.000			
	Yes	-0.1797	0.0575	-0.1801	(-0.2936, -0.0689)

Table 4.8: Posterior mean estimates for Bayesian semi-parametric model

Fixed effects	**Categories**	**Estimate**	**Std. Error**	**Median**	**95%CI**
Intercept		0.6697	0.1341	0.6702	(0.4027,0.9301)
Mother Anaemia	Not Anaemic(ref)	0.000			
	Anaemic	0.6728	0.0536	0.6740	(0.5597, 0.7761)
Wealth index	Poor(ref)	0.000			
	Middle	-0.4322	0.0808	-0.4327	(-0.5919, -0.2695)
	Rich	-0.3621	0.0757	-0.3614	(-0.5107, -0.2152)
Mother age	15-24 (ref)	0.00			
	25-34	-0.0694	0.0718	-0.0704	(-0.2127 ,0.0744)
	35-49	-0.2915	0.0862	-0.2913	(-0.4577, -0.1221)
Toilet facility	Not Improv. (ref)	0.0000			
	Improved	-0.1987	0.0673	-0.1964	(-0.3304, -0.0724)
Mother education level	No educ.(ref)	0.00			
	Primary	-0.1047	0.0669	-0.1038	(-0.2291,0.0332)
	Second.and above	-0.3480	0.1049	-0.3458	(-0.5458, -0.1363)
Sour. of drinking water	Not improv.(ref)	0.000			
	Improved	0.1318	0.0621	0.1329	(0.0077, 0.2488)

No. child U5 in the HH	0-3(ref)	0.00			
	4-6	0.5651	0.1881	0.5669	(0.1820 ,0.9181)
Size of child at birth	S. than aver. (ref)	0.00			
	Average	-0.1597	0.0624	-0.1601	(-0.2873, -0.0408)
	L. than averages	-0.1700	0.0686	-0.1662	(-0.3040, -0.0316)
Household size	1-4(ref)	0.000			
	5-8	0.1868	0.0656	0.1878	(0.0562 ,0.3160)
	9 and above	0.3195	0.1023	0.3202	(0.1122, 0.5129)
Fever in last two weeks	No(ref)	0.000			
	Yes	0.2067	0.0766	0.2073	(0.0622, 0.3519)
Currently breast	No (ref)	0.000			
Feeding status	Yes	-0.1353	0.0565	-0.1362	(-0.2421, -0.0195)
Mother Occupation	No (ref)	0.000			
	Yes	-0.1725	0.0595	-0.1726	(-0.2979 -0.0550)
Smooth terms variances of non-linear Metrical covariates					
		Estimate	Sd	Median	95% CI
sx(Mother BMI)		0.0007	0.0006	0.0005	(0.0002,0.0021)
sx(Current age of child in month)		0.0010	0.0008	0.0008	(0.0002,0.0031)

Table 4.9: Association between childhood anaemia and socio-economic, demographic and environmental factors.

Variable Names	**Pearson's Chi-squared**	**p-value**
Region	364.240[a]	.000
Residence	20.815[a]	.000
Gender of Child	.619[a]	.432
Mother occupation	6.823[a]	.009
Mother Education level	14.672[a]	.001
Currently breastfeeding	53.917[a]	.000
Taking iron pills, sprinkles/syrup	.698[a]	.403

Fever in last two weeks	33.304[a]	.000
Had diarrhea recently	8.472[a]	.004
cough in last two weeks	4.696[a]	.030
No. of child U5in HH	14.134[a]	.000
Birth size of child	21.827[a]	.000
Wealth Index	84.284[a]	.000
Mother age	42.936[a]	.000
House Hold size	10.219[a]	.006
Husband occupation	10.363[a]	.001
Mother Anaemia level	182.683[a]	.000
Types of toilet facility	30.457[a]	.000
Sour. of drinking water	8.349[a]	.004
Husb. education level	9.563[a]	.023

