Flexible Statistical Modeling of Childhood Malnutrition in Malawi

Mzwakhe Magagula

, 2019

Flexible Statistical Modeling of Childhood Malnutrition in Malawi

by

Mzwakhe Magagula

A thesis submitted to the
University of KwaZulu-Natal
in fulfilment of the requirements for the degree
of
MASTER OF SCIENCE
in
STATISTICS

Thesis Supervisor: Prof. Shaun Ramroop

UNIVERSITY OF KWAZULU-NATAL
SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
PIETERMARITZBURG CAMPUS, SOUTH AFRICA

Declaration - Plagiarism

I, Mzwakhe Magagula, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowlegded as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then
   (a) their words have been re-written but the general information attributed to them has been referenced, or
   (b) where their exact words have been used, then their writing has been placed in italics and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference sections.

[redacted] 30 / 07 / 2020

Mzwakhe Magagula (Student) Date

[redacted] 30/07/2020

Prof. Shaun Ramroop (Supervisor) Date

Disclaimer

This document describes work undertaken as a Masters programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institution.

Acknowledgements

First and foremost, I would like to express my great appreciation to my supervisor. Prof. Shaun Ramroop. He suggested the topic for this thesis and provided invaluable advice and constant encouragement throughout the course of my research. Without his support and guidance, this work would not have been completed. Shaun is more than an academic supervisor to me and I shall always appreciate his guidance, which led me to the wonderful world of statistics.
I am also grateful to Dr Faustin Habyarimana, for his help and guidance, sharp comments and warm encouragement that accompanied me throughout the research study process.

I also wish to thank my family for all the support; without their love towards me, it would never have been possible for me to go on. Last, but not least, I would also like to extend my gratitude to my friends for helping me out in many ways and for making my life in PMB colourful and enjoyable.

To God I give all the glory for this wonderful work, with Him every step of the process was made possible. I cannot leave out my family in Christ, both from home and in PMB (Church on Campus), for the spiritual support and for helping me to be kept under the hand of God, I thank you very much.

Contents

List of Figures

List of Tables

Abstract

Childhood malnutrition is one of the most significant health problems affecting public health departments, mainly in developing countries. The development of proper assessment of malnutrition is one of the challenges faced by policy makers in many countries across the globe. Therefore, the current study was undertaken with the primary objective of assessing and determining all possible determinants of malnutrition in Malawi, using the Demographic and Health Survey (DHS) data 2015/16. Different types of statistical models were adopted to allow variety in methodology and to find the most accurate results among the models used. As a point of departure, the study utilized Generalized Linear Models (GLM) to account for the ordering of the outcome variable (severe, moderate and nourished). Furthermore, we noticed that it would be substantial to extend the ordinal logistic regression to include random effects and therefore to consider the variability between the primary sampling units or villages. Furthermore, we adopted a class of models that allows flexible functional dependence of an outcome variable on covariates by using nonparametric regression. Hence, the use of the generalized additive mixed model (GAMM), which relaxes the assumption of normality and linearity inherent in linear regressions.

Analyses of childhood stunting have mainly used mean regression, yet modelling using quantile regression is more appropriate than using mean regression in that the former provides flexibility to study the impact of predictors on different desired quantiles of the response distribution, whereas the latter allows only studying the impact of predictors on the mean of the response variable. Therefore, quantile regression models were adopted for the provision of a complete picture of the relationship between the outcome variable (stunting) and the predictor variables on different desired quantiles of the response distribution. This study fitted a Bayesian additive quantile regression model with structural spatial effects for childhood stunting in Malawi, using 2015/16 DHS data. Inference was fully Bayesian, using the new integrated nested Laplace approximation (INLA), purely because of its much faster computation as compared to Markov chain Monte Carlo (MCMC). Further-

more, different types of quantile regression models were fitted and compared according to each Deviance Information Criteria (DIC) for determination of the best model among them.

Each of these models has inherent strengths and weaknesses. The choice of one depends on what the research is trying to accomplish and the type of data one has. In this study, we combined the results from different models, mainly from our quantile regression models. The significant determinants of childhood stunting in Malawi were found to be the age of the child, the education level of parents (mother and father), the family's place of residence, gender of the child, incidence of recent fever, incidence of recent diarrhoea, multiple births, mother's age at the birth, body mass index of the mother, wealth index of the family, source of drinking water and districts. Furthermore, from the spatial quantile regression model, a map was generated showing the distribution of malnutrition in a district level of Malawi. This map gave us an overview on how stunting is distributed in Malawi and from the map we were able to visualize and assess affected districts.

Background

Proper nutrition allows children to grow, develop, learn, play, participate and contribute, while malnutrition robs children of their future and leaves young lives hanging in the balance [1]. In 2017, the WHO, UNICEF and the World Bank group reported that nearly half of all deaths in children under five years of age were attributable to undernutrition, translating to a loss of about 3 million young lives per year. Undernutrition puts children at greater risk of dying from common infections, increases the frequency and severity of such infection, and contributes to delayed recovery. It was also reported that undernutrition in the first 1,000 days of a child's life can lead to stunted growth, which is associated with impaired cognitive ability and subsequent reduced school and work performance [1].

Though childhood malnutrition remains a global concern, it is observed to be more dominant in developing countries. For example, in 2016, more than half of all stunted children under five years old lived in Asia and more than one third lived in Africa; almost half of all overweight children under five lived in Asia and one quarter lived in Africa; and more than two thirds of all wasted children under five lived in Asia with more than one quarter living in Africa. Basically, Africa and Asia have the greatest share of all forms of malnutrition [2].

Undernutrition is commonly assessed through the measurement of a child's anthropometry (height, weight), as well as through screening for biochemical and clinical markers [3]. Wasting, stunting and underweight are expressions of under nutrition and the anthropometric indicators for the assessment of a child's nutritional status. Stunting is seen when a child is too short for his or her age. It is the failure to grow both physically and cognitively and is the result of chronic or recurrent malnutrition. The devastating effects of stunting can last a lifetime [2]. Overweight refers to when a child is too heavy for his or her height. This form of malnutrition results from expending too few calories for the amount consumed from food and drinks and increases the risk of developing noncommunicable diseases later in life. Wasting refers to a when a child is too thin for his or her height [2]. Also known as or

acute malnutrition, it is the result of recent rapid weight loss or the failure to gain weight. A child who is moderately or severely wasted has an increased risk of death, but treatment is possible [3].

Although an increased risk of dying is the most serious and alarming consequence of malnutrition, recent reports have revealed that malnutrition has other significant health and economic effects that include increased risk of illness and limited cognitive development, which result in lower levels of educational attainment [4]. These are critical findings because of the serious implications as far as development of a country. It follows that if a country has a high rate of malnutrition, the number of educated people will decrease which will imply lack of knowledge and skills in that country. Now when there is a lack of skills in the country, the production will also decrease, and when there is insufficient production, the country's GDP will be low, with far reaching effects, like weakening the currency of the country when compared to other countries. Therefore, it is clear that the effects of childhood malnutrition can be regarded as long-term in nature, that will later on determine the state of a country economically.

Chapter 1

Introduction

Sub-Saharan Africa has one of the highest levels of child malnutrition globally. Therefore, a critical look at the distribution of malnutrition within its subregions is required to determine the factors that escalate the worst nutritional status in these areas. This study uses regression models to identify the determinants of childhood (under-five years) malnutrition in Malawi.

Determinants of child malnutrition remain an interest of many researchers. In 2017, according to the World Health Statistics data visualizations dashboard (2018), the regional average of prevalence of stunting in the regions of Africa was found to be 33.6% above the global average, which is 22.2%. Moreover, in 2017, globally, there were 151 million children under five years of age who were stunted, 51 million who were wasted and 38 million overweight children [5]. Hence, we can observe that the prevalence of stunting remains the most problematic form of malnutrition globally.

A number of studies, globally, have attempted to establish factors that determine child malnutrition in order to directly improve the situation. For example, the World Health Organization (WHO) has also implemented a plan of action: the United Nations Decade of Action on Nutrition. In addition, the WHO's member states have endorsed global targets for improving maternal, infant and young child nutrition and are committed to monitoring progress. These targets are vital for identifying priority areas for action and for catalysing global change as stipulated in the global nutrition targets 2025. For instance, increasing exclusive breastfeeding in the first six months resulted in up to 50% and 40% reduction in the number of stunted under-five children [6]. These are some areas, according to WHO, which need to be improved in order to achieve the target.

In response to the call of the United Nations Decade of Action on Nutrition, there

is positive and impressive global implementation to fight the issue of malnutrition, one of a number such cases is Norway becoming the first country to establish an action network as part of the United Nations (UN). Decade of Action on Nutrition 2016-2025 on 6 June 2017. Another is tThe Global Action Network on Sustain- able Food from the Ocean for Food Security and Nutrition callings for higher priority to be given to fisheries and aquaculture in an efforts to improve global food security [7]. The Second International Conference on Nutrition (ICN2) adopted the ICN2 Rome Declaration on Nutrition and its Framework for Action, and governments committed to eradicate hunger and prevent all forms of malnutrition worldwide. Among other action, member states committed to enhance sustainable food systems by developing coherent public policies from production to consumption across relevant sectors to provide year-round access to food that meets people's nutrition needs, promote safe and diversified healthy diets [8].
The governments of Italy, Japan, the Russian Federation and the United Kingdom and Northern Ireland co-sponsored a special event on strengthening national commitment to end malnutrition in all its forms. This event was supported by the Food and Agriculture Organization of the United Nations and the World Health Organization, and took place on 20 September 2016 at the UN Headquarters [9]. Many other reports presented by the WHO show how fighting malnutrition is being fought around the world and researchers are also working to identify all the key factors.

The African region has the largest population (33.6%) of stunted children and a minimum of 20.0% prevalence of wasting in all the regions, according to the WHO's (2018) latest report. As a result, a number of studies on the maturational status of children under-five years have been carried out across the regions of Africa, using different methods. The purpose of these studies is to determine the risk factors of childhood malnutrition on the continent. One researcher, Habyarimana (2017), used the proportional odds model with a complex sampling design to identify key determinants of malnutrition in under-five children in Rwanda [10]. Another recent study conducted by Taluder (2017) attempted to uncover the associated factors of malnutrition among under-five Bangladeshi children by using BDHS data with an application of the PO model [11]. A number of researchers have already revealed that factors such as mother's education level, father's education level, wealth index, mother's BMI, place of residence, division, antenatal care service during pregnancy, and birth interval are common causes of a poor nutritional status among children under the age of five across the African continent [10, 11, 12, 13].

A descriptive and econometric analysis done by Kabubo-Mariara (2008) in Kenya, augmented by policy simulations, was employed to investigate the impact of child,

parental, household and community characteristics on children's height and on the probability of stunting [14]. Key findings from this study were that boys suffer more from malnutrition than girls, and children of multiple births are more likely to be malnourished than singletons. Other studies have discovered that even the age of the child and childhood illness are significant variables affecting child nutritional status in Africa [12, 4].

The prevalence of stunting in under-five children is very high in Malawi (37.1% on average), which is above the regional average in Africa (33.6%), according to the joint child malnutrition estimates by the United Nations Children's Fund, the World Health Organization, and the World Bank Group [5]. According to the UNICEF global site (2018), in Malawi, 4% of children, especially those below five years of age, suffer from acute malnutrition, and more than half of Malawian children suffer from chronic malnutrition, resulting in stunting (being too short for one's age). These figures therefore imply that Malawi is one of the countries with a high malnutrition incidence in eastern and southern Africa. There are several causes that have been attributed to this such as national and household food insecurity, heavy workloads and poor nutrition of mothers, frequent infections, poor eating habits, and HIV infections leading to repeated illness [4]. Malnutrition is devastating and the single biggest contributor to child death. In Malawi, unfortunately there has been only a small change in children's nutritional status (stunting) since 1992 (48.7%) and stunting rates remain unacceptably high thus far (37.1%), according to the WHO.

In Malawi, similar to what other countries are doing around the world, different stakeholders are attempting to mitigate the impact of malnutrition. During the period 2012–2018, UNICEF focused on prevention of faltering and stunting, both of which are major stumbling blocks to children's survival and development. Focus was directed towards adequate nutrition for pregnant women and newborn babies during their first 24 months of life. Some strategies to prevent stunting in the country include supporting the government to develop the Nutrition Act and supporting partners in 15 districts to achieve household and behavioural change for maternal nutrition, and infant and young child feeding practices [15].All these efforts seem to be promising in making sure that the average level of stunting is decreasing in the country.

In addition to these strategies aimed at preventing the disease, a significant approach is to know the exact factors that are escalating the occurrence of the disease. A study by Chirwa (2008) was conducted in Malawi, using a multivariate analysis to investigate factors that determine child malnutrition. Using the three anthropometric

measures of malnutrition, WAZ for underweight, HAZ for stunting and WHZ for wasting, Chirwa (2008) discovered that stunting is the greater of the malnutrition problems. The age of the child, sex of the child, regular sickness, source of water, education level of household head, household land size and ability to produce own food were all factors associated with child nutritional status derived from multivariate analysis [4].

Many different statistical methods can be used to uncover the factors of malnutrition. Some researchers have considered the response variable as a binary (nourished and undernourished) variable, hence the binary logistic regression model was applied. However, the child nutritional status is ordered in three main categories (severely malnourished, moderately malnourished and nourished). In taking this ordinal nature of childhood malnutrition status into consideration of this ordinal nature of childhood malnutrition status, the use of ordinal logistic regression models are used to discover the factors that are associated with childhood mal- nutrition is used. The use of Ordinal logistic regression is preferred because ordinary logistic regression, unlike polytomous regression, takes into account any inherent ordering of the levels in the disease or outcome variables, thus making fuller use of the ordinal information [16]. Alternatively, a multinomial logistic regression can be imposed by ignoring the order of information that is present in the response variable [12, 13].

The current study sets out to identify risk factors associated with the prevalence of childhood malnutrition in Malawi by developing various statistical models namely; proportional odds model (POM), mainly to deal with ordered categorical nature of the data; generalized additive mixed models (GAMM), mainly to allow for flexible functional dependence of an outcome variable on covariates by using nonparametric regression and to also account for variability between primary sampling units; quantile regression models, this model provides the flexibility to analyse the impact of predictor variables on the different quantiles of interest of childhood anthropometric index (stunting) instead of the mean distribution. In addition, this model allows the possible nonlinearity effects of continuous covariates to be accounted for, the possible structured spatial effects on childhood stunting to be captured and possible heterogeneity among the variables to be incorporated. In addition, a spatial quantile additive model is computed and produced a map, showing the distribution of childhood stunting on districts level in Malawi, using the new integrated nested Laplace approximations (R-INLA) approach to statistical inference for latent Gaussian models.

In the beginning of every chapter, we begin by giving a brief introduction of the

model to be fitted and advantages thereof. We seek to test different statistical methods which in turn can help to propose suitable techniques and to appropriately fit future work from the demographic and health survey data (DHS) and any other related data. This study will use only one anthropometric measure of malnutrition, HAZ for stunting, all the factors considered in other studies are tested, then other variables which have never been considered before are added. The prime objective of this study is to use the results to recommend a need to re-examine and rewrite the existing policies on child malnutrition in Malawi.

Chapter 2

Exploratory data analyses (EDA)

Exploratory data analysis (EDA) is used to understand the basic structure and statistics of the data. It is where a bird's eye view is taken of the data and an attempt is made to gain some sense of it. For many researchers, it is often the first step in data analysis implemented before any formal statistical techniques are applied, although, in the current study, we have applied some specific statistical techniques (Chi-Square test) to check the association of our response variables with the selected predictor variables. EDA is a complement to inferential statistics, which tends to be fairly rigid with rules and formulas. Alternatively, EDA involves the analyst trying to 'get a feel' for the dataset, often using their own judgment to determine what the most important elements in the dataset are. This study has used EDA for the following purposes: to check for missing data and the total number of observed variables, to gain maximum insight into the dataset and its underlying structure and to check assumptions associated with any model fitting or hypothesis test. This section will provide a clear view with more details of the dataset worked with, before going further to fit any model of interest.

2.1 Data and Definition of Variables

The data used in this study are from the nationwide Malawi Demographic Health survey 2015 (MWDHS-2015). It includes an anthropometric component where all children aged under five listed in the Household Questionnaire were weighed and measured. The study utilized 5 092 children, whose complete and plausible anthropometric data were available.

The dependent variable in this study is the anthropometric measure of height-for-age z-score as an indicator of chronic malnutrition known as stunting. Childhood stunting is the best overall indicator of children's well-being [17]. Moreover, stunt-

ing is the most prevalent form of child malnutrition, with an estimated 161 million children worldwide in 2013 falling below 2 SD from the height-for-age World Health Organization Child Growth Standards median.

In the present study we considered only the height-for-age anthropometric index instead of weight-for-height (wasting) and weight-for-age (underweight) to determine the child nutrition status. We calculated the Z-scores of height-for-age for each child to access the association between nutritional status and other selected variables. The child nutrition status was considered as a category with three levels: severely malnourished (Z-score $<$ -3.0), moderate malnourished (-3.0 $\leq$ Z-score $<$ 2.0) and nourished (Z-score $\geq$ -2.0). Hence we recognized that our nutritional status is an ordinal response variable grouped from continuous variables.

The explanatory variables considered are child characteristics and recent child illness, household characteristics, and community characteristics. The child characteristics include the age of the child, gender of the child, birth order and whether the child is a twin or not. The sex of the child is recorded by a dummy variable equal to 1 for a male and 2 for a female child. If there is gender inequality with respect to the care of children, we expect female children to be better nourished than male children, if female children are favoured and vice-versa. In other studies, female children were found to have a better nutrition status as compared to male children [4]. Then we also took into consideration the characteristic of whether a child is from a multiple birth or not. It was recorded by a dummy variable equal to 1 for a child from a multiple birth and zero for a single born child.
The age of the child has an upper boundary of five years, which was measured in months, since stunting is the failure to grow optimally and is first detected in children who are considered short for their age group when they are two years old [17]. Therefore, we expect the children who are two years or more to have worse nutrition status compare to those who are younger than 2 years. The birth order of the child and recent illness of the child child are also characteristics that we are to access against the nutrition of status of the child. The child's illness were the recent incident of diarrhea and whether the child had fever in the last two weeks or not. both were captured with dummy variable equal to zero for no incident of diarrhea or fever, and dummy variable equal to 1 represent a yes there was incident of diarrhea or fever recently.

