



Addis Ababa University  
Office of Graduate Program

Faculty of Science  
Department of Statistics

# Bayesian Approach to Identify Predictors of Children Nutritional Status in Ethiopia

By  
Tesfaye Mesele

A Thesis submitted to the Office of Graduate Programs of Addis Ababa  
University in Partial fulfillment of the requirement for the Degree  
of Master Science in Applied Statistics

*June, 2009*

Addis Ababa University  
Office of Graduate Program

Faculty of Science  
Department of Statistics

Bayesian Approach to Identify Predictors of Children Nutritional Status in Ethiopia

By  
Tesfaye Mesele

Approve by the Board of Examiners:

Fentaw Abegaz, (Ph.D)

Advisor

-----  
Signature

\_\_\_\_\_  
Internal Examiner

-----  
Signature

\_\_\_\_\_  
External Examiner

-----  
Signature

---

Tel. 251-1-239465  
Ababa, Ethiopia

P.O.Box 1176 Addis

June, 2009

## TABLE OF CONTENTS

Acknowledgment.....	iii
Abstract.....	iv
List of Abbreviations.....	v

<b>Table of Contents</b>	<b>Pages</b>
1. INTRODUCTION.....	1
1.1 Background.....	1
1.2 Statement of the problem.....	6
1.3. Objective of the study.....	6
1.4. Significance of the Study.....	7
1.5. Limitation of the study.....	7
2. LITRATURE REVIEW.....	9
2.1 General.....	9
2.2 Factors associated with Children’s Malnutrition.....	11
2.3.1 Household economic status.....	12
2.3.2 Education of Mother.....	13
2.3.3 Employment Status of Mothers.....	14
2.3.4 Source of Water and availability of Toilet Facility.....	14
2.3.5 Body Mass Index (BMI).....	15
2.3.6 Age of child.....	15
2.3.7 Sex of Child.....	16
2.3.8 Birth Order.....	16
2.3.9 Birth interval of the child.....	16
2.3.10 Environmental Factor.....	17
3. DATA AND METHODOLOGY.....	18
3.1 Data.....	18
3.1.1 Variables.....	20
3.1.1.1 The Response Variable.....	20
3.1.1.2 Explanatory Variables/factors.....	20
3.2 Methodology.....	22

3.2 Methodology.....	22
3.2.1 Bayesian structured additive regressions models based on MCMC techniques.....	24
3.2.2 Bayesian Linear regression (Parametric) Model.....	26
3.2.3. Bayesian semiparametric Model.....	27
3.2.4 Prior distribution.....	28
3.2.5 Posterior Inference.....	34
3.3 Model Selection.....	36
3.4 Software.....	38
4. ANALYSIS AND RESULT.....	40
4.1 Descriptive and Explanatory Analysis of Variables.....	40
4.1.1 Categorical Covariates.....	40
4.1.2 Continuous Covariates.....	45
4.2. Determinants of Children malnutrition using Bayesian structured Additive regression models.....	48
4.2.1 Bayesian linear regression (parametric) model.....	49
4.2.2 Bayesian linear regression (parametric) model with some transformation.....	51
4.2.3 Bayesian semiparametric regression model.....	51
4.3 Model Selection.....	53
4.4 Results of Bayesian semiparametric model.....	54
4.4.1. Fixed (linear) effects.....	54
4.4.2. Continuous (non linear) effects.....	56
5. DISCUSSION, CONCLUSION AND RECOMMONDATION.....	61
5.1 Discussion.....	61
5.2 Conclusions.....	65
5.3 Recommendation.....	66
REFERENCES.....	67
ANNEX.....	73
Annex A1: BayesX code for Bayesian Semi Parametric regression model.....	73
Annex: A2 MCMC SIMULATION.....	75
Annex A4: Posterior mean for fixed effects.....	81
Annex A5: Estimation results for the smoothing parameter:.....	82
Annex B1: Plots of autocorrelation functions for samples obtained with MCMC in BayesX.....	85
Sample Path in the MCMC simulation.....	86
Annex B2: Sample Path in the MCMC simulation for Mother's Body Mass Index.....	86
Annex B3: Sample Path in the MCMC simulation for Child Age.....	86
Annex B4: Sample Path in the MCMC simulation for Child Age.....	87



## Acknowledgment

I am really and ever grateful to the Almighty God for blessing me with His help and enabling me to achieve this milestone in my academic life.