Appendix A

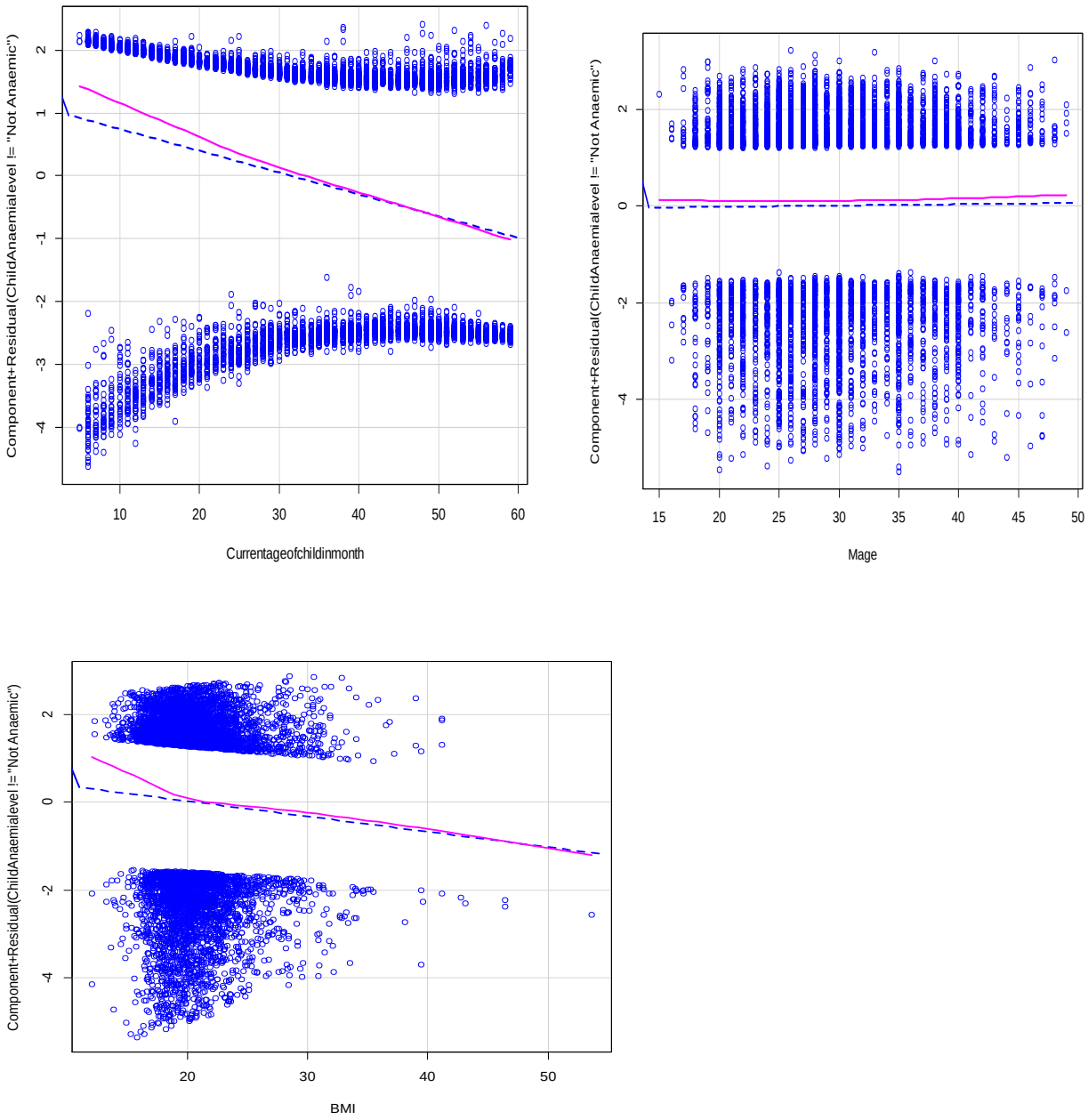


Figure 4.9: Component-plus-residual plot (partial-residual plots) for child age in month, Mothe age, and Mother Body Mass Index