The household characteristics include wealth status of the household, mother's BMI, mother's age at birth of respective child, mother's working status, mother's highest level of education ,father's highest level of education and father's main occupation

grouped. The wealth index of household economic status was constructed in the MWDHS-2015 report by using the information on household ownership of assets and dwelling characteristics. The wealth index was recorded with dummies representing the three categories, of poor, middle and rich. Mother's BMI was captured in two categories, Thin (BMI$\leq$18.5) and Normal (BMI$>$18.5). The expectation was that an underweight mother would result in a worse nutrition status for the child, since she would lack the fat needed to produce adequate milk for her child.

The age of the female parent at the birth of the respective child was also categorized in five different levels. There is evidence elsewhere that children born to mothers below 18 and above 34 years of age are more likely to be malnourished when compared with the children born to mothers aged 18 to 34 [18]. The current working status of the mother was captured by a dummy variable equal to one for the currently working mother and 0 for a non-working mother. Also, the highest level of education of a mother was measured with dummies representing four categories: no education, primary, secondary and higher education. The highest level of education for a father was captured in the same manner. Education affects care-giving practices through the ability to process information, to acquire skills and to model behaviour. It can be hypothesized, therefore, that educated parents are associated with a high nutritional status of the child, as they are better able to use healthcare facilities and ensure a high standard of environmental sanitation [4].
Five dummy variables of a father's occupation measure different types of occupation. This independent variable is thus categorized as: not working, professional, business, agriculture and other occupational sectors. The prevalence of stunting was found to be significantly lowest among the children of fathers who were service holders in a study of predictors of chronic child malnutrition in Bangladesh by Das(2011). One would expect children of fathers who hold professional positions (such as managerial positions) to have better nutritional status, since in professional occupations, parents usually get maternity leave to care for their newborn baby, which may help them to ensure an early healthy lifestyle for the child.

The community variables that are used in this study include water sources and type of residence of the household. Source of water is captured with dummy variables for piped water, well water and other sources (such as rain, tank, etc.). The type of residence household is captured by dummy variables equal to one for urban and two for rural areas. The expectation is that children from urban areas would be associated with better nutritional status, since the availability of health and educational facilities is greater in urban areas compared with rural areas. The duration of breastfeeding was also added to the independent variables to enable an assessment of the

effect of breastfeeding. Breastfeeding is captured by dummies that are categorized in three groups: ever breastfed, then stopped, never breastfed and still breastfeeding. Table 2.1 below presents all the selected variables along with their levels of coding.

Table 2.1: Variables and their level of coding from 2015-16 Malawi DHS-dataset

Variable	Description	level of coding	Type
Nutrition status	Level of stunting	1 = Severe 2 = Moderate 3 = Nourished	Ordinal response variable
Child's age	Age of the child in months	1 = 0 - 11 months 2 = 7 - 11 months 3 = 24 -35 months 4 = 36 -47 months 6 = 48 -59 months	Categorical covariate
Sex	Gender of the child	1 = female 2 = male	Categorical covariates
Birth type	Child is of single/multiple birth	1 = singleton 2 = multiple	Categorical covariates
Birth order	The number of birth line for child	1 = 1^{st} born 2 = $2^{nd} - 3^{rd}$ born 3 = $4^{th} - 5^{th}$ born $4 = 6^{+}$ born	Categorical covariates
Diarrhea	Child suffered from diarrhea in the last two weeks	1 = Yes 2 = No	Categorical covariate
Flue	Child suffered from flue in the last two weeks	1 = Yes 2 = No	Categorical covariate
Mother's education	Highest level of education for mother	1 = No education 2 = Primary 3 = Secondary 4 = Higher	Categorical covariate
Mother's BMI	Body mass index of mother	1 = Thin 2 = Normal	Continuous covariate
Mother's age (years)	Age of mother at birth of child	1 = 0 - 18 yrs 2 = 19 - 23 yrs 3 = 24 - 29 yrs 4 = 30 - 34 yrs 5 = 35^{+} yrs	Categorical covariate
Father's education	Highest level of education for father	1 = No education 2 = Primary 3 = Secondary 4 = Higher	Categorical covariate

Continued on the next page

Table 2.1 – *Continued from the previous page*

Variable	Description	level of coding	Type
Father's occupation	Type of work for the father	1 = Not working 2 = Professional 3 = Business 4 = Agriculture 5 = Other	Categorical covariate
Wealth index	Household wealth index	1 = Poor 2 = Middle 3 = Rich	Categorical covariate
Type of residence	Type of residence	1 = Rural 2 = Urban	Categorical covariate
Source of water	Source of water for household	1 = Piped water 2 = Well water 3 = Rainwater, tank and other	Categorical response variable
Duration of breastfeeding	Consistency in breastfeeding of the child	1 = Ever, then stopped 2 = Never breastfed 3 = Still breastfeeding	Categorical response variable

2.2 Descriptive statistics

The proportion of stunted children was 28.1% with 7.2% severely stunted in 2015. The prevalence of stunting according to selected independent variables are shown in Table 2.2. The proportion of severely stunted and moderately stunted children is higher among the children aged 36–47 months (10.8% and 23%). Moreover, the level of stunting is also lowest for children aged less than, or equal to, 6 months. According to the WHO, stunting is often picked up when children are two years or older. Table 2.2 also reveals to us that the level of stunting is significantly associated with characteristic of multiple/twin born child ($\chi^2 = 38.041$ and *p-value* = 0.000). The prevalence of stunting was higher for multiple/twin born children (48.4%) compared to single born children (27.5%). This could be explained by low birth weight, inadequate breastfeeding and competition for nutritional intake, which tends to afflict children of multiple births more than singletons.

The type of residence had a significant association with the level of stunting ($\chi^2 = 37.05$, *p-value* = 0.000). Prevalence of stunting was higher among rural (29.9%) than

among urban children (19.4%). As expected, living in an urban area has a significantly positive effect on avoiding the worst nutritional status of children compared to rural areas. This may be attributable to access to health facilities and information from local doctors in urban areas. Moreover, the increased access to higher paying jobs and opportunities to earn a higher income to support a family may be the reason for this urban bias. The prevalence of malnutrition is also observed to decrease with the increase in the level of education of parents. For mothers and fathers with higher education levels the proportion of stunting was found to be lowest (8.8% and 10.5% respectively), while the proportion of stunting was highest for mothers and fathers with no education at all (35.2% and 31.1% respectively). Mother's and father's educational attainment has been shown elsewhere as evidence of a high inverse association with the level of stunting [13].

There was no direct significant association between a mother's current working status and level of stunting (χ^2 = 3.3, *p-value* = 0.193), however, the father's occupation has a significant association with the prevalence of childhood nutritional status (χ^2 = 25.0, *p-value* = 0.002) (Table 2.2). The prevalence of stunting was significantly lowest among the children of fathers who hold professional occupations (20.4%). As hypothesized, fathers who are working in professional occupations are expected to be associated with good nutritional status in their children. This is because their jobs are not as demanding as other occupations, so they have more time to ensure the wellbeing of their families. Both severe and moderate stunting was higher for children of fathers working in the agricultural sector and fathers not working (unemployed).

Most studies on child health suggest that children born of young mothers, teenage mothers more so, are more likely to suffer ill health than children born of adult women [13, 14]. The results in Table 2.2 also confirm that children of mothers who are less than 24 years of age are associated with a poor nutritional status. The prevalence of stunting is observed to be lower for children of mothers who are aged 24–29 and 30–34 (25.8% and 25.7% respectively) compared to those aged less than 18, 19–24 and 35+ (27.5%, 31.3% and 30.2% respectively). The results from Table 2.1 also show that a mother's BMI is associated with malnutrition. Well-nourished mothers with a BMI of 18.5 or more were associated with a lower prevalence of stunting. The results are consistent with the study from Bangladesh by Das [13].

The growth of infants and young children throughout the world is always related to the socio-economic environment in which they live [13]. Likewise, the results of this study in Malawi show that children from richer families were less likely to be

malnourished compared with those from poorer families. The results also show a significant association between the household source of drinking water and child nutritional status (χ^2 = 52.8, *p-value* = 0.000). The prevalence of stunting was very low for children from households using piped water (20.1%) as their source of drinking water compared with those using well water and water from other sources.

On the other hand, the prevalence of diarrhoea and fever showed no association with the prevalence of stunting. However, the results from Table 2.2 show a strong and significant association between the duration of breastfeeding and child nutritional status (χ^2 = 36.0, *p-value* = 0.000). The prevalence of stunting was found to be higher for children who were first breastfed and had mothers who later decided to stop breastfeeding. Low levels of stunting were found among children who got milk from their mothers in a consistent manner, with 76.3% of them being nourished

2.3 Summary

In the current chapter (EDA) the study utilized a quantitative research method appropriate for analyzing the relationship between two or more variables known as cross-tabulation. The use of cross-tabulation enabled us to examine associations within our selected independent variables and the ordered categorical response variable (child nutrition status). The Pearson's chi-square test, or the chi-square test of association, was used to discover if there is a relationship between the level of stunting and each independent variable. We have an increased understanding of the data as we were able to isolate the most important variables. To examine, these further it is necessary that additional statistical methods be used. The results show that some of our independent variables are highly significant with our response variables. Therefore, we will examine these variables more closely when fitting our models in the subsequent chapters.

Table 2.2: Children's nutritional status according to selected independent variables

Characteristics	Level of Stunting (Height-for-Age)				
	Severe (z-score:< -3.00)	**Moderate** (z-score:-3.00 to -2.01)	**No stunting** (z-score:≥ 2.00)	**Total**	**Chi-square (p-value)** (χ^2)
Basic Predictors of Child Malnutrition					
Child's Age (in months)					
$\leq$6	11 (1.9)	43 (7.5)	518 (90.6)	572	175.8(0.000)
7-11	14 (3.2)	63 (14.4)	359 (82.3)	436	
12-23	81 (7.7)	280 (26.6)	690 (65.7)	1051	
24-35	70 (6.9)	219 (21.5)	729 (71.6)	1018	
36-47	114 (10.8)	243 (23.0)	701 (66.3)	1058	
48-59	78 (8.2)	217 (22.7)	662 (69.2)	957	
Sex					
Male	190 (7.6)	526 (21.0)	1784 (71.4)	2500	1.15(0.562)
Female	178 (6.9)	539 (20.8)	1875 (72.3)	2592	
Birth order					
1	87 (6.9)	279 (22.0)	901 (71.1)	1267	7.5(0.279)
2-3	134 (6.8)	394 (20.0)	1439 (73.2)	1967	
2-3	83 (7.2)	235 (20.5)	831 (72.3)	1149	
6^+	64 (9.0)	157 (22.1)	488 (68.8)	709	
Residence					
Rural	327 (7.7)	947 (22.2)	2996 (70.2)	4270	37.5(0.000)
Urban	41 (5.0)	118 (14.4)	663 (80.7)	822	
Mother's education level					
No education	73 (11.6)	150 (23.8)	408 (64.7)	631	62.7(0.000)
Primary	246 (7.4)	725 (21.8)	2355 (70.8)	3326	
Secondary	49 (4.6)	183 (17.3)	823 (78.0)	1055	
Higher	0 (0.0)	7 (8.8)	73 (91.3)	80	
Father's education level					
No education	30 (7.6)	95 (23.9)	272 (68.5)	397	50.0(0.000)
Primary	199 (8.4)	516 (21.7)	1658 (69.9)	2373	
Secondary	69 (5.2)	252 (19.2)	994 (75.6)	1315	
Higher	5 (2.6)	15 (7.9)	171 (89.5)	191	

Continued on the next page

Table 2.2 – *Continued from the previous page*

Wealth status					
Poor	214 (9.5)	552 (24.6)	1480 (65.9)	2246	93.9(0.000)
Middle	61 (6.1)	223 (22.2)	721 (71.7)	1005	
Rich	93 (5.1)	290 (15.8)	1458 (79.2)	1841	
Mother's BMI					
Thin (BMI < 18.5)	24 (9.3)	66 (25.6)	168 (65.1)	258	6.1(0.047)
Normal (BMI ≥ 18.5)	342 (7.1)	998 (20.7)	3477 (72.2)	4817	
Mother's age at birth of respective child					
$\leq$18	17 (8.8)	36 (18.7)	140 (72.5)	193	19.4(0.013)
19-23	109 (8.0)	319 (23.3)	941 (68.7)	1369	
24-29	96 (6.0)	315 (19.8)	1177 (74.1)	1588	
30-34	64 (6.5)	190 (19.2)	738 (74.4)	992	
35^{+}	82 (8.6)	205 (21.6)	663 (69.8)	950	
Mother's working status					
Currently working	248 (7.4)	727 (21.6)	2393 (71.1)	3368	3.3(0.193)
Not working	120 (7.0)	338 (19.6)	1266 (73.4)	1724	
Father's occupation					
Not working	31 (8.6)	72 (19.9)	258 (71.5)	361	25.0(0.002)
Professional	17 (5.0)	52 (15.4)	269 (79.6)	338	
Business	18 (6.4)	47 (16.7)	216 (76.9)	281	
Agriculture	127 (7.5)	399 (23.6)	1168 (68.9)	1694	
Other^{+}	113 (6.9)	316 (19.4)	1199 (73.6)	1628	
Birth type					
Singleton	342 (6.9)	1015 (20.6)	3578 (72.5)	4935	38.0(0.000)
Multiple birth	26 (16.6)	50 (31.8)	81 (51.6)	157	
Duration of breastfeeding					
Ever, then stopped	262 (8.5)	691 (22.4)	2127 (69.1)	3080	36.0(0.000)
Never breastfed	4 (4.9)	19 (23.5)	58 (71.6)	81	
Still breastfeeding	102 (5.3)	355 (18.4)	1474 (76.3)	1931	
Source of water					
Piped water	48 (4.3)	175 (15.8)	887 (79.9)	1110	52.8(0.000)
Well water	284 (7.8)	821 (22.4)	2558 (69.8)	3663	
Rainwater, tank, other	36 (11.3)	69 (21.6)	214 (67.1)	319	

Continued on the next page

Table 2.2 – *Continued from the previous page*

Had diarrhea in last two weeks					
No	258 (7.3)	730 (20.6)	2558 (72.1)	3546	0.8(0.663)
Yes	109 (7.1)	334 (21.7)	1096 (71.2)	1539	
Had fever in last two weeks					
No	289 (7.2)	823 (20.5)	2903 (72.3)	4015	1.8(0.402)
Yes	77 (7.3)	236 (22.3)	743 (70.4)	1056	

Chapter 3

Generalized linear models (GLM)

3.1 Introduction

Nelder and Wedderburn (1972) pioneered the theory of generalized linear models (GLMs). This class of models is an extension of classical linear models that allows the mean of a population to depend on a linear predictor through a non-linear link function where the response probability distribution is a member of an exponential family of distributions. Generalized linear models were formulated by John Nelder and Robert Wedderburn as a way of unifying various other statistical models, including linear regression, logistic regression and poisson regression [19]. Many other books and journal articles followed the seminal work by Nelder andWedderburn (1972). McCullagh and Nelder (1989) (the original text was published in 1983) provided a detailed introduction to GLMs. The books by Aitkin et al. (1989) and Dobson (1990) are also excellent references with many examples of applications of GLMs.

GLMs are quickly becoming the premier statistical analysis method for different types of data from many sources. In the current study we consider the special case of a GLM model that deals with ordinal scales in which the categories are ordered much like the ordinal numbers, 'first', 'second', and so on. Agresti (1990) provides an excellent simple outline for dealing with categorical data using ordinary logistic regression models, which are special cases of GLMs. In this chapter we study the origin of the GLM and the special case of one that is imposed when we are dealing with categorical ordered data.

3.2 Model fitting

We now consider fitting a GLM to the data. The GLMs have been considered for a very long time by many researchers and are considered to be a family of flexible

statistical models. The model-fitting process that we use in this study involves four steps [20]:

1. Model specification – A model is specified in two parts: an equation linking the response and explanatory variables through the probability distribution of the response variable.
2. Estimation of the parameters of the model.
3. Checking the adequacy of the model – How well it fits or summarizes the data.
4. Inference – Calculating confidence intervals and testing hypotheses about the parameters in the model and interpreting the results.

In the subsequent section we briefly discuss the theory of GLMs and their extensions. From the exploratory data analysis we learnt that our response variable is both categorical with 3 categories and ordinal. Hence, We focus more on GLMs which are more appropriate for categorical data, especially when there is a natural order on the categorical response variable.

3.3 Component of generalized linear models

Generalized linear models are an extension of classical linear models, so that the latter form a suitable starting point for discussion. A vector of observations $\mathbf{y}$ having n components is assumed to be a realization of a random variable $\mathbf{Y}$ whose components are independently distributed with means $\boldsymbol{\mu}$. The systematic part of the model is a specification for the vector $\boldsymbol{\mu}$ in terms of a small number of unknown parameters $\beta_1, \cdots, \beta_p$. In the case of ordinary linear models, this specification takes the form

$$\boldsymbol{\mu} = \sum_{1}^{p} \boldsymbol{x_j} \beta_j \tag{3.1}$$

where the βs are parameters whose values are usually unknown and have to be estimated from the data. If we let i index the observations then the systematic part of the model may be written as

$$E(\boldsymbol{Y_i}) = \mu_i = \sum_{1}^{p} \boldsymbol{x_{ij}} \beta_j; i = 1, 2, \cdots, n \tag{3.2}$$

where $\boldsymbol{x_{ij}}$ is the value of the jth covariate for observation i. In matrix notation (where $\boldsymbol{\mu}$ is $n \times 1$, $\boldsymbol{X}$ is $n \times p$ and $\boldsymbol{\beta}$ is $p \times 1$) we may write

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$$

where $\boldsymbol{X}$ is the model matrix and $\boldsymbol{\beta}$ is the vector of parameters. This completes the specification of the systematic part of the model.
For the random part we assume independence and constant variance of errors. These assumptions are strong and need checking, as far as is possible, from the data themselves. Nelder and Wedderburn (1972) outline different types of excellent model checking techniques which are not included in this study.
A further specialization of the model involves the stronger assumption that the errors follow a Gaussian or Normal distribution with constant variance σ^2. We may thus summarize the classical linear model in the form:
The components of $\boldsymbol{X}$ are independent Normal variables with constant variance σ^2 and

$$E(\boldsymbol{Y}) = \boldsymbol{\mu}, \text{where } \boldsymbol{\mu} = \boldsymbol{X\beta} \tag{3.3}$$

3.3.1 The generalization

To simplify the transition to generalized linear models, we shall rearrange equation slightly to produce the following three-part specification:

1. The *random component*: the components of $\boldsymbol{Y}$ have independent Normal distributions with E($\boldsymbol{X}$) = $\boldsymbol{\mu}$ and constant variance σ^2;
2. The *systematic component*: covariates $x_1, x_2, \cdots, x_p$ produce a linear predictor $\boldsymbol{\eta}$ given by

$$\boldsymbol{\eta} = \sum_{1}^{p} \boldsymbol{x_j}\beta_j;$$

3. The *link* between the random and systematic components:

$$\boldsymbol{\mu} = \boldsymbol{\eta}$$

This generalization introduces a new symbol η for the linear predictor and the third component then specifies that μ and η are in fact identical. If we write

$$\eta_i = g(\mu_i),$$

then $g(\cdot)$ will be called the *link function*. In this formulation, classical linear models have a Normal (or Gaussian) distribution in component 1 and the identity function for the link in component 3. Generalized linear models allow two extensions;
First the distribution in component 1 may come from an exponential family other than the Normal, and
Secondly the link function in component 3 may be any monotonic differentiable

function.
Now in simple terms we can see that GLMs are a natural generalization of classical linear models that allow the mean of a population to depend on a linear predictor through a (possibly nonlinear) link function. This allows the response probability distribution to be any member of the exponential family of distributions.