I would like to express my gratitude to my advisor, Dr Fentaw Abegaz, for his unflinching and generous supervision as well as easy accessibility at all time. His constructive criticism and suggestions have helped me in widening my research abilities. In spite of his demanding schedule, he made himself available to me at any time for useful discussions, which helped in finalizing this work.

I am also grateful to Girmay Medhin, of the Institute of Aklilu lemma department of Pathobiology, for his insightful comments, helpful advice and his discussions which helped me very much in finding problem areas and guiding in accessing of research resources. To him I also say a *special thank you*.

My thanks go to Dejen Tesfaw, of Addis Ababa University Department of Statistics, for his insightful comments, helpful advice and his discussions which helped me very much in editing the data.

I am highly indebted to my sweetie Meklit, for her unreserved support and encouragement throughout my study.

Finally, I really would like to express my deep gratitude and appreciation to my mother, Mama your prayers are always a source of strength. She suffers so much for me. I am deeply indebted to her for her patience, concern and love. I cannot praise her enough for her support. I would also like to thank my father and my sister for their support. To them all I have dedicated this thesis.

## **Abstract**

Child Mortality and Undernutrition are the two major public health problems in the developing world. The associated phenomena are household living style, type of residence and awareness in health and family planning programs. Ethiopia, in particular, suffers from worst forms of malnutrition due to access to health care and nutrition. This thesis explores the most dominant socio economic, demographic and environmental factors in children nutritional status. We used recently developed Bayesian structured additive models to flexibly model the effects of included covariates on the EDHS 2005 datasets of Ethiopia. Inference is fully Bayesian based on recent Markov chain Monte Carlo techniques. These models allow us to analyze usual linear effects of categorical covariates and nonlinear effects of continuous covariates within a unified semiparametric Bayesian framework for modeling and inference. Most of the socioeconomic, demographic and environmental determinants included in the study were found to be statistically significant with the exception of covariates household economic status and mother education.

## List of Abbreviations

ACC/SCN	Administration Committee on Coordination–Sub-Committee on Nutrition
BMI	Body Mass Index
CDC	Centre for Disease Control and Prevention
DHS	Demographic and Health Survey
HKI	Helen Keller International
HSB	Health Seeking Behavior Index
MDG	Millennium Development Goals
MOPED	Ministry of Planning Economic Development
NCHS	National Centre for Health Statistics
NIPS	National Institute for Population Studies
EDHS	Ethiopian Demographic and Health Survey
U5M	Under five Mortality
UNDP	United Nations Development Program
UNICEF	United Nations Children Fund
USAID	United States Agency for International Development
WHO	World Health Organization

# CHAPTER ONE

## 1. INTRODUCTION

### 1.1 Background

Nutrition is one of the basic requirements for good health and is crucial for the attainment of the first goal of the Millennium Development Goals (MDG). However, progress on the MDGs over the last few years indicate that the world is not on track to halve the proportion of underweight children by 2015 a key indicator set in the MDG fight against poverty and hunger. UNICEF's, May 2006 Progress for Children reports that 27 per cent of children under five in the developing world are underweight. Some 5.6 million under-fives die of causes related to undernutrition each year.

Undernutrition may be defined as insufficient intake of energy and nutrients to meet an individual's needs to maintain good health. Additionally, it may indicate insufficient absorption of nutrients due to ill health. The term "malnutrition" is sometimes also used synonymously for undernutrition. However, strictly speaking, malnutrition includes both undernutrition as well as over-nutrition. Over nutrition simply refers to excess intake of macronutrients and micronutrients (Maleta, 2006). In the developing world, it is the undernutrition which is of greater concern because it is alleged to be one of the leading causes of morbidity in children and contributes to more than half of child deaths (Helen Keller International, 2001). Nutritional deprivation in early life can have long lasting effects on growth, educational attainment and productivity. Usually, the undernutrition in children (under age five) is used as a measure for determining the extent of this particular public health problem in a population (WHO, 1995; Kandala et. al., 2001).