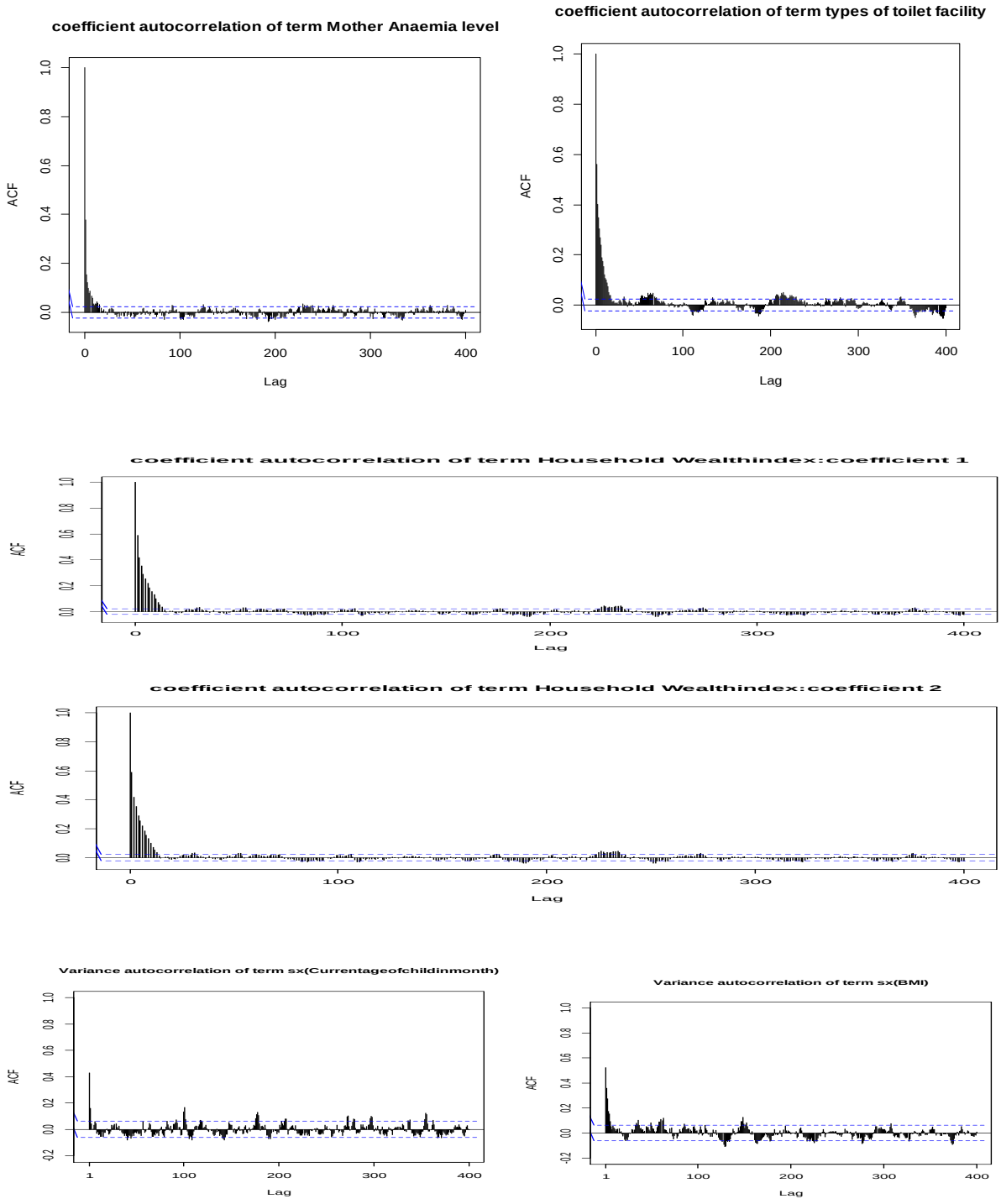


Figure 4.10: Autocorrelation plot of convergence check.