3.3.2 Likelihood functions for GLMs

We assume that each component of $\boldsymbol{Y}$ has a distribution in the exponential family, taking the form

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \tag{3.4}$$

for some specific functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$. If ϕ is known, this is an exponential-family model with canonical parameter θ. It may or may not be a two-parameter exponential family if ϕ is unknown.

We write $l(\theta,\phi;y) = \log f_Y(y;\theta,\phi)$ for the log-likelihood function considered as a function of θ and ϕ, y being given. The mean and variance of $\boldsymbol{Y}$ can be derived easily from the well known relations

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0 \tag{3.5}$$

and

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0 \tag{3.6}$$

We have from (3.5) that

$$l(\theta;y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)$$

hence

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \tag{3.7}$$

and

$$\frac{\partial^2 l}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)}. \tag{3.8}$$

From (3.5) and (3.7) we have

$$0 = E\left(\frac{\partial l}{\partial \theta}\right) = \{\mu - b'(\theta)\}/a(\phi)$$

so that

$$E(\boldsymbol{Y}) = \mu = b'(\theta).$$

Similarly from (3.6), (3.7) and (3.8) we have

$$0 = \frac{-b''(\theta)}{a(\phi)} + \frac{var(Y)}{a^2(\phi)}$$

so that

$$var(Y) = b''(\theta)a(\phi)$$

Thus the variance of Y is the product of two functions; one, $b''(\theta)$, depends on the canonical parameter (and hence on the mean) only and will be called the *variance function*, while the other is independent of θ and depends only on ϕ. The variance function considered as a function of μ will be written $V(\mu)$.
The function $a(\phi)$ is commonly of the form

$$a(\phi) = \phi/w$$

where ϕ, also denoted by σ and called the dispersion parameter, is constant over observations, and w is a known prior weight that varies from observation to observation.
Table 3.1 below gives the different model components of the GLMs that are commonly used.

Table 3.1: Examples of GLMs

Y$\backsim$	**Normal(μ, σ^2)**	**Gamma(α, β)**	**Poisson(λ)**	**Bin($(m, q)/m$)**
$Link(g)$	identity	reciprocal	log	logit
E(Y) = $\mu(\theta)$	$\theta = \mu$	$-\theta^{-1} = \frac{\alpha}{\beta}$	$e^\theta = \lambda$	$\frac{e^\theta}{1+e^\theta} = q$
$V(Y) = V(\mu)\alpha$	σ^2	$\frac{1}{\theta^2\alpha} = \frac{\alpha}{\beta^2}$	$e^\theta = \lambda$	$\frac{q(1-q)}{m}$
$V(\mu)$	1	θ^{-2}	$e^\theta = \lambda$	$q(1-q)$
ϕ	σ^2	α^{-1}	1	$1/m$
$c(y, \phi)$	$-\frac{1}{2}[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)]$	$\alpha\ln(\alpha y) + \ln y + ln\Gamma(\alpha)$	$-\ln(y!)$	$\ln\begin{pmatrix} m \\ my \end{pmatrix}$

3.4 Estimation methods using GLM

When the response variable is categorical, with more than two categories, then there are two options from GLMs. One relies on generalizations of logistic regression from dichotomous responses to nominal or ordinal responses with more than two categories. The current study will use this approach to create a model for under-five-year-old child malnutrition. The other option is to model the frequencies or counts for the covariate patterns as the response variables with Poisson distributions. The second approach, called log-linear modelling, is covered in this study.

For nominal or ordinal logistic regression, one of the measured or observed categorical variables is regarded as the response, and all other variables are explanatory variables. For log-linear models, all the variables are treated alike. The choice of which approach to use in a particular situation depends on whether one variable is clearly a 'response' (for example, the outcome of a prospective study) or whether several variables have the same status (as may be the situation in a cross-sectional study). Additionally, the choice may depend on how the results are to be presented and interpreted. Nominal and ordinal logistic regression yields odds ratio estimates, which are relatively easy to interpret if there are no interactions (or only fairly simple interactions). Log-linear models are good for testing hypotheses about complex interactions, but the parameter estimates are less easily interpreted [20].

In the next section we begin with a short summary of multinomial distribution, which provides the basis for modelling categorical data with more than two categories. Then the various formulations of ordinal logistic regression models are discussed, including the interpretation of parameter estimates and methods for checking the adequacy of a model.

3.4.1 Multinomial distribution

Consider a random variable $\boldsymbol{Y}$ with J categories. Let $\pi_1, \pi_2, \cdots, \pi_J$ denote the respective probabilities, with $\pi_1 + \pi_2 + \cdots + \pi_J = 1$. If there are n independent observations of $\boldsymbol{Y}$ which result in y_1 outcomes in category 1, y_2 outcomes in category 2, and so on then let

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}, \text{ with } \sum_{j=1}^{J} y_j = n.$$

The multinomial distribution is

$$f(y|n) = \frac{n!}{y_1!y_2!\cdots y_J!}\pi_1^{y_1}\pi_2^{y_2}\cdots\pi_J^{y_J} \tag{3.9}$$

If $J = 2$, then $\pi_2 = 1 - \pi_1, y_2 = n - y_1$ and (3.9) is the binomial distribution.
Agresti (1990) states that for multinomial distribution (3.9),
$E(Y_j) = n\pi_j, var(Y_j) = n\pi_j(1 - \pi_j)$, and $cov(Y_j, Y_k) = -n\pi_j\pi_k$ [21].

We also have Nominal logistic regression models as an option to deal with categorical data. However, nominal logistic regression can be only be used when there is no natural order of the response categories. Hence, the current study do not consider nominal logistic regression models because there is an obvious natural order among the response categories (nutrition status), so we sought models where the ordinal nature can be taken into account in the model specification.

The current study has a situation where the response variable is a continuous variable z which measures the severity of disease (stunting). Under-five children with small values of z are classified as having 'severe disease' or 'moderate disease' and those with large values of z are classified as having 'no disease'. The cut-points say $C_1, \cdots, C_{J-1}$ define J categories with associated probabilities $\pi_1, \cdots, \pi_J$ (with $\sum_{j=1}^{J}\pi_j = 1$). For ordinal categories, there are several different commonly used models and some of them are described in the next section.

3.4.2 Ordinal logistic regression models

Ordinal regression models are closely related to logistic models for dichotomous outcomes. With only two categories for an outcome variable, logistic regression is used to model the likelihood of one of the outcomes, usually termed the success, as a function of a set of independent variables [22]. When the possible responses for an outcome variable consist of more than two categories and are ordinal in nature, the notion of "success" can be conceived of in many different ways. Regression models for ordinal response variables are designed for just this situation and are extensions of the logistic regression model for dichotomous data. The complexity in fitting ordinal regression models arises in part because there are so many different possibilities for how "success," and the consequent probability of "success," might be modelled [21, 23]. Moreover, when there is a need to take several split into consideration, special multivariate analysis for ordinal data is the natural alternative. There are many approaches, such as the use of mixed models or another class of models, for example, the probit link model, but the ordinal logistic regression model has been widely

used in most of the previous research works [10, 11, 12].
The analysis that imitate this method of dichotomizing the outcome, in which the successive dichotomizations form cumulative "splits" to the data, is referred to as proportional or cumulative odds (CO) [21]. It is one way to conceptualize how the data might be sequentially partitioned into dichotomous groups, while still taking advantage of the order of the response categories [22]. The ordinal nature of this approach is so appealing because of its similarity to logistic regression [24]. The goal of the cumulative odds model is to simultaneously consider the effects of a set of independent variables across these possible consecutive cumulative splits to the data.

A simplifying assumption is made of the data when applying ordinal regression models, and that is the assumption of proportional, or parallel odds. This assumption implies that the explanatory variables have the same effect on the odds, regardless of the different consecutive splits to the data[16]. Both SAS and SPSS provide a score test for the proportional odds assumption within their ordinal regression procedures, but this omnibus test for proportionality is not a powerful test and is anticonservative[25]; the test nearly always results in very small p values, particularly when the number of explanatory variables is large [26], the sample size is large, or continuous explanatory variables are included in the model [27]. By default, PROC LOGISTIC fits the proportional odds model combined with the cumulative logit link when the researcher has more than two response levels.
Therefore, conclusions about rejecting the null hypothesis of proportionality of the odds, based solely on the score test, should be made cautiously. Rejection of the assumption of parallelism (proportional odds) for the particular ordinal model being investigated implies that at least one of the explanatory variables may have a differential effect across the outcome levels, that is, there is an interaction between one or more of the independent variables and the derived splits to the data [28, 27]. The vital point is to be able to discover which variable(s) may result for the rejection of this overall test.

Now should the assumption of proportionality be violated, that is when we have a situation whereby one set of effects $\boldsymbol{X}$ has p_1 parameters that satisfy the parallel lines assumption (that is, they have equal slopes), but the remaining set $\boldsymbol{Z}$ has p_2 parameters that do not and instead require the general model (that is, they have unequal slop), hence a valid solution is to fit a partial proportional model (PPOM) [12]. There are two types of partial proportional model (PPOM), one without restriction (PPOM-UR) and another with restriction (PPOM-R). This approach relaxes the proportional odds assumption for some explanatory variables [29]. The interpretation of the proportional odds parameters is independent of the response function; inter-

pretation of the general parameters depends on the response function. When you fit the partial proportional odds model, you must be especially careful to ensure that the cumulative logits remain ordered for the data being modeled [30]. The functional form of POM and PPOM (Restricted and Unrestricted) has the form described below [12].

Proportional Odds Model (POM)

$$\boldsymbol{\lambda(\vec{x})} = \ln\left\{\frac{Pr(Y=1|\vec{x})+\cdots+Pr(Y=j|\vec{x})}{Pr(Y=j+1|\vec{x})+\cdots+Pr(Y=k|\vec{x})}\right\} = \ln\left\{\frac{\sum_1^j Pr(Y=j|\vec{x})}{\sum_{j+1}^k Pr(Y=j|\vec{x})}\right\}$$

$$\boldsymbol{\lambda(\vec{x})} = \alpha_j + (\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p), j = 1, 2, \cdots, k-1 \tag{3.10}$$

Unrestricted Partial Proportional Odds Model (PPOM-UR)

$$\boldsymbol{\lambda(\vec{x})} = \ln\left\{\frac{Pr(Y=1|\vec{x})+\cdots+Pr(Y=j|\vec{x})}{Pr(Y=j+1|\vec{x})+\cdots+Pr(Y=k|\vec{x})}\right\} = \ln\left\{\frac{\sum_1^j Pr(Y=j|\vec{x})}{\sum_{j+1}^k Pr(Y=j|\vec{x})}\right\}$$

$$\boldsymbol{\lambda(\vec{x})} = \alpha_j + [(\beta_1+\gamma_{j1})x_1 + \cdots + (\beta_q+\gamma_{jq})x_q + (\beta_{q+1})x_{q+1} + \cdots + \beta_p x_p], j = 1, \cdots, k-1 \tag{3.11}$$

Restricted Partial Proportional Odds Model (PPOM-R)

$$\boldsymbol{\lambda(\vec{x})} = \ln\left\{\frac{Pr(Y=1|\vec{x})+\cdots+Pr(Y=j|\vec{x})}{Pr(Y=j+1|\vec{x})+\cdots+Pr(Y=k|\vec{x})}\right\} = \ln\left\{\frac{\sum_1^j Pr(Y=j|\vec{x})}{\sum_{j+1}^k Pr(Y=j|\vec{x})}\right\}$$

$$\boldsymbol{\lambda(\vec{x})} = \alpha_j + [(\beta_1+\gamma_{j1})x_1 + \cdots + (\beta_q+\gamma_{jq})x_q + (\beta_{q+1})x_{q+1} + \cdots + \beta_p x_p], j = 1, \cdots, k-1 \tag{3.12}$$

Where Y is response variable, $\vec{x}$ is the vector of explanatory variables $= (x_1, x_2, \cdots, x_p)$, α_j are intercepts and $(\beta_1, \beta_2, \cdots, \beta_p)$ are logit coefficients. Equation (3.10) is valid when originally a continuous response variable was subsequently grouped and the proportional odds assumption is satisfied, Equation (3.11) is valid when the proportional odds assumption is not valid and equation (3.12) is used when the proportional odds assumption is not satisfied and linear relationship for odds ratio between covariate and the response variable are not valid [10]. One can also consider another simple and valid approach to analyze the data by dichotomizing the ordinal response variable by means of several cut-off points and use separate binary logistic regression models for each dichotomous response variable [31]. However, Gameroff (2005) suggested that this second procedure should be avoided if possible because of the loss in statistical power and the reduced generality of the analytical solution. [32, 12].

3.5 Application of proportional odds model to determine risk factors for stunting in under-five children in Malawi

The fitted response variable (nutritional status) is ordinal in nature (derived from continuous variable height-for-age anthropometric index); hence, the PO model was formed. The chi-square score test for the proportional odds assumption was used to see whether the main model assumption was violated or not. As the score test is often anti-conservative (i.e the resulting p-value are far too small) [12], we used another technique to investigate the proportional odds assumption. The single score tests for each covariates was calculated as an alternative method of testing whether the proportional odds assumption is violated [33]. The estimated effects are displayed in Table 3.2.

Firstly, from Table 3.2, the score test of the proportional odds assumption is found to be insignificant at 5% level of significance (χ^2= 36.94, $p - value = 0.3347$) indicating that the data satisfy the proportional odds assumption. In addition, this indicates that for each of the chosen covariates a single parameter can be used to model the effect of a covariate to the separate logits of cumulative probabilities. However, the p-value of the score test is small (0.3347). To confirm the conclusion regarding the proportional odds assumption, single score tests for each covariate were conducted. The p-values of the single score tests are shown in the last column of Table 3.2. The results indicate that all the covariates were found to be insignificant (p-value ≤ 0.005) and hence, they satisfy the odds assumption. From the latter results, we were able to make a final decision that our dataset satisfies the proportional odds assumption for POM.

The column labeled odds ratio in Table 3.2 displays the values of the estimated adjusted odds ratios. Therefore, the interpretation of our results is based on adjusted odd ratios. From the results of POM in Table (3.2), is evidence that the risk of having worse nutrition status were 3.62, 4.19 and 3.91 higher among children belonging to the age group 12-23, 36-47 and 48^+ months respectively, when compared with the infants. This results supports our earlier hypothesis; that stunting can be clearly identified when the child is two years old. Moreover, this result supports earlier evidence from a study on stunting that was done by Das (2008) in Bangladesh. It can be surmised that when a child is growing, the mother's milk alone becomes insufficient for feeding and as a result, we can confidently state that the reason for increased stunting by 12 months of age may be due to deficiency of supplementary foods when the breastmilk alone is no longer adequate. The risk of stunting in under-five children can worsen with age, but that occurs only up to a critical age (47

months), beyond which a child's stunting status improves.

Children of multiple births are likely to be more malnourished than singletons (Table 3.2), the multiple births had 3.66 times greater risk of having poor nutrition status when compared with children of single births (p-value = $<.0001$). This could be explained by low birth weight, inadequate breastfeeding and competition for nutritional intake, which tend to afflict children of multiple births more than singletons. In addition, parents can take better care of fewer children, which can be added as a justification for the high risk of stunting in multiple birth children. We also applaud the Human Fertilization and Embryology Authority in its efforts to reduce multiple births, the study suggest that it is time for African nations to follow other nations in imposing a single embryo transfer policy – this alone will play a significant role in decreasing the risk of childhood malnutrition, hence the prevalence of stunting.

Our results from Table (3.2) reveal that the difference in the level of a mother's education plays a significant role in the risk of poor nutritional status of a child. The children of illiterate mothers (no education) have a high likelihood of having a poor nutritional status (about 2.0 times) when compared with mothers with higher educational qualifications. Compared with children of fathers with higher levels of education, the risk of stunting was found to double (2.144) for children of fathers with only primary school as their highest level of education. These results highlight the importance of education as an important determinant of child malnutrition in Malawi. This finding is common and a number of studies have attempted to explain this difference in malnutrition in relation to education levels of parents (e.g. Das and Rahman, 2011; Chirwa and Ngalawa 2008). There is a belief that educated mothers may be more conscious about the health of their children. As a result of their educated status, they are easily able to research, identify and adopt new feeding practices, which may help to significantly improve the nutritional status of their children.

Household factors make a strong contribution to nutritional status of under-five children. Usually children in households with the lowest income have the worst nutritional status and this improves with the increased wealth status of the household [13, 12]. In the current study, the results indicate that children from households with a poor wealth index status are 1.55 times more at risk of a poor nutrition status as compared with children from households classified under the rich wealth index status (Table 3.2). However, there was no significant difference in the likelihood of a poor nutrition status between children from households with either moderate or rich wealth index status. The results support the statement that the growth of infants

and younger children throughout the world is related to the socioeconomic environment in which they live [13].