Hunger and malnutrition are devastating problems, particularly for the poor and underprivileged. Child disease and malnutrition reflect a country's level of socio-economic developing and quality of life. According to the study by the Ethiopian Ministry of Economic Development and Cooperation, 50 percent of the Ethiopian population are living below the food poverty line and cannot meet their daily minimum nutritional requirement of 2200 calories (MOPED, 1999). The prevalence of stunting in children below five years in East Africa averages about 48 percent (ACC/SCN 2000), which is the highest in the world. Evidence also showed that the situation in Ethiopia is worse than in other East African countries.

In Ethiopia, as in many other developing countries, undernutrition is one of the leading causes of childhood morbidity and mortality. Childhood undernutrition affects physical and cognitive growth, impairs the immune system, and increases the risk of morbidity and mortality. In developing countries around the world, an estimated 148 million children are stunted, 127 million are underweight, and 46 million are wasted (UN, 2004). According to a recent comparative risk assessment by the World Health Organization, under-nutrition is estimated to be, by far, the largest contributor to the global burden of disease (WHO, 2002). Many people in developing countries still live in extreme poverty and in these countries economic growth tends to benefit only a small group of advantaged and affluent people and causes growing inequality in health and nutrition that affects particularly vulnerable groups of the population, such as children. Economic well-being at the household level operates mainly through availability of better food, more hygienic living conditions, and better access to health services in affecting the health and nutritional status of children (Mishra et. al, 1999).

While there are numerous studies on childhood malnutrition in Ethiopia and other countries majority of these studies have looked at the contributions of individual-level (socioeconomic and family planning) characteristics. A growing body of literature considers the importance of understanding of determinants of childhood malnutrition through an integrated analysis that considers linkages between demographic, household, and community structures. Thus, the contextual aspect of child malnutrition needs to be explored to understand the process of malnourishment as a whole. To expand our understanding of the most common and consistent factors on the risk of childhood malnutrition, the study argue that it is necessary to consider expected determinants for malnutrition using Bayesian approach which is not considered so far in such study. Therefore, the purpose of this paper is to develop and test a model of childhood malnutrition that includes all examined factors by previous researcher through Bayesian Semi Parametric regression model. The second aim is to determine whether there is significant variation in the factors on childhood malnutrition identified using other approaches.

The 2005 EDHS included information on the nutritional status of children less than five years of age for the three indices, namely, weight-for-age, height-for-age and weight-for-height, taking age and sex into consideration. Weight measurements were taken using a lightweight electronic SECA scale designed and manufactured under the guidance of UNICEF, and height measurements were carried out using a measuring board produced by Shorr Productions. Children younger than 24 months were measured lying down (recumbent length) on the board, while standing height was measured for older children. The scale allowed for the weighing of very young children through an automatic mother child adjustment that eliminated the mother's weight while she was standing on the scale with her baby.

As recommended by WHO, the anthropometric measurements of children in the survey were compared with an international reference population defined by the U.S. National Centre for Health Statistics (NCHS) and accepted by the U.S. Centers for Disease Control and Prevention (CDC). Each of the three nutritional status indicators described below are expressed in standard deviation units (Zscores) from the median of the reference population. The use of this reference population is based on the finding that well nourished young children in all population groups (for which data exist) follow very similar growth patterns. The reference populations are useful for comparison, facilitating the examination of differences in the anthropometric status of subgroups in a population and changes in nutritional status over time. In any large population, there is variation in height and weight; this variation approximates a normal distribution.

Each of these indices *height-for-age*, *weight-for-height*, and *weight-for-age* provides different information about growth and body composition, which is used to assess nutritional status. The height-for-age index is an indicator of linear growth retardation and cumulative growth deficits. The weight-for-height index measures body mass in relation to body length and describes current nutritional status. Weight-for-age is a composite index of height-for-age and weight-for-height. It takes into account both acute and chronic malnutrition.