The odds of having poor nutritional condition was found to be significantly different between children of mothers with normal (BMI ≥ 18.5) and thin (BMI < 18.5) body mass index (BMI). Children of mothers with a BMI indicating underweight, compared with those with a normal body mass index, are about 1.5 times more likely to be severely stunted. The odds of having malnutrition are lower for children born to mothers over 18 years of age birth when compared with mothers who gave birth when under 18 (teenaged). Results in Table(3.2) reveal that children born to mothers aged 19-23, 24-29 and 30-34 were 0.89, 0.577 and 0.57 less likely to have chronic malnutrition (stunting) when compared with children of mothers who gave birth when they were below the age of 18 years. The results support the inherent results that generally, the children born to mothers below 18 and above 34 are more susceptible to malnourishment than children born to mothers aged 18-34 [12, 18].

Children who suffered from diarrhoea within the two weeks prior to the survey were found to be at 1.2 times higher risk of being malnourished when compared with children who did not suffer from diarrhoea during the same time period (Table 3.2). This suggests that the illness history of the child can serve as one of the determinants of malnutrition.

3.6 Summary

In summary of the current chapter and results from the proportional odds model, we got an evidence that the age of the child, birth type (single or multiple birth), education level of mother and father, type of residence, mother's BMI, mother's age at birth of the respective child and a child's experience of diarrhea in the last two weeks before the survey were found to be independent predictors of under-five child nutrition status.

Table 3.2: Results of the multiple POM for nutrition status

Covariate	Regression coefficient	Standard error	p-value	Odds ratios	Single score test (p-value)
$Intercept_1$	-2.1226	0.2037	<.0001	-	-
$Intercept_2$	-0.4293	0.1993	0.0313	-	-
Child's age (in months) [0-11 months as reference]					
12-23	0.2632	0.0894	0.0032	3.620	**0.056**
24-35	0.0074	0.0776	0.9236	2.803	
36-47	0.4106	0.0818	<.0001	4.195	
48-59	0.3420	0.0877	<.0001	3.917	
Sex of child [male as reference]					
Female	-0.0241	0.0351	0.4932	0.953	**0.382**
Birth order [first as reference]					
2-3	-0.0227	0.0685	0.7397	1.087	**0.673**
4-5	0.0248	0.0758	0.7433	1.140	
6+	0.1043	0.1089	0.3381	1.235	
Type of birth [singleton as reference]					
Multiple	0.6485	0.0901	<.0001	3.658	**0.718**
Type of residence [rural as reference]					
Urban	-0.0143	0.0656	0.8279	0.972	**0.446**
Mother's education level [higher as reference]					
No education	0.3083	0.1437	0.0319	1.735	**0.130**
Primary	-0.0188	0.1247	0.8804	1.251	
Secondary	-0.0471	0.1298	0.7168	1.216	
Father's education level [higher as reference]					
No education	0.1389	0.1197	0.2458	1.921	**0.358**
Primary	0.2488	0.0879	0.0047	2.144	
Secondary	0.1263	0.0896	0.1585	1.897	
Household wealth index [rich as reference]					
Poor	0.2279	0.0516	<.0001	1.551	**0.274**
Middle	-0.0173	0.0591	0.7698	1.213	
Mother's BMI [normal as reference]					
Thinness (BMI < 18.5)	0.1867	0.0775	0.0160	1.453	**0.862**

Continued on the next page

Table 3.2 – *Continued from the previous page*

Mother's age at birth [$\leq$18 as reference]					
19-23	0.2347	0.0895	0.0087	0.894	**0.565**
24-29	-0.2028	0.0820	0.0134	0.577	
30-34	-0.2150	0.0991	0.0301	0.570	
35+	-0.1638	0.1211	0.1762	0.600	
Mother's working status [not working as reference]					
Currently working	0.0343	0.0407	0.3992	1.071	**0.575**
Mother's age at birth [professional as reference]					
Not working	-0.0136	0.1135	0.9049	0.921	**0.530**
Business	-0.0855	0.1259	0.4972	0.857	
Agriculture	0.0401	0.0711	0.5731	0.972	
Other	-0.0093	0.0679	0.8910	0.925	
Source of water [piped water as reference]					
Well water	-0.0002	0.0620	0.9975	1.141	**0.161**
Other, rainwater, tank, etc.	0.1327	0.0970	0.1711	1.304	
Duration of breastfeeding [still breastfeeding as reference]					
Ever, then stopped	0.0581	0.1081	0.5908	0.966	**0.291**
Never breastfed	-0.1510	0.1916	0.4306	0.784	
Had diarrhea recently [no as reference]					
Yes	0.0898	0.0453	0.0472	1.197	**0.503**
Had fever in last two weeks [no as reference]					
Yes	0.0105	0.0392	0.7883	1.021	**0.489**
Score test for the Proportional Odds Assumption: Chi-Square (χ^2) = 36.94, df = 34, p-value = 0.3347					
Goodness-of-fit of overall model (Likelihood Ratio): Chi-Square (χ^2) = 330.60, df = 34, p-value = $<$.0001					
Sample size: 4244					

Chapter 4

Generalized additive mixed Models (GAMMs)

4.1 Introduction

The generalized linear model (GLM) (McCullagh and Nelder 1989) neatly synthesizes likelihood-based approaches to regression analysis for a variety of outcome measures. In Chapter 3 we analyzed the malnutrition data using the generalized linear model through the use of a proportional odds model (POM). GLMs, as parametric models, offer a strong tool for modelling the relationship between the response variable and predictor variables when their assumptions meet [34]. However, these models may suffer from inflexibility in modelling complex relationship between the response variable and the predictor variables in some applications. The parametric mean assumption may not always be desirable, as suitable functional forms of the predictor variables may not be known in advance and the response variable may depend on the covariates in a complicated manner [35].

Generalized additive mixed models (GAMMs) are proposed for over dispersed and correlated data, which arise frequently in studies involving clustered, hierarchical and spatial designs. GAMMs are extension of GAMs that include random effects when fitted. This class of models allows flexible functional dependence of an outcome variable on covariates by the use of nonparametric regression, while accounting for correlation between observations by using random effects. The interesting aspect is that the GAMM relaxes the assumption of normality and linearity inherent in linear regression. The flexibility of nonparametric regression for continuous predictor variables, coupled with linear models for predictor variables, offers ways to reveal structure within the data that may miss linear assumptions [34].
By imposing the flexibility of GAMM, in the current Chapter we use semi-parametric

logistic mixed model to assess the risk factors associated to under-five years children malnutrition. There exists many nonparametric regression models and a number of techniques of smoothing for independent data on which we are going to utilize herein.

A smoother is a tool for summarizing the trend of a response measurement Y as a function of one or more predictor measurements $X_1, \cdots, X_p$. It produces an estimate of the trend that is less variable than Y itself; hence the name *smoother*. An important property of a smoother is its nonparametric nature: it does not assume a rigid linear form for the independence of Y on $X_1, \cdots, X_p$. For this reason, a smoother is often referred to as a tool for nonparametric regressions [36].The most commonly used are splines smoothers, kernel smoothers, locally-weighted running-line smoothers and running-mean smoothers. These methods are well detailed in Hastie and Tibshirani (1990); Hardle (1999) and Green and Silverman (1993) [36, 37, 38].

4.2 Generalized additive mixed model (GAMM)

GAMs are extensions of GLMs in which a link function describing the total explained variance is modeled as a sum of the covariates. The terms of the model can in this case be local smoothers or simple transformations with fixed degrees of freedom (e.g. Maunder and Punt 2004) . In general the model has a structure of:

$$g(\mu_i) = X_i + f_1(x_{i1}) + f_2(x_{i2}) + f_3(x_{i3}) + \cdots$$

Where $\mu_i = E(Y_i)$ and X_i has an exponential family distribution. Y_i is a response variable, X_i is a row for the model matrix for any strictly parametric model component, θ is the corresponding parameter vector, and the are smooth functions of the covariates, x_k.
In regression studies, the coefficients tend to be considered fixed. However, there are cases in which it makes sense to assume some random coefficients. These cases typically occur in situations where the main interest is to make inferences on the entire population, from which some levels are randomly sampled. Consequently, a model with both fixed and random effects (so called mixed effects models) would be more appropriate. In the present study we utilize GAMM to investigate the effects of covariates on childhood malnutrition.
A fundamental feature of GAMM (8.1) over GAM is that the additive nonparametric functions are used to model covariate effects and random effects are used to model the correlation between observations [35]. The GAMM used in this study was first formulated by Zin and Zhang (1999). Suppose that the j^{th} observation of k^{th} units

consists of an outcome variable y_j and p covariates $x_j = (1, x_{j1}, \cdots, x_{jp})^T$ associated with fixed effects and $q \times 1$ of covariates z_j associated with random effects. GAMM is formulated as [35]:

$$g(\mu_j) = \beta_0 + f_1(x_{j1}) + \cdots + f_p(x_{ip}) + z_j b \tag{4.1}$$

where g(·) is a monotonic differentiable link function, $\mu_j = E(y_j|b)$, $f_j(\cdot)$ is a centered twice-differentiable smooth function, the random effect b is assumed to be distributed as $N\{0, K(\theta)\}$ and θ is $c \times 1$ vector of variance components. If $f_j(\cdot)$ is a linear function, then GAMM (4.1) reduces to generalized linear mixed model (GLMM) of Breslow and Clayton (1993).
For a given variance component θ, the integrated log-quasi-likelihood function of $(\beta_0, f_i, \theta, i = 1, 2, \cdots, k)$ is given by [35];

$$\exp[l\{y; \beta_0, f_1(\cdot), \cdots, f_p(\cdot), \theta\}] \propto |D|^{-\frac{1}{2}} \int \exp\left\{-\frac{1}{2\phi}\sum_{j=1}^{n} d_j(y_j; \mu_j) - \frac{1}{2} b^T D^{-1} b\right\} db, \tag{4.2}$$

where $y = (y_1, \cdots, y_n)^T$ and

$$d_i(y_i; \mu_i^b) \propto -2 \int m_i(y_i - u)/v(u) d(u)$$

defines the conditional deviance function of $\{\beta_0, f_1(\cdot), \cdots, f_p(\cdot)\}$ given b. For simplicity, D can be assumed to be a full rank matrix, otherwise Moore-Penrose generalized inverse may be used.
Statistical inference in the GAMM (4.1) involves inference on the nonparametric functions $f(\cdot)$, which often require the estimation of smoothing parameters, say λ, and inference on the variance component θ.

4.3 Inference on the nonparametric functions

4.3.1 Natural cubic smoothing spline estimation

A smoothing spline estimates the non-parametric regression function $f_j(\cdot)$ using a piecewise polynomial function with all the observed covariate values $\{X_j\}$ used as knots, where smoothness constraints are assumed at the knots [38, 39]. The most commonly used smoothing spline is the natural cubic smoothing spline, which assumes $f_j(\cdot)$ is piecewise cubic function, continuous and twice differentiable with step function third derivative at the knots X_i. Given the values of λ and θ and noting that $f_j(\cdot)$ are infinite dimension unknown parameters and that equation (4.1) is continuous linear $functional$ of the $f_j(\cdot)$, the results of O'Sullivan $etal.$ (1986). The

natural cubic smoothing spline estimator of the $f_j(\cdot)$ maximize the penalized log-quasi-likelihood given by;

$$l\{y;\beta_0,f_1(\cdot),\cdots,f_p(\cdot),\theta\}-\frac{1}{2}\sum_{j=1}^{p}\lambda\int_{s_j}^{t_j}f_j''(x)^2dx=l(y;\beta_0,f_1,\cdots,f_p,\theta)-\frac{1}{2}\sum_{j=1}^{p}\lambda_j f_j^T K_j f_j, \tag{4.3}$$

where (s_j,t_j) defines the range of the jth covariate and $\lambda=(\lambda_1,\cdots\lambda_p)^T$ is a vector of smoothing parameters and K_j is the smoothing spline penalty matrix [35], $f_j''(x)$ denotes the second derivative of $f_j(x)$. f_j is an $r_j\times 1$ unknown vector of the values of $f_j(\cdot)$ evaluated at the r_j ordered distinct values of the $x_{ij}(i=1,\cdots,n)$.
In matrix notation, with $\mu^b=(\mu_1^b,\cdot,\mu_n^b)^T$, $g(\mu^b)=\{g(\mu_1^b),\cdot,g(\mu_n^b)\}^T$ and $Z=(z_1,\cdots,z_n)^T$, GAMM (Equation 4.1) can be written as

$$g(\mu^b)=\mathbf{1}\beta_0+N_1f_1+\cdots+N_pf_p+Zb, \tag{4.4}$$

where $\mathbf{1}$ is a an $n\times 1$ vector of 1s and N_j is an $n\times 1$ incidence matrix defined in a way similar to that given in Green and Silverman (1994) [38], such that the ith component of N_jf_j is $f_j(x_{ij})$. It is often difficult to determine the full natural cubic smoothing spline estimator of f_j by directly maximizing expression (4.3). However, an approximating is therefore proposed in the next section by using Monte Carlo simulation.

4.3.2 Double penalized quasi-likelihood (DPQL)

Laplace approximation [40], some calculation shows that approximate natural cubic smoothing spline estimators $(\hat{\beta}_0,\hat{f}_1,\cdots,\hat{f}_p)$ can be obtained by maximizing the following DPQL with respect to $(\beta_0,f_1,\cdots,f_p)$ and b:

$$-\frac{1}{2\phi}\sum_{i=1}^{n}d_j(y_i;\mu_i^b)-\frac{1}{2}b^TD^{-1}b-\frac{1}{2}\sum_{j=1}^{p}\lambda_jf_j^TK_jf_j. \tag{4.5}$$

We want to show that the DPQL (4.5) estimator $\hat{f}_j$ can be easily obtained by fitting a GLMM using existing statistical softwares. Following Green (1987), Equation (4.2) and Zhang (1998), Equation (10) and also noting that f_j is a centered parameter vector, therefore we can parametrise f_j in terms of β_0 and $a_j((r_j-2)\times 1)$ via a one-to-one transformation as

$$f_j=X_jB_j+B_ja_j, \tag{4.6}$$

where X_j is an $r_j\times 1$ vector containing the r_j centered ordered *distinct* values of the x_{ij} $(i=1,\cdots,n)$, and $B_j=L_j(L_j^TL_j)^{-1}$ and L_j is an $r_j\times(r_j-2)$ full rank matrix satisfying $K_j=L_jL_j^T$ and $L_jX_J=0$. Using the identity $f_j^TK_jf_j=a_j^Ta_j$, DPQL (4.5)

becomes (compare Breslow and Clayton (1993)) [40], Equation (6)

$$-\frac{1}{2\phi}\sum_{i=1}^{n} d_i(y;\mu_i^b) - \frac{1}{2}b^T D^{-1} b - \frac{1}{2}a^T \Lambda^{-1} a, \tag{4.7}$$

where $a = (a_1^T, \cdots, a_p^T)$ and Λ=diag($\tau_1\mathbf{I}, \cdots, \tau_p\mathbf{I}$) with $\tau_j = 1/\lambda_j$. A small value of $\tau = (\tau_1, \cdots, \tau_p)^T$ corresponds to over-smoothing.
Plugging Equation (4.6) into Equation (4.4), Equation (4.7) suggest that, given θ and τ, the DPQL estimators $\hat{f}_j$ can be obtained by fitting the following GLMM by using Breslow and Clayton's (1993) [40] penalized quasi-likelihood approach:

$$g(\mu^b) = X\beta + Ba + Zb, \tag{4.8}$$

where $X = (1, N_1X_1, \cdots, N_pX_p)$, $B = (N_1B_1, \cdots, N_pX_p)$, $\beta = (\beta_0, \cdots, \beta_p)^T$ is a $(p+1)\times 1$ vector of regression coefficients and a and b are independent random effects with distribution $a \sim N(0, \Lambda)$ and $b \sim N(0, D)$.
The DPQL estimator $\hat{f}_j$ is calculated as $\hat{f}_j = X_j\hat{\beta}_j + \hat{B}_j a_j$, which is a linear combination of Breslow and Clayton (1993) [40] penalized quasi-likelihood estimators of the fixed effect $\hat{\beta}_j$ and the random effects $\hat{a_j}$ in the working GLMM (4.8) and be obtained by fitting the working GLMM (4.8) by using existing statistical software, such as the SAS macro GLIMMIX [35].

The maximization of expression (4.7) with respect to (β, a, b) can proceed by using the Fisher scoring algorithm to solve

$$\begin{pmatrix} X^TWX & X^TWB & X^TWZ \\ B^TWX & B^TWB+\Lambda^{-1} & B^TWZ \\ Z^TWX & Z^TWB & Z^TWZ+D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ a \\ b \end{pmatrix} = \begin{pmatrix} X^TWY \\ B^TWY \\ Z^TWY \end{pmatrix} \tag{4.9}$$

Where $Y = \beta_0 1 + \sum_{j=1}^{p} N_j f_j + Zb + \Delta(y-\mu^b)$ is a modified generalized additive model working vector, Δ=diag$\{g'(\mu_j)\}$ and W=diag$[\{\phi m_i^{-1} v(\mu_i^b) g'(\mu_i^b)^2\}^{-1}]$ is a modified generalized additive model working weight matrix.
An examination of equation (4.9) shows that it corresponds to the normal equation of the best linear unbiased predictors (BLUPs) of β and (a, b) under the linear mixed model

$$Y = X\beta + Ba + Zb + \epsilon \tag{4.10}$$

where a and b are independent random effects with $a \sim N(0, \Lambda)$ and $b \sim N(0, D)$ and $\epsilon \sim N(0, W^{-1})$. This suggests that the DPQL estimators $\hat{f}_j$ and the random effect estimators $\hat{b}$ can be easily obtained using the BLUPs by iteratively fitting model (4.10) to the working vector Y [35, 34].

To compute the covariance matrix of $\hat{f}_j$, it is more convenient to compute β and a by using

$$\begin{pmatrix} X^T R^{-1} X & X^T R^{-1} B \\ B^T R^{-1} X & B^T R^{-1} B + \Lambda^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ a \end{pmatrix} = \begin{pmatrix} X^T R^{-1} Y \\ B^T R^{-1} Y \end{pmatrix}, \tag{4.11}$$

where $R = W^{-1} + ZDZ^T$. Denoting by H the coefficient matrix on the left-hand side of equation (4.11) and $H_0 = (X, B)^T R^{-1}(X, B)$, the approximate covariance matrix of $\hat{\beta}$ and $\hat{a}$ is

$$cov(\hat{\beta}, \hat{a}) = H^{-1} H_0 H^{-1}, \tag{4.12}$$

It follows that the approximate covariance matrix of $\hat{f}_j$ is $(X_j, B_j) cov(\hat{\beta}_j, \hat{a_j})(X_j, B_j)^T$, where $cov(\hat{\beta}_j, \hat{a_j})$ can be easily obtained from the corresponding blocks of $H^{-1} H_0 H^{-1}$. We assume that the $f_j(\cdot)$ are fixed smooth functions in calculating the covariance of the f_j.

4.4 Inference on the smoothing parameters and the variance components

In the previous section (4.3), the smoothing parameters λ and the variance components θ were assumed to be known for making inferences on the nonparametric functions f_j. In the current section we discus one way to estimate λ and θ jointly by using a marginal quasi-likelihood by extending the REML approach of Wahba (1985), Kohn *et al.* (1991) and Zhang *et al.* (1998) [39, 41, 42]. The AQ key feature of this approach is that λ and θ can be easily obtained by fitting the working GLMM (4.8) via iteratively fitting the working linear mixed model (4.10) using REML with $\tau = (1/\lambda_1, \cdots, 1/\lambda_p)^T$ treated as extra variance component in addition to θ [35].

Under the classical nonparametric regression model

$$y = f(x_i) + \epsilon_i \tag{4.13}$$

where the ϵ_i are independent random errors following $N(0, \sigma^2)$. Wahba (1985) and Kohn *et al.* (1991) [39, 41] proposed to estimate the smoothing parameter λ by maximizing a marginal likelihood. The marginal likelihood of $\tau = 1/\lambda$ is constructed by assuming that $f(x)$ is estimated using a cubic smoothing spline in the form of equation (4.7) with a $a \sim N(0, \tau I)$ and a flat prior for β and integrating out a and β as follows:

$$\exp\{l_M(y; \tau, \sigma^2)\} \propto \tau^{-1/2} \int \exp\left\{ l(y; \beta, a, \sigma^2) - \frac{1}{2\tau} a^T a \right\} \mathrm{d}a \mathrm{d}\beta \tag{4.14}$$

where $l(y;\beta,a,\sigma^2)$ is the log-likelihood (normal) of f under model (4.13). Wehba (1985) [39] called the maximum marginal likelihood estimator of τ the generalized maximum likelihood estimator. Thompson (1985) pointed out that this marginal likelihood (4.14) of τ is in fact the REML under the linear mixed model

$$y = \mathbf{1}\beta + X\beta + Ba + \epsilon, \tag{4.15}$$

where $a \sim N(0,\tau I)$ and $\epsilon \sim N(0,\sigma^2 I)$, and B is the same as defined in Section (4.3); τ is regarded as a variance component. Hence the maximum marginal likelihood estimator of τ is an REML estimator. The extensive simulation study of Kohn **et al.** (1991) [41] showed that the maximum marginal likelihood estimator of τ has similar and often better performance compared with the generalized cross-validation (GCV) estimator in estimating the nonparametric function.
Zhang *et al.* (1998) [42] extended these researchers' results to estimate the smoothing parameter λ and the variance component θ jointly by using REML for longitudinal data with normally distributed outcomes and a nonparametric mean function. A representative special case of their model can be written as

$$y = f(X) + Zb + \epsilon, \tag{4.16}$$

where $f(X)$ denotes the values of nonparametric function $f(\cdot)$ evaluated at the design points of $X(n\times 1), b \sim N\{0, D(\theta)\}$ and $\epsilon \sim N\{0, G(\theta)\}$. If $f(\cdot)$ is estimated by using a natural cubic smoothing spline, using equation (4.7), Zhang *et al.* (1998) [42] rewrote model (18) as a linear mixed model

$$y = \mathbf{1}\beta_0 + X\beta_1 + Ba + Zb + \epsilon \tag{4.17}$$

where $a \sim N(0,\tau 1)$ and the distribution of b and ϵ are the same as those in model (4.16). They therefore proposed to treat τ as an extra variance component in addition to θ in model (4.17) and to estimate τ and θ by using REML. Using the results of Harville (1974) [43], this REML equations corresponds to the marginal likelihood of (τ,θ) constructed by assuming that f takes the form (4.7) with $a \sim N(0,\tau I)$ and a flat prior for β and integrating out a and β as follows:

$$\exp\{l_M(y;\tau,\theta)\} \propto |D|^{-1/2}\tau^{-1/2}\int \exp\left\{l(y;\beta,a,b) - \frac{1}{2}b^T D^{-1} b - \frac{1}{2\tau}a^T a\right\} \mathrm{d}b\mathrm{d}a\mathrm{d}\beta \tag{4.18}$$

where $l(y;\beta,a,b) = l(y;f,b)$ is the conditional log-likelihood (normal) of f given the random effect b under model (4.16). The marginal log-likelihood $l_M(y;\tau,\theta)$ in expression (4.18) has a closed form expression.

4.4.1 The marginal quasi-likelihood

Lin and Zhang (2002) [35] proposed to extend the marginal likelihood approach of Wahba (1985), Kohn *et al.* (1991) and Zhang *et al* (1998) [39, 41, 42] to GAMM (4.1) and to estimate τ and θ jointly by maximizing a marginal quasi-likelihood. They noted that the GLMM representation (4.8) suggests that τ can be treated as an extra variance components in addition to θ. Similarly to the REML (4.18), marginal quasi-likelihood of (τ, θ) under GAMM (4.1) can be constructed by assuming that f_j takes the form (4.7) with $a_j \sim N(0, \tau_j I)(j = 1, \cdots, p)$ and a flat prior for β and integrating the a_j and β out as follows:

$$exp\{l_M(y;\tau,\theta)\} \propto |\Lambda|^{-1/2} \int \exp\left\{l(y;\beta,a,\theta) - \frac{1}{2}a^T\Lambda^{-1}a\right\} \mathrm{d}b\mathrm{d}\beta$$

$$\propto |D|^{-1/2}|\Lambda|^{-1/2} \int \exp\left\{\sum_{i=1}^{n} -\frac{1}{2\phi}d_i(y;\mu_i^b) - \frac{1}{2}b^TD^{-1}b - \frac{1}{2}a^T\Lambda^{-1}a\right\} \mathrm{d}b\mathrm{d}a\mathrm{d}\beta \tag{4.19}$$

where $(y;\beta,a,\theta) = l(y;f_1,\cdots,f_p)$.
Since the evaluation of the marginal quasi-likelihood $l_M(y;\tau,\theta)$ in expression (4.19) often involves high dimensional integration [35], $l_M(y;\tau,\theta)$ is approximated using Laplace method. Specifically, taking a quadratic expansion of the exponent of the integrand of expression (4.19) about its mode before integration and approximating the deviance statistic $d_i(y;\mu_i^b)$ by the Pearson χ^2-statistic (Breslow and Clayton, 1993) [40], the derivation by Lin and Zhang (2002) [35] give the marginal quasi-likelihood as

$$l_M(y;\tau,\theta) \approx -\frac{1}{2}\log|V| - \frac{1}{2}\log|X^TV^{-1}X| - \frac{1}{2}(Y - X\hat{\beta})^TV^{-1}(Y - X\hat{\beta}), \tag{4.20}$$

where $V = B\Lambda B^T + ZDZ^T + W^{-1}$. An examination of Equation (4.20) shows that it corresponds to the REML log log-likelihood of the working vector Y under the linear mixed model (4.10) with both a and b as random effects and τ and θ as variance components. It follows that that we can easily estimate τ and θ by iteratively fitting model (4.10) using REML.

4.5 Application of GAMM to the determinants of risk factors of stunting in under-five children from Malawi

In the previous Chapter (4) we studied the risk factors of stunting (height-for-age) through the use of a generalized linear model, the proportional odds model (POM). Proportional odds models, discussed in Chapter 4 assume all independent variables

are fixed effects, hence they exclude the possibility of modeling a model with random effects. Another major limitation of these ordinal logistic models (Chapter 4) is that they are parametric. In these methods (logistic regression models) the relationship of the mean of a response to covariates is always assumed fully parametric. Although such parametric mean models may enjoy simplicity, hence, as result, the parametric models have suffered from inflexibility in modeling complicated relationships between the response and covariates in various longitudinal studies.
The latter deficiency in logistic regression models has motivated the current study to use alternative models to improve the results from the previous chapter (4). We had the option of trying to improve our logistic regression model (GLM) using a model that accommodates the random effect factor (GLMM).However, after considering the small difference between the two models, and since we will still be working with parametric models, we decided on other options.
In this chapter we have applied some data exploration by accessing the relationship between stunting (response) and the continuous covariates following Zuur *et al.*(2010) [44, 45]. Our main objective was to apply a nonparametric model if a nonlinear relationship between response and covariates was observed, using a generalized additive mixed model (GAMM). The use of this model is motivated by its ability to accommodate both fixed and random effects on its formulation and the ability to model the covariates, which show a non-linear relationship with the response nonparametrically, while keeping other covariates parametric. Table 4.1 below present the relevant variables to be used in fitting the model.

4.5.1 Relationships

Figure 4.1 shows the relationship between stunting (nutrition status) and the selected continuous variables of child's age in months, mother's age and mother's body mass index. In the previous chapter (4) we only fitted a linear model (POM) to these data, and therefore the relationships between the responses and covariates were assumed to be linear. However, if we look at the scatter plots (figure 4.1) between stunting and the three independent variables (child's age, mother's age and mother's BMI), we can clearly see a pattern that does not appear linear. For this reason, if we want to obtain the best possible results in terms of modelling stunting, we may need to employ models for a non-linear relationship. As a result, we attempted to model the relation between stunting and these covariates with a nonparametric smoother.

Table 4.1: Relevant variables from Malawi DHS-dataset

Variable	Description	Type
Child's age	Age of the child in months	Continuous covariate
Birth type	Child is of single/multiple birth	Categorical covariates
Child suffered from diarrhea	Child suffered from diarrhea in the last two weeks	Categorical covariate
Mother's education	Highest level of education of mother	Categorical covariate
Mother's BMI	Body mass index of mother	Continuous covariate
Mother's age	Age of mother at birth of child	Continuous covariate
Father's education	Highest level of education of father	Categorical covariate
wealth index	Household wealth index	Categorical covariate
Type of residence	Type of residence	Categorical covariate
Nutrition status	Stunting	Binary response variable

4.5.2 Model formulation

In formulating the model to use, we first took into account that the data exploration done above (section 4.5.1) indicated a potential non-linear relationship be- tween stunting and child's age, mother's age and mother's BMI. These covariates enter the model as nonparametric effects. Summary statistics for these covariates appear in Table (2.2). The remaining discrete covariates in Table (4.1) enter the model as parametric effects, they are also summarized in Table (2.2).
Secondly, because the data were collected using multi-stage sampling, where the primary sampling units or villages were selected at random, this may result in some variability among these primary sampling units. Therefore, in order to account for the possible variability between the primary sampling units, GLMM can be imposed. If this is indeed the case, then we can consider models of the form:

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1 \times \text{Birth type}_i + \beta_2 \times \text{Mother's education}_i \\ &+ \beta_3 \times \text{Father's education}_i + \beta_4 \times \text{Type of Residence}_i \\ &+ \beta_5 \times \text{Child suffered from diarrhea}_i + \beta_6 \times \text{wealth index}_i \\ &+ f_1(\text{Child's age}_i) + f_2(\text{Mother's age}_i) + f_3(\text{Mother's BMI}_i) + b_{0i} \end{aligned} \tag{4.21}$$

where $g(\cdot)$ is the logit link function, β's are parametric regression coefficients, f_i's are centered smooth functions and b_{0i} is the random effect distributed as $N(0, K(\theta))$. In

coming up with the final model (4.21), we have considered a possible interaction effects, one way to compare these models is via the Akaike Information Criterion (AIC), thus AIC for each model was examined and also the inference of smooth function along with the p-value of the individual term was examined. Hence the final model (4.21) was selected accordingly following Zuur *et al*.(2013) [46].

R software was used to fit the model. **nlme, mgcv and gamm4** are the libraries that can call GAMM models in R. In the terminology of R categorical variables are entered as factors, so a variable like mother's education that has four distinct levels accounts for 4 model parameters.

4.5.3 Interpretation of the results from GAMM

When fitting the model of this chapter, we had the problem of choosing whether we should fit a GLMM or a GAMM. One of the considerations discussed in Zuur et al. (2014) is to start with GAMM and if cross-validation gives one degree for a smoother, then this covariate is fitted as a parametric term. If all terms are parametric, then the model is a GLMM. The results from our GAMM (4.21) are presented in Table 4.2, Table 4.3 and Figure 4.2.
The results in Table 4.2 reveals that birth type (singleton/multiple) of a child is a highly significant effect to nutrition status of under-five years children in Malawi (p-value = 6.19×10^{-12}). A child born from multiple births is found to be 3.9 ($e^{1.3693}$) times more likely to suffer from stunting when compared to a child born from single birth. This results are consistent from what was observed in chapter 4 on POM.
We also can discern that a mother's level of education significantly affects nutrition status (stunting) of children in Malawi. The extent of the risk for a child to suffer from malnutrition (stunting) decreases with the increase in a mother's educational attainment. Numerically, the findings tell us that children born to mothers with primary or secondary education as their highest level of training are 0.7219 or 0.6890 ($e^{-0.3258}$ or $e^{-0.3725}$) times less likely to be at risk of producing a stunted child when compared to children born to mothers with no education at all and the effect was found significant a p-value = 0.0051 and p-value = 0.0147, respectively. Furthermore, the level of education of the father was also found to have a significant effect on the nutritional status of the child. Children born to fathers with higher or tertiary level education are 0.5403 ($e^{-0.6156}$ & p-value = 0.0477) times less likely to report bad nutrition status when compared to children born to fathers with no education.
The household wealth index status significantly affects the height-for-age (stunting) of a child in Malawi. From Table 4.2 we observe that children born from households with a middle wealth index status are 0.7787 times less at risk of being stunted than

children born to poor families and the effect is significant (p-value = 0.0024). Likewise, children from rich families are 0.6258 less likely to be stunted when compared to children from poor families (p-value = 2.41×10^{-6}).

Table 4.3 presents the results of fitted continuous covariates, which were fitted non-parametrically. To be more precise, the table is giving the approximate significance of smooth terms. The smoother for child's age is reported as more than six estimated degrees of freedom and is highly significant (p-value = $< 2 \times 10^{-16}$) in opposition to the assumption of a linear relationship between stunting and child's age. Likewise, mother's age and BMI also appeared to be significant against the assumption of a linear relationship with stunting. Mother's age reported test statistics of 136.49 with six estimated degrees of freedom and BMI reported test statistics of 15.03 with almost three estimated degrees of freedom. Furthermore, all of the smooth terms in Table 4.3 reported degrees of freedom that are greater than one; hence the significance of the smoothers is supported.

The estimated smoothers of the covariates; child,s age, mother's age and mother's BMI are presented in Figure 4.2. The shape of each smoother indicates a non-linear effect on stunting, which justifies the application of a GAMM instead of a GLMM. The shape of the child's age smoother shows a sharp linear increase in stunting up to 18 months and from 19-23 (months) the risk of stunting is observed to be at peak for under-five children in Malawi. A slight decrease in the risk of stunting is observed in children from 23-29 months old. Thereafter, a slight increase is observed up to 40 months. The smoother support observations of medical institutions that children aged 12–24 months face a high risk of stunting [17].

Figure 4.2 also presents a smoother for mother's age effect on stunting. We observe a high level of stunting corresponding to mothers less than 20 years of age and the stunting level declines with an increase of a mother's age up to age 30 years. Thereafter, a slightly linear effect of a mother's age is observed. As expected, children born to teenagers face a high-risk negative, or bad nutrition status. Moreover, the smoother curve of a mother's BMI indicates a decrease in stunting with increasing BMI. The effect of a mother's BMI is almost linear when seen visually, but statistically, it was found to be a non-linear effect. Hence, the results from this smoother are almost similar to what we observed in parametric regression models.

4.6 Summary

The main objective of the current chapter was to improve the model (GLM) used in the previous one. We improved the model by using a model that accommodates both fixed and random effects (GLMM). Moreover, we noticed that GLM and GLMM are both parametric, and they may also suffer from inflexibility to model complicated relationships between the outcome variable and predictor variables in some application. After assessing the relation between the outcome variable and the covariates we noticed some non-linear trends and were concerned about the application of GLMM. Therefore, we opted to improve our model even further to a model that is suitable for modelling any relationship between outcome variable and covariates: the generalized additive mixed model (GAMM).
For the parametric part of the model, some of the results from the GAMM were consistent with the findings of the previous chapter, but some were not. The effect of type of residence was found insignificant at 5% level of significance in GAMM and the effect of history of diarrhoea for a child was also found insignificant. All other covariates were significant and the trends were almost consistent with the results from the previous chapter (4).
The results from the nonparametric part model also validated the findings on child's age, mother's age and mother's BMI.

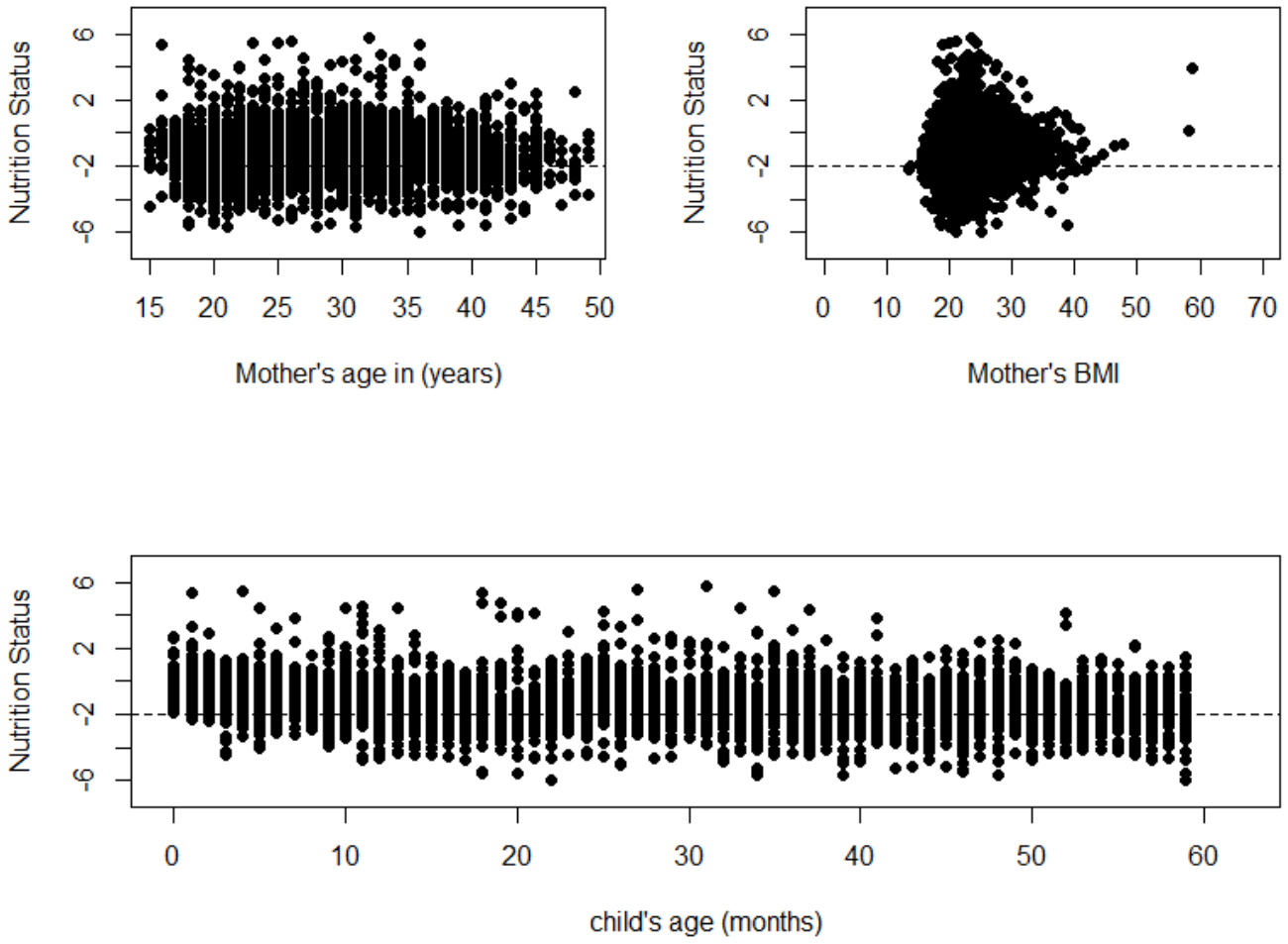

Figure 4.1: Scatter plots of the response variable (stunting; nutrition status of a child) versus selected covariates (child's age, mother's age and mother's BMI).

Table 4.2: The parameter estimates of the fixed effects of GAMM on malnutrition (stunting) for children under-five in Malawi

Co-variates	Regression coefficient	Standard error	z-value	p-value
Intercept	-0.771739	0.186704	-4.133	3.57×10^{-5}***
Type of birth [Singleton as reference]				
Multiple	1.369344	0.199168	6.875	6.19×10^{-12}***
Mother's education level [No education as reference]				
Primary	-0.325879	0.116355	-2.801	0.00510 **
Secondary	-0.372460	0.152634	-2.440	0.01468 *
Higher	-0.582283	0.475074	-1.226	0.22032
Father's education level [No education as reference]				
Primary	0.091030	0.130989	0.695	0.48709
Secondary	-0.008557	0.149923	-0.057	0.95449
Higher	-0.615618	0.310911	-1.980	0.04770 *
Type of residence [Urban as reference]				
Rural	0.083898	0.128048	0.655	0.51234
Had diarrhea [No as reference]				
Yes	0.175684	0.093363	1.882	0.05987
Household wealth index [Poor as reference]				
Middle	-0.250120	0.096692	-2.587	0.00969 **
Rich	-0.468785	0.099406	-4.716	2.41×10^{-6}***

Table 4.3: Approximate significance of smooth terms

Co–variate	edf	Chi.sq	p-value
s(Child's age)	6.296	136.49	$< 2 \times 10^{-16}$ ***
s(Mother's age)	2.627	15.03	0.00235 **
s(Mother's BMI)	6.296	136.49	5.36×10^{-16} ***

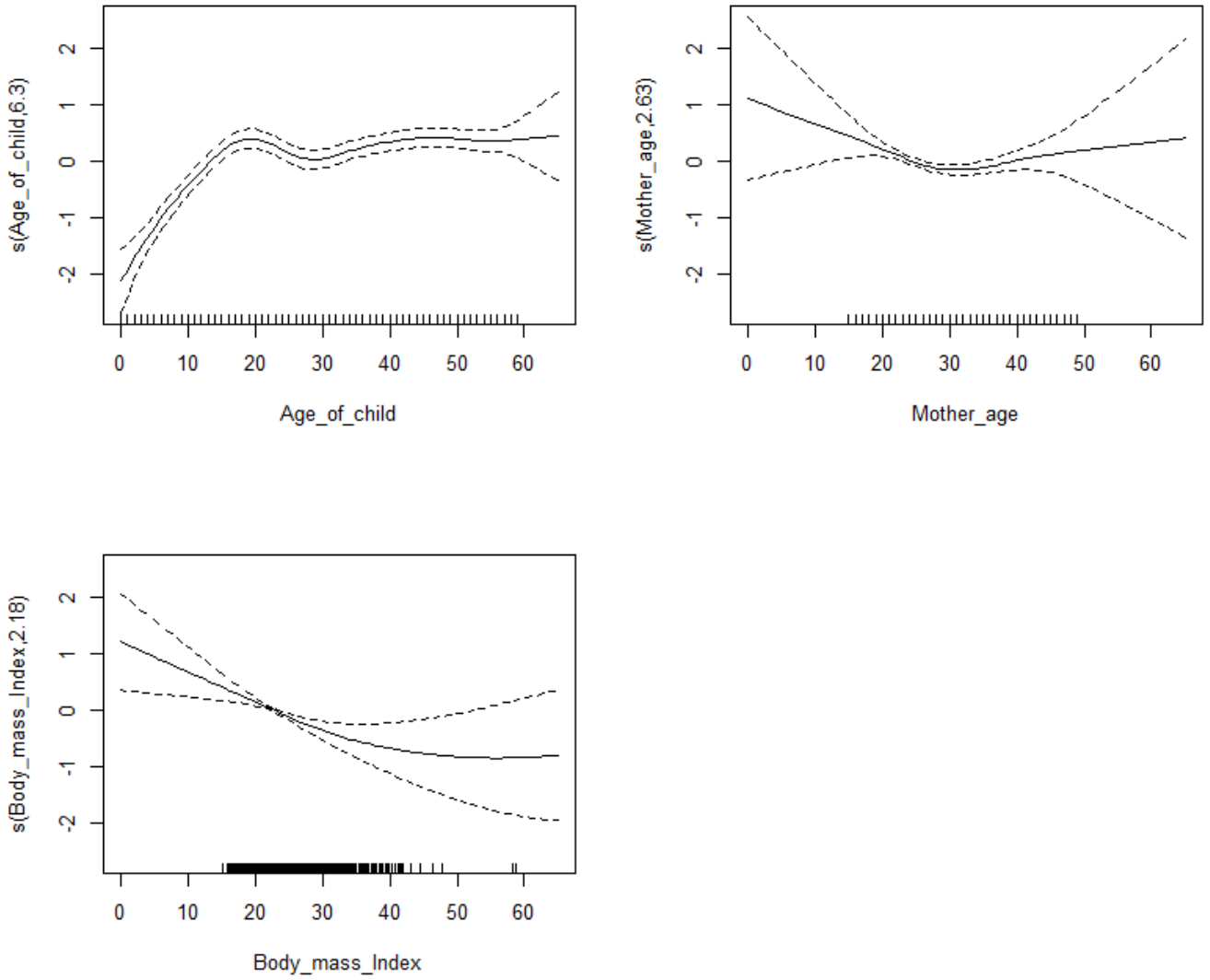

Figure 4.2: Estimarted smoother (solid line) for child's age, mother's age and BMI along with 95% point-wise confidence interval

Chapter 5

Quantile regression models

5.1 Introduction

In the preceding chapters, we discussed generalized linear models (GLM) and generalized additive mixed models (GAMM) to assess the risk factors of malnutrition in under-five children in Malawi. However, we realize that all these techniques summarize the average relationship between a set of regressors and the outcome variable based on the conditional mean function $E(y|x)$. This provides only a partial view of the relationship, especially when we are interested in reporting the relationship at different points in the conditional distribution of y. Analogous to the conditional mean function of linear regression, we consider the relationship between the regressors and outcome, using the conditional median function, where the median is the 50^{th} percentile, or quantile q, of the empirical distribution. Quantile regression provides that capability.

In this chapter, we discuss quantile regression that allows for studying the impact of predictors on different desired quantiles of the response distribution, and thus it provides a complete picture of the relationship between the response and predictor variables. Quantile regression is desired if conditional quantile functions are of interest. One advantage of quantile regression, relative to the ordinary least squares regression, is that the quantile regression estimates are more robust against outliers in the response measurements. In addition, the main attraction of quantile regression is the flexibility of the model, in the sense that it does not involve a link function that the variance and the mean of the response variable. It is also a robust method in the sense that it makes no assumption about the distribution of the error terms in the model.

The current study uses quantile regression models to characterize the impact of se-

lected independent variables on the whole distribution of the nutritional status of under-five children in Malawi. We fit structured additive quantile regression model and we use the geographical data from Malawi DHS data to draw a map using the spatial effects via integrated nested Laplace approximation (INLA).

5.2 Model formulation

The quantile regression model is defined in Koenker and Basett (1978) [47] as

$$y_i = x_i'\beta_0 + u_{q_i} \tag{5.1}$$

with

$$Q_q(y_i|x_i) = x_i'\beta_0 \tag{5.2}$$

and

$$Q_q(u_{q_i}|x_i) = F_u^{-1}(q|x_i) = 0 \tag{5.3}$$

where y_i is the i^{th} observation of the outcome variable, X_i is a vector of predictor variables, β_0 is a vector of unknown regression parameters and u_{q_i} are independent identically distributed error terms with unspecified distribution; the quantile $Q_q(y_i|x_i)$ and $Q_q(u_{q_i}|x_i)$ mean the q^{th} conditional quantile (percentile) of y_i and u_{q_i}, given x_i, respectively.
The q^{th} sample quantile is given by $Q_Y(q) = \xi_\tau, 0 \leq q \leq 1$, of a random variable Y is the inverse of the cumulative distribution function written as $F_Y(y) = q$ defined as

$$Q_Y(q) = F_Y^{-1}(q) = inf\{y : F_Y(y) \geq q\}. \tag{5.4}$$

5.3 Structured Additive Quantile Regression

Structured additive regression models are a flexible and extensive used class of models, see for example Fahrmeir and Tutz (2001) [48] for a detailed account. In these models, the observation (or response) variable y_i is assumed to belong to an exponential family, where the mean μ_i is linked to a structured additive predictor η_i through a link-function $g(\cdot)$, so that $g(\mu_i) = \eta_i$, The structured additive predictor η_i accounts for effects of various covariates in an additive way:

$$\begin{aligned} y_j &= \eta_j + \epsilon_j,\ \epsilon_j \sim F(\epsilon_j|\theta),\ j \in J \\ \eta_i &= x_j^T + \sum_{k=j}^{n_f} g_k(u_{ki}) + b_i, i = 1, \cdots, n \end{aligned} \tag{5.5}$$

where y_j denotes the j^{th} observation and η_j is its predictor; J is its predictor of $\{1, \cdots, n\}$, that is, not necessarily all latent η_j are observed through the data y_i; F is an unknown distribution function of error ϵ_i and may depend on some additional parameters $\boldsymbol{\theta}$; $\boldsymbol{x_i}$ is a vector of n_β covariates with a linear effect and $\boldsymbol{\beta}$ is the corresponding unknown vector; b_i are unstructured random effects; $g_k, k = 0, \cdots, n_g$, comprise usual nonlinear effect of continuous covariates. This is a structured addictive regression model (STAR) In some application we might have a situation where upper or other quantiles of y_i may depend on the covariates quite differently from the center. Hence, inference on conditional quantiles can provide a more complete description of functional changes than focusing sorely on the mean.The STAR models, unfortunately, only estimate conditional mean. However, to overcome this limitation, we extend the STAR models to quantile regression context [49].

Quantile regression, a completely distribution free approach, has emerged as a useful supplement to ordinary mean regression [49]. Given a fixed and known quantile $q \in (0, 1)$, assuming that the q^{th} quantile of the error distribution in (5.5) is zero, i.e., $F^{-1}(q|\boldsymbol{\theta}) = 0$. Then the corresponding quantile function of the continuous response variable Y_i is

$$Q_Y(q|\boldsymbol{x_i}, \boldsymbol{u_i}) = \eta_{qi} = \boldsymbol{x_i^T}\boldsymbol{\beta_q} + \sum_{k=1}^{n_g} g_{qk}(u_{ki}) + b_{qi} \tag{5.6}$$

where predictor η_{qi} is composed of linear effect β_q, a sum of nonlinear effects g_{qk} for both univariate and bivariate covariates, and an unstructured random effect $\boldsymbol{b_q}$.
An alternative representation of model (5.6) can be achieved via the minimization problem:

$$\arg\min_{\eta_q} \sum_{i=1}^{n} \rho_q(y_i - \eta_{qi}), \quad \text{where} \quad \rho_q(u) = \begin{cases} nq & n \geq 0 \\ u(q-1) & u < 0 \end{cases}, \tag{5.7}$$

this equation (5.7) is called, 'check function' [47] and η_{qi} has the same structure as (5.6). Thus the check function is the appropriate loss function for quantile regression problems regarded from a decision theoretical point of view [49]. In the case of simple linear quantile regression, Equation (5.7) can be estimated using a linearization problem [50] , but in the case of the non-linear case such as the structured additive quantile model, the estimation requires the Bayesian inferential approach. The method is based on assuming Y_i have iid (identically independent distributed) asymmetric Laplace distributions and taking appropriate Gaussian type priors on the additive components in the predictor. Markov chain Monte Carlo (MCMC) and integrated nested Laplace approximation (INLA) are two different tools proposed for such Bayesian quantile inference. This study will use the INLA method because

of the quicker convergence on its solutions when compared to MCMC in case of quantile regression models [51].
In the Bayesian analysis framework, the unknown functions for the metric, the spatial effects, regression coefficients for categorical covariates and all variance parameters have to be assigned suitable prior distributions [52]. The next sections are arranged in the following order. In Section 5.4 we discuss necessary information about asymmetric Laplace distribution on which we would be used as likelihood in our Bayesian quantile regression models; Section 5.5 specifies all the priors assigned for the latent parameters in model (5.5); Then we discuss the INLA inference approach in Section 5.6. Application of structured quantile regression models are presented in Section 5.7.

5.4 Asymmetric Laplace distribution

Bayesian inference requires likelihood. We need an assumption on data distribution for Bayesian quantile inference because the classical quantile regression has no such restriction. A possible parametric link between the minimization problem (5.7) and the maximum likelihood theory is the asymmetric Laplace density. We say that a random variable Y has asymmetric Laplace distribution with parameters $\eta.\delta_0$ and q, and write it as $Y \sim ALD(\eta, \delta_0, q)$, if the corresponding density function is given by [49]

$$[y|\eta, \delta_0, q] = q(1-q)\delta_0 \exp\{-\delta_0 \rho_q(y-\eta)\}, \tag{5.8}$$

where $\eta \in \Re$ is the location parameter, $\delta_0 \in \Re^+$ is the scale parameter, and $0 < q < 1$ is the skewness parameter. When assuming that $y_i \sim iddALD(\eta_i, \delta_0, q)$, then the likelihood for n independent observation is proportional to

$$L(\boldsymbol{\eta}, \delta_0; \boldsymbol{y}, \boldsymbol{q}) \propto \delta_0^{n/2} \exp\left\{-\delta_0 \sum_{i=1}^{n} \rho_q(y_i - \eta_i)\right\} \tag{5.9}$$

now we see that the maximization of the likelihood in (5.9) with respect to η given δ_0 is equivalent to the minimization of the loss function in (5.7). Thus, the asymmetric Laplace distribution is useful in terms of unifying the likelihood and the inference for quantile regression estimation.

5.5 Prior distributions

5.5.1 Smoothness priors for functions

Let $\boldsymbol{g} = (g(u_1), \cdots, g(u_n))^T$, a random vector of the response at $u_i, i = 1, \cdots, n$. $\boldsymbol{g}$ is said to be a Gaussian Markov random fields (GMRF) with mean $\boldsymbol{\mu}$ and precision (the inverse covariance) matrix $\delta \boldsymbol{R}$ if and only if it has density of form

$$\pi(\boldsymbol{g}|\delta) \propto \delta^{(n-m)/2} \exp\left(-\frac{\delta}{2}(\boldsymbol{g}-\boldsymbol{\mu})'\boldsymbol{R}(\boldsymbol{g}-\boldsymbol{\mu})\right), \tag{5.10}$$

where $\boldsymbol{R}$ is a semidefinite matrix of constants with rank $n - m(m \geq 0)$. The properties of a particular GMRF are all reflected through matrix $\boldsymbol{R}$. For instance, the Markov properties of GMRFs totally depend on the various sparse structure that the matrix $\boldsymbol{R}$ may have. The current study will use two kinds of GMRFs: continuous random walk (CRW) models (detailed in Wecker and Ansley, 1983 [53]) for metrical covariates and intrinsic autoregressive models (see Besag and Kooperberg, 1995 [54]) for spatial covariates. The two GMRFs share the form (5.10) but with different structures of $\boldsymbol{R}$.

Metrical covariates

Let $u_1 < u_2 < \cdots < u_n$ be the set of continuous locations and $z_i = g(u_i)$ be the function evolutions at u_i for $i = 1, \cdots, n$. The construction of CRW model is based on a discretely observed continuous time process $z(u)$ that is a realization of an $m-1$ fold integrated Wiener process given by

$$z(u) = \int_0^u \frac{(u-h)^{m-1}}{(m-1)!} dW(h), \tag{5.11}$$

where W(h) is a standard Wiener process. One can show that density of $z(u)$ is Gaussian with zero mean and completely dense matrix $\boldsymbol{R}$. Since factorizing an $n \times n$ dense matrix is at cost $\mathcal{O}(n^3)$, such matrix $\boldsymbol{R}$ would make Bayesian computation cumbersome with a huge dataset [49].

Spartial covariates

We now look at the spatial covariate u, where the values of u represent the location or site in connected geographical region. A common way to deal with special covariates is based on a set of predefined neighbors for each u_i. For geographical data as considered in the current study we assume that two sites u_i and u_j are neighbors if they share common boundary. Letting n_i denote number of neighbors of

site u_i, we assume the following spatial smoothness prior for function evaluations $g(u_i), i = 1, \cdots, n$;

$$g(u_i)|\{g(u_j)j \neq i\}, \delta \sim N(\frac{1}{n_i}\sum_{j:j\sim i} g(u_j), \frac{1}{n_i\delta}), \tag{5.12}$$

where $j \sim i$ denotes that site u_i and u_j are neighbors. Thus the conditional mean of $g(u_i)$ is unweighted average of evaluations of neighbors sites. The joint density of $\boldsymbol{g}$ in (5.12) has the same expression as in (5.10) with mean zero and matrix $\boldsymbol{R}$ such that

$$R_{ij} = \begin{cases} n_i & i = j, \\ -1 & i \sim j, \\ 0 & \text{otherwise.} \end{cases} \tag{5.13}$$

It is easy to see that the sparse matrix $\boldsymbol{R}$ has rank $n - 1$.
Let $g_k, k = 1, \cdots, n_f$, be a vector of kth function component in (5.5). Since they are of the same form, we assume the priors on functions $\boldsymbol{g_k}$ are

$$\pi(\boldsymbol{g_k}|\delta_k) \propto \delta_k^{(n-m_k)/2} \exp\left(-\frac{\delta_k}{2}\boldsymbol{g_k'Q_kg_k}\right), \tag{5.14}$$

where $\boldsymbol{Q_k}$ has rank $n - m_k$ and its structure depends on the type of covariates.

5.6 Posterior inference by integrated nested Laplace approximations (INLA)

The current section present the INLA approach for approximating the posterior marginals of the latent Gaussian field $\pi(x_i|\boldsymbol{y}), i = 1, \cdots, n$. Integrated nested Laplace approximations (INLA) is a new approach to statistical inference for latent Gaussian models introduced by Rue and Martino (2007) and Rue et al. (2009) [51, 55]. The main advantage of the INLA approach over MCMC is that it is much faster to compute; it gives answers in minutes and seconds where MCMC requires hours and days [49]. The approximation is computed in the three following steps:

5.6.1 Exploring $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$

The first step of the INLA approach is to compute our approximation to the posterior marginal of $\boldsymbol{\theta}$:

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^\star(\boldsymbol{\theta})} \tag{5.15}$$

where $\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation to the full conditional of $\boldsymbol{x}$, and $\boldsymbol{x}^\star(\boldsymbol{\theta})$ is the mode of the full conditional for $\boldsymbol{x}$, for a given $\boldsymbol{\theta}$. The main use of $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is to integrate out the uncertainty with respect to $\boldsymbol{\theta}$ when approximating the posterior marginal of x_i (see equation (5) of Rue and Martino, 2009). For this task, there is no need to represent $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ parametrically, but rather to explore it sufficiently well to be able to select good evaluation points for the numerical integration.

Step 2 Locate the mode of $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$, by optimizing $\log \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ with respect to θ. This can be achieved using quasi-Newton method which builds up an approximation to the second derivative of $\log \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ using the difference successive gradient vectors. The gradient is approximated using finite differences. Let $\boldsymbol{\theta}^\star$ be a modal configuration.

Step 3 At the mode configuration $\theta^\star$ compute the negative Hessian matrix $\mathbf{H} > 0$, using finite differences. Let $\boldsymbol{\Sigma} = \mathbf{H}^{-1}$, which would be the covariance matrix for $\boldsymbol{\theta}$ if the density were Gaussian. To aid the exploration, use standardized variable $\mathbf{z}$ instead of $\boldsymbol{\theta}$. Let $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\boldsymbol{T}}$ be the eigen-decomposition of $\boldsymbol{\Sigma}$, and define $\boldsymbol{\theta}$ via $\boldsymbol{z}$, as follows

$$\boldsymbol{\theta}(\boldsymbol{z}) = \boldsymbol{\theta}^\star + \boldsymbol{V}\boldsymbol{\Lambda}^{\mathbf{1/2}}\boldsymbol{z}$$

If $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is a Gaussian density, then $\boldsymbol{z}$ is $\boldsymbol{\aleph(0, I)}$. This parametrization corrects for scale and rotation, and simplifies numerical integration [55].

5.6.2 Approximating $\pi(\boldsymbol{x_i}|\boldsymbol{\theta}, \boldsymbol{y})$

Using Gaussian Approximations

The simplest (and cheapest) approximation to $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation $\tilde{\pi}_G(x_i|\theta, y)$, where the mean $\mu_i(\boldsymbol{\theta})$ are derived using the recursions (7), and possibly correcting for linear constraints. During the exploration of $\tilde{\pi}(\theta|y)$ in Section (5.6.1), we already compute $\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$, so only marginal variances need to be additionally computed. The Gaussian approximation gives often reasonable results, but there can be errors in the location and or errors due to the lack of skewness [55].

Using Laplace approximations

The natural way to improve the Gaussian approximation is to compute the Laplace approximation [55]

$$\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}, \boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{\pi}_{GG}(\boldsymbol{x_{-i}}|x_i, \boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x_i} = \boldsymbol{x}^\star_{\boldsymbol{-i}}(x_i, \boldsymbol{\theta})} \tag{5.16}$$

where $\tilde{\pi}_{GG}$ is the Gaussian approximation to $\boldsymbol{x}_i|x_{-i}, \boldsymbol{\theta}, \boldsymbol{y}$ and $\boldsymbol{x_{-i}}(x_i, \theta)$ is the model configuration. Moreover, we note that $\tilde{\pi}_{GG}$ is different from the conditional density corresponding to $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$.
Sadly, (5.16) implies that $\tilde{\pi}_{GG}$ must be recomputed for each value of x_i and $\boldsymbol{\theta}$, since its precision matrix depends on x_i and θ. This is far too costly, as it requires n factorization of the full precision matrix. Rue and Martino, (2009) [55] proposed two modifications to (5.16) to make it computationally feasible.
The first modification consists of avoiding the optimization step in computing $\tilde{\pi}_{GG}(\boldsymbol{x_{-i}}|\boldsymbol{x_i}, \boldsymbol{\theta}, \boldsymbol{y})$ by approximating the model configuration,

$$x_i^\star(x_i, \boldsymbol{\theta}) \approx E_{\tilde{\pi}_G}(\boldsymbol{x_i}|x_i). \tag{5.17}$$

The right-hand side is evaluated under the conditional density derived from the Gaussian approximation $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$. The computational benefits is immediate. First, the conditional mean can be computed by rank-one update from the unconditional mean, (using Equation (8) of Rue and Martino, 2009). This rank-one update is computed only once for each i, as it is linear in x_i. Although their settings are slightly different, Hsiao et al. (2004) [56] show that deviating from the conditional mode does not necessarily degrade the approximation error.

The next modification materializes the following intuition: only those x_j that are 'close' to x_i should have an impact on the marginal of x_i. If the dependency between x_j and decays as the distance between nodes i and j increases, only those x_j's in a 'region of interest' around $i, \boldsymbol{R_i}(\boldsymbol{\theta})$, determine the marginal of x_i, The conditional expectation in (5.17) implies that

$$\frac{\boldsymbol{E}_{\tilde{\pi}_G}(x_j|x_i) - \mu_j(\boldsymbol{\theta})}{\sigma_j(\boldsymbol{\theta})} = a_{ij}(\theta)\frac{x_i - \mu_i(\boldsymbol{\theta})}{\sigma_j(\boldsymbol{\theta})} \tag{5.18}$$

for some $a_{ij}(\theta)$ when $j \neq i$. Hence, a simple rule for constructing the set $R_i(\boldsymbol{\theta})$ is

$$R_i(\boldsymbol{\theta}) = \{j : |a_{ij}(\boldsymbol{\theta})| > 0.001\}, \tag{5.19}$$

The most important computational saving using $R_i(\boldsymbol{\theta})$ comes from the calculation of the denominator of (5.16), where we now only need to factorize a $|R_i(\boldsymbol{\theta})| \times |R_i(\boldsymbol{\theta})|$ spare matrix.

5.7 Application of quantile regression models to risk factors of malnutrition

To investigate risk factors of childhood malnutrition we consider determinants of children's heights in Malawi. The data comes originally from the Demographic and Health Surveys (DHS) conducted regularly in more than 75 countries. We use the dataset for children in Malawi from our sample size of 4 244. All children in the sample are between the ages of 0 and 5 years (0–59 months). We consider three covariates entering as additive nonparametric effects in addition to the response variable height-for-age: age of a child, mother's BMI and her age. Summary statistics for these variables are presented in Table (2.2). There are also a number of discrete variables that enter the model as parametric effects: gender, birth type, child having suffered from diarrhoea, mother's education level, father's education level, household wealth index and type of residence. These variables are also summarized in Table (2.2). Lastly, we also have one spatial covariates which we use for mapping the effect of malnutrition in the districts of Malawi.

5.7.1 Model fit

In order to identify the factors associated with childhood malnutrition in Malawi, the current study attempted to do so by fitting the following models;

$$\text{Model 1} : n_{qi} = x_i'\beta_q$$

Model 1 is known as a linear quantile regression model, where x_i present all categorical variables and their effects is summarized in β_q.

$$\begin{aligned}\text{Model 2} : n_{qi} &= x_i'\beta_q + g_{q1}(\text{Child's age}_i) + g_{q2}(\text{Mother's age}_i) \\ &+ g_{q3}(\text{Mother's BMI}_i)\end{aligned}$$

We note that Model 2 combines the fixed effects and non-linear effects, x_i and β_q are the same as in Model 1. Now Model 2 is known as additive quantile regression.

$$\text{Model 3} : n_{qi} = x_i'\beta_q + g_q(\text{District}_d),\ d = 1, 2, \cdots, 28$$

Model 3 combines the fixed effects and spatial effects. Districts considered as a spatial covariate and this model is known as spatial quantile regression.

$$\begin{aligned}\text{Model 4} : n_{qi} &= x_i'\beta_q + g_{q1}(\text{Child's age}_i) + g_{q2}(\text{Mother's age}_i) \\ &+ g_{q3}(\text{Mother's BMI}_i) + g_{qd}(\text{District}_d),\ d = 1, 2, \cdots, 28\end{aligned}$$

Model 4 upgrades Model 2 to include structured spatial effects, and therefore, Model 4 is known as a structured spatial additive quantile regression.

R (version 3.5.1) statistical software packages were used to fit all the above quantile regression models. The statistical inference was fully Bayesian using the "INLA" approach for quantile regression in R. For brevity, only three quantiles (q = 0.20, 0.40 and 0.50) were assessed. The choice of the quantiles to assess was motivated by the cut-points of childhood stunting (severe, moderate and nourished) based on WHO standards such that $q = 0.2$ is equivalent to severe (haz < -3), $q = 0.4$ is equivalent to moderate (-3 $\leq$ haz < -2) and $q = 0.5$ quantile is equivalent to haz = -1.88 which is within the range of normal adjusted childhood height-for-age, according to the WHO standards.

5.7.2 Model Selection

In order to select the best model fit, the method Deviance Information Criterion (DIC) for comparing models was imposed, which is related to other information criteria and has an approximate decision theoretic justification. The summary of DICs for all models that were fitted in this study is presented in Table 5.1. From Table 5.1, D stands for "Deviance evaluated at the posterior mean", pD stands for "Effective number of parameters", and DIC stands for "Deviance Information Criterion". The mathematical relationship used to compute DIC is given by DIC = D + 2p, where D is minus twice the log likelihood (-2LL). The rule of thumb is that the smaller DIC values correspond to better model fit, that is, the model with a smaller DIC is better than a model with a larger DIC [57].
In a comparison of the DICs from Table (5.1), we observe that Model 4 corresponds to the smallest DIC. This observation implies that the final model to be considered in the analysis of this study is going to be Model 4, the called **structured additive quantile regression model**.

Table 5.1: Model comparison based on Deviance Information Criteria (DIC).

Statistics	Model 1	Model 2	Model 3	Model 4
D	36515.96	36119.26	36507.35	36101.08
pD	13.02	49.87	17.60	57.96
DIC	36542.54	36219.78	36542.55	36217.69

Using the best fitting Model 4, we analyzed and assessed the fixed effects, non-linear effects, and structured spatial effects.s.

5.7.3 Fixed effects

The summary of fixed effects on children stunting is displayed in Table 5.2. Stunting and adjusted height-for-age are negatively associated variables; hence, care was taken in interpreting the effects on childhood stunting. From the results presented in Table 5.2, children from multiple births showed a significant negative effect on height-for-age at all observed quantile levels, implying that this had a significant positive effect on severe, moderate and median childhood stunting. In simple words, stunting among children in Malawi is more attributable to children of multiple births than to singletons. The observations are consistent with what was observed in the previous chapters of this study when mean models were fitted, that multiples births are at higher risk of stunting when compared to singletons.

Table 5.2: Summary of the fixed effects on childhood stunting in Malawi

Co-variates	Posterior Mean	Standard Deviation	0.2 quant	0.4 quant	0.5 quant
Intercept	4.4186	1.3888	3.2500	4.0675	4.4195
Type of birth [Singleton as reference]					
Multiple	-2.1723	1.3696	-3.3261	-2.5199	-2.1724
Mother's education level [No education as reference]					
Primary	-0.6440	0.7788	-1.3001	-0.8417	0.6441
Secondary	0.8609	0.9812	0.8609	1.1098	1.6875
Higher	0.4181	2.2178	0.4181	0.9808	2.2864
Father's education level [No education as reference]					
Primary	-0.8811	0.8626	-1.6078	-1.1000	0.8811
Secondary	1.7649	0.9698	1.7649	2.0109	2.5818
Higher	2.7187	1.6160	2.7188	3.1288	4.0800
Type of residence [Urban as reference]					
Rural	-0.9259	0.7309	-1.5416	-1.1114	-0.9259
Had diarrhea [No as reference]					
Yes	-0.1157	0.6074	-0.6275	-0.2699	-0.1158
Household wealth index [Poor as reference]					
Middle	0.6689	0.6327	0.6690	0.8295	1.2019
Rich	0.8423	0.6238	0.8423	1.0005	1.3678

The household wealth index of the family is observed to have a significant positive effect on adjusted height-for-age for children in Malawi. In general, the adjusted height-for-age of under-five children in Malawi increases with the betterment of the wealth index of the family. Furthermore, the effect (coefficients) increases with in-

creasing the quantile level. This result implies that a child born to a family with a rich and middle wealth index is less likely to be stunted compared with a child born to a poor family.

Diarrhoea is found to be a significant predictor of child stunting. The results in Table 5.2 show that a recent incidence of diarrhoea has a significant negative effect on adjusted height-for-age in all quintiles. In essence, the results show that the incidence of a fever in the two weeks prior to the survey has a negative effect on adjusted height-for-age across all the observed quantiles, where negative effect decreases with increasing the quantile. From this observation, we conclude therefore that a child who had diarrhoea in the two weeks prior to the survey is more likely to be stunted when compared with a child who did not have fever in the same time period.

The results of Table 5.2 show that the type of residence has a significant effect on childhood stunting. It was noted from the results that child adjusted height-for-age is negatively affected among those born to households in rural areas compared with adjusted height-for-age of those living in urban areas. This implies that children living in rural areas possess a higher risk of being stunted than children living in urban areas.

The results of Table 5.2 also reveal that the mother's and father's education level has a significant effect on childhood adjusted height-for-age. Moreover, we observe that child adjusted height-for-age increases with the increase in the level of education in both the mother and father of the child. Another notable observation is that the effect increases with increasing the quantile in both mother and father's education level variables. In addition, we note that children born to educated parents are less likely to be stunted than children born to uneducated parents.

5.7.4 Nonlinear Effects

Figure 5.1 shows the summary of observed nonlinear effects. The summary of nonlinear effects of child's age in months on childhood stunting was displayed on the top left corner in Figure 5.1. We observe that the adjusted height-for-age remained high, which implies low childhood stunting for about the first four months, after which it deteriorated, implying increasing childhood stunting until about 20 months. The adjusted height-for-age remained constantly low, implying a high risk of stunting from 20 months upward.

The summary of nonlinear effects of mother's age in years on children stunting is displayed along the top right of figure 5.1. We found that the general relationship of the effects of a mother's age with adjusted height-for-age followed a bell shape. Young mothers (below 18 years) and very old mothers (above 40 years) are associ-

ated with a lower height-for-age, implying an increase of childhood stunting. Moreover, mothers between 18 and 40 years are associated with high child height-for-age, implying children born of mothers between 18 and 40 are less likely to be stunted. Body mass index continued increasing the adjusted height-for-age (reducing the risk of stunting) beyond 40 kg/m^2. The results reveal that thin mothers are associated with low height-for-age, implying that their children are more likely to be stunted than those of normal weight mothers.

5.7.5 Spatial Effects

The spatial variation of childhood malnutrition (stunting) was examined. Figure 5.2 (map) presents the posterior means of structured spatial effects on average adjusted childhood height-for-age. Looking at the map, dark blue corresponds to the districts where there is a positive effect on adjusted height-for-age for children under the age of five and thus a low risk of childhood stunting. The light blue on the map corresponds to the districts where location of the child has a negative effect on adjusted height-for-age; hence, a high risk of childhood stunting.
The map shows that most districts in Malawi possess a higher risk of childhood stunting, except for 9 out of 28 districts that show a positive effect on child height-for-age, implying less likelihood of stunting. The findings derived from this map seem to be consistent with the fact that Malawi is one of the African regions with a high rate of malnutrition. This high rate of prevalence of childhood malnutrition in some districts of Malawi may be due to the poverty of household members.

5.8 Summary

Using the fitted quantile regression models, we concluded as follows: The fixed effects of multiple births, household wealth index, recent history of diarrhoea and parent's education level had significant effects on childhood stunting. The general relationship of the effects of a child's age with adjusted height-for-age showed that risk of stunting increases with the increase of the age of a child. The general relationship of the effects of a mother's age with adjusted height-for-age followed an inverse U-shape, while a mother's age and adjusted height-for-age showed a U-shape. In general, of the districts, only Chikwawa, Thyol, Mulanje, Chiradzulu, Blantyi, Mwanza, Zombi, Phalombe and Nkhotakota districts depicted positive structured spatial effects on adjusted height-for-age. Finally, the current chapter provided evidence that structured spatial additive quantile regression is more appropriate (smallest DIC).

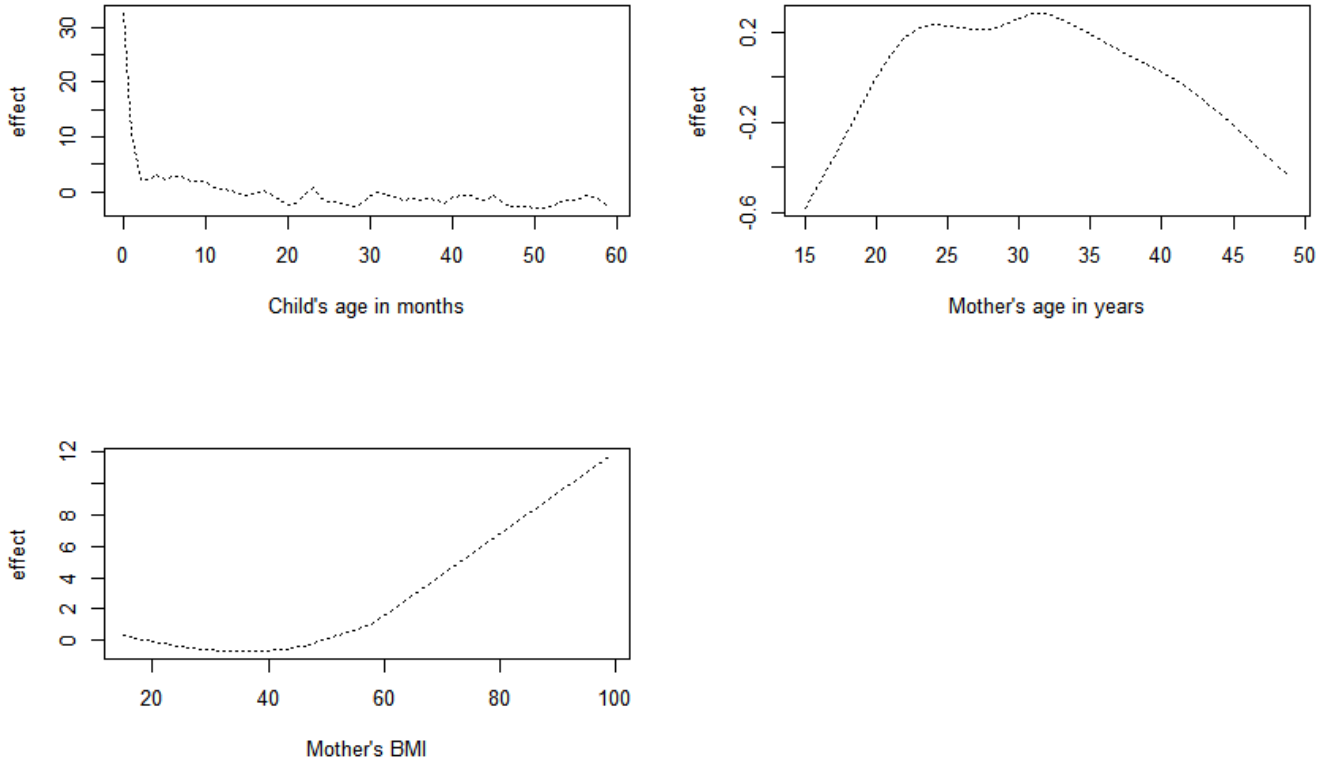

Figure 5.1: Nonlinear effects on height-for-age for under-five children in Malawi

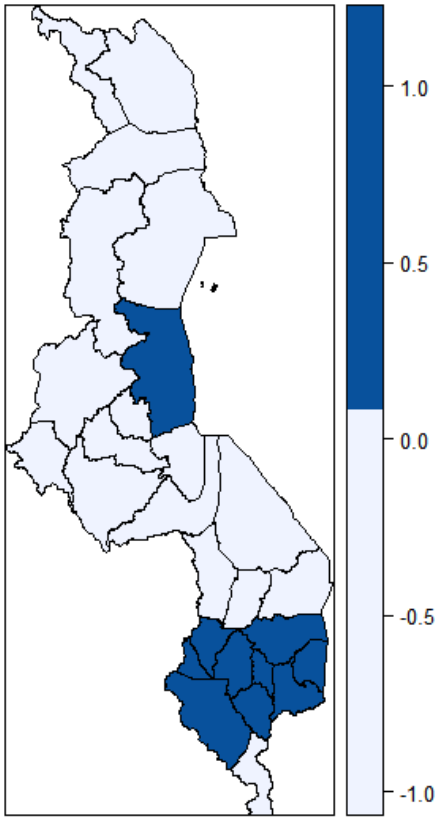

Figure 5.2: Structured spatial effects on childhood height-for-age where regions with dark blue colour depicts lower risk and light blue depicts higher risk of stunting

Chapter 6

Discusion and Conclusion

The study provides a broad overview on unfolding key determinants of childhood malnutrition by making use of different types of statistical models. The use of these models was to assess how each, from basic models like ordinary logistic regression models to more complex models, allows for more information to be considered, differs in their results and thereby examine the various statistical methods which are suitable to identify the risk factors associated with the prevalence of malnutrition in under-five children.
To achieve our objective, in Chapter 3, we fitted generalized linear models to get an overview of how the predictors are related to the mean value of the outcome variable. A cumulative odds model was utilized to simultaneously consider the effects of a set of independent variables across the possible consecutive cumulative splits on the data while still taking advantage of the order of our response categories.

The generalized linear models (GLMs) are parametric models, offering a strong tool for modelling the relationship between the response variable and predictors when their assumptions meet. However, these models may suffer from inflexibility in modelling complex relationships between the response and predictor variables in some applications. In order to relax the assumption of normality and linearity inherent in linear regression models, a generalized additive mixed model (GAMM) (semi parametric) was utilized, where the categorical covariates were modelled parametrically, while the continuous non parametrically after assessing the relationship with the response variable. In our case, it was also a vital addition, as it also came with an opportunity to add the random effects in the model. The use of GAMM revealed certain information that may have been hidden by the use of parametric models.

Most studies in literature used a binary response variable among others or a categorical ordered response variable; some of them were ordinary logistic regression or

spatial analysis [10-13]. All these studies have considered the mean average of the response variable or Gaussian distribution but it might be possible that the upper or lower quantiles of the outcome variables depend on the covariates differently from the mean. In addition, there may be a need to analyse it as a quantile of interest and this is not possible in all above mentioned models. An inference on conditional quantiles can allow the influence of explanatory variables to be assessed and give detailed information on quantiles of interests of the response variable rather than focusing solely on the mean.

Therefore, the current study used structured additive quantile regression that provides the flexibility to analyse the impact of predictor variables on the different quantiles of interest of childhood anthropometric index (stunting) instead of the mean distribution. In addition, this model allows the possible nonlinearity effects of continuous covariates to be accounted for, the possible structured spatial effects on childhood stunting to be captured and possible heterogeneity among the variables to be incorporated.

Upon the utilization of the above mentioned variety of statistical models, the current study revealed that the key determinants of malnutrition of children under five years of age in Malawi are a child's age in months, birth weight, recent incidence of fever and diarrhoea, a mother's education level, a father's education level, a mother's age at the birth, a mother's body mass index, districts, source of drinking water, multiple births and wealth index of the household. In addition, a spatial quantile additive model was fitted and produced a map of Malawi, using an R-INLA package showing the prevalence of stunting in the districts of Malawi.
These results are almost consistent with results from a study done by Chirwa (2008) which was also conducted in Malawi, using a multivariate analysis to investigate factors that determine child malnutrition. Three anthropometric measures of malnutrition, WAZ for underweight, HAZ for stunting and WHZ for wasting were used. Chirwa (2008) discovered that stunting is the greater of the malnutrition problems. The age of the child, sex of the child, regular sickness, source of water, education level of household head, household land size and ability to produce own food were all factors associated with child nutritional status derived from multivariate analysis [4].
Number of stakeholders in Malawi are attempting to come up with effective different strategies to be put in place in a hopes of achieving low level of childhood malnutrition. However, the general theme of the strategies put in place seem to be promising in making sure that the average level of stunting is decreasing in the country and these strategies aimed at preventing the disease (malnutrition), How-

ever, in my opinion, significant approach is to know the underlying factors that are escalating the occurrence of the childhood malnutrition. It is therefore hoped that the results from this study will help to emphasize the need for policy makers and other public health related institutions to properly understand the determinants of childhood stunting in Malawi and visualize the spatial components of childhood malnutrition at a district level. Upon understanding all the significant underlying factors of stunting, they will be in a good position to improve the current strategies to reduce its prevalence.

The prevalence of stunting in under-five children is very high in Malawi (37.1% on average), which is above the regional average in Africa (33.6%), according to the joint child malnutrition estimates by the United Nations Children's Fund, the World Health Organization, and the World Bank Group [5]. According to the UNICEF global site (2018), in Malawi, 4% of children, especially those below five years of age, suffer from acute malnutrition, and more than half of Malawian children suffer from chronic malnutrition, resulting in stunting (being too short for one's age). These figures therefore imply that Malawi is one of the countries with a high malnutrition incidence in eastern and southern Africa. There are several causes that have been attributed to this such as national and household food insecurity, heavy workloads and poor nutrition of mothers, frequent infections, poor eating habits, and HIV infections leading to repeated illness [4]. Malnutrition is devastating and the single biggest contributor to child death. In Malawi, unfortunately there has been only a small change in children's nutritional status (stunting) since 1992 (48.7%) and stunting rates remain unacceptably high thus far (37.1%), according to the WHO.

In light of the articulation of the above statistics, it is enough to call for serious structural solutions from the health authorities in Malawi, i.e. put more emphasis on projects that will come up with robust and reliable solutions to reduce and prevent the prevalence of childhood malnutrition. This may include but not limited to improving the education level of women with respect to the manner at which they should raise their children and improving their healthy life style during pregnancy. All possible initiatives should be considered in very extensive and enthusiastic manner that will drive the success with respect to the children's wellness. Over and above everything, it is very imperative that these solutions are implemented in a collaborative regime between the public and private stakeholders to ensure efficiency and availability of the resources.

Study limitations

First and foremost, the sample size of the study was limited to the variable with no missing information, the variables with missing data were ignored in the context of the current study, hence relatively small sample size was generated. Moreover, we also remain cognisant of the fact that in order to precisely measure the change of stunting in under-five children ,it is imperative that a longitudinal studies be done over a period of five years, this may help to precisely study the trends in time and track whether there's any improvement from the implementation of any intervention strategies put in place by the government and any other authorities in Malawi. However, due to relatively inevitable circumstances, the study used a cross-sectional date. In addition, the use of cross-sectional data appeared as a limitation considering the fact that this data cannot determine the causality, therefore a longitudinal study appear as better alternative that can relatively address this problem.

Future studies

One of the avenues for future research, would be to use quasi likelihood matching or a Mata-analysis from data collected in Malawi over a period of time but using different subject and different households. Given that the current study neglected the missing values, another method may be on how to improve the missing values by incorporating missing values analysis, one may choose to perform imputations based on multiple imputation via chained equations. The latter method may lead into a great precision on the parameter estimates.

References

[1] United Nations International Children's Emergency Fund (UNICEF). Monitoring the situation of children and women, 2018.

[2] Hayashi C, Krasevec J, Kumapley R, and Mehra V. Levels and trends in child malnutrition. *Joint Malnutrition Estimates*, 2017.

[3] United Nations International Children's Emergency Fund (UNICEF). Improving child nutrition: the achievable imperative for global progress, 2013.

[4] Chirwa EW and Ngalawa PE. Determinants of child malnutrition in Malawi. *South African Jounal of Economics*, 2008.

[5] United Nations Children's Emergency Fund, World Health Organization, and World Bank Group. Joint child malnutrition estimates - levels and trends in child malnutrition, 2018.

[6] World Health Organization. Global targets 2025, 2018.

[7] World Health Organization. Norway announces the establishment of a sustainable fisheries action network, 2017.

[8] FAO/WHO. International symposium on sustainable food systems for healthy diets and improved nutrition, 2017.

[9] World Health Organization. Decade of action on nutrition at the UN General Assembly (71st session), 2016.

[10] Habyarimana F, Zewotir T, and Ramroop S. A proportional odds model with complex sampling design to identify key determinants of malnutrition of children under five years in rwanda. *Mediterranean Journal of Social Sciences*, 2017.

[11] Talukder A. Factors associated with malnutrition among under-five children. *Children*, 2017.

[12] Das S. Application of ordinal logistic regression analysis in determining risk factors of child malnutrition in banladesh. *Nurition Journal*, 2011.

[13] Das S and Islam A. Predictors of child chronic malnutrition in bangladesh. *Proceeding of the Pakistan Academy of Sciences*, 2008.

[14] Kabubo-Mariara J, Ndenge GK, and Mwabu DK. Determinants of children's nutritional status in Kenya: Evidence from demographic and health surveys. *Journal of African Economies*, 2008.

[15] United Nations Children's Emergency Fund global site. Nutrition key result and indicator, 2018.

[16] Kleinbaum DG and Klein M. *Logistic Regression-A Self-Learning Text Second Edition*. 2002.

[17] Mohamed RS. Malnutrition, stunting and the importance of a child's first 1000 days. *The conversation Academic rigour journalistic flair*, 2015.

[18] Mosley WH and Chen LC. Analytical framework for the study of child survival in developing countries. *Population and development review*, 1984.

[19] J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 1972.

[20] Annette C and Dobson J. *An Introduction to Generalized Linear Models*. 2 edition, 2002.

[21] Agresti A. *An Introduction to Categorical Data Analysis*. 2 edition, 1996.

[22] O'Connell A. *Logistic Regression Models for Ordinal Response Variables*. 2006.

[23] Agresti A. *Categorical data analysis*. 1990.

[24] Sarkar D and Haldar S. Socioeconomic determinants of child malnutrition in India evidence from NFHS-III. *Epidemiology*, 2005.

[25] Peterson B and Harrell Jr FE. Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society*, 1990.

[26] Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 1990.

[27] Allison PD. Logistic regression using the SAS system: Theory and application, 1999.

[28] Armstrong BG and Sloan M. Ordinal regression models for epidemiologic data. *American Journal of Epidemiology*, 1989.

[29] Ananth CV and David G. Regression models for ordinary responses: A review of methods and applications. *International journal of epidemiology*, 1997.

[30] Bob Derr and SAS Institute Inc. Ordinal response modeling with the LOGISTIC procedure. *3 Statistics and Data Analys, SAS Global Forum 2013*, 2013.

[31] Bender R and Grouven U. Using binary logistic regression models for ordinal data with non-proportional odds. *J Clin Epidemiol*, 1998.

[32] Gameroff MJ. Using the proportional odds model for health-related outcomes: Why, when, and how with various SAS procedures. *Statistics and Data Analysis*, 2005.

[33] Adeleke KA and Adepoju AA. Ordinal logistic regression model: An application to pregnancy outcomes. *J Math and Stat 2010*, 2010.

[34] Habyarimana F. Measuring poverty and child malnutrition with their determinants from household survey data, 2016.

[35] Lin and Zhang. Inference in generalized additive mixed models by using smoothing splines. *J. R. Statistic Society B*, 2002.

[36] Hastie T and Tibshirani R. *Generalized Additive Model*. 1990.

[37] Härdle W and Kneip A. Testing a regression model when we have smooth alternatives in mind. *Board of the Foundation of the Scandinavian Journal of Statistics*, 26:221–238, 1999.

[38] Green P and Silverman B. *Nonparametric Regression and Generalized Linear Models; A roughness penalty approach*. 1999.

[39] Wahba G. *Spline Models for Observational Data*. 1990.

[40] Breslow N and Clayton D. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

[41] Kohn R, Ansley C, and Tharm D. The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. *Journal of applied statist. Ass.*, 86(416):1042–1050, 1991.

[42] Zhang D, Lin X, Raz J, and Sowers M. Semi-parametric stochastic mixed models for longitudinal data. *Journal of american statistical association*, (442):710–719, 1991.

[43] Harville DA. Bayesian inference for variance components using only error contrasts. *Biometrika*, (2):383–385,, 1974.

[44] Zuur A, Ieno E, and Elphick C. A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution*, 1(1).

[45] Zuur A, Saveliev AA, and Ieno E. *Beginner's Guide to GAMM with R*.

[46] Zuur A, Ieno E, and Hilbe JM. *Beginner's Guide to GLM and GLMM with R*.

[47] Koenker R and Bassett Jr. G. *Econometrica*, 46(1).

[48] Fahrmeir L and Tutz G. *Models for multicategorical responses*. Springer New York, 2001.

[49] Yu Y and Havard R. Bayesian inference for structured additive quantile regression models. 2009. [Online; accessed October-2018].

[50] Koenker R and D'Orey V. Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society*, 36(3), 1987.

[51] Rue H and Martino S. Bayesian inference for hierarchical gaussian markon random fields models. *s. J. Stat. Plan. Inference*, 137, 2007.

[52] Habyarimana F, Zewotir T, and Ramroop S. Structured additive quantile regression for assessing the determinants of childhood anemia in rwanda. *International Journal of Environmental Research and Public Health*, 2017.

[53] Wecker WE and Ansley CF. The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78(381), 1983.

[54] Besag J and Kooperberg C. On conditional and intrinsic autoregression. *JBiometrika*, 82(4), 1995.

[55] Rue H and Martino S. Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations. *The Norwegian University for Science and Technology, Trondheim, Norway*, 2009.

[56] Hsiao CK, Huang SY, and Chang CW. Bayesian marginal inference via candidate's formula. *Statistics and Computing*, 14(1), 2004.

[57] Spiegelhalter DJ, Best NG, Carlin BP, and van-der Linde A. Bayesian measures of model complexity and fit. *J. R. Statistics Society*, 64(4), 2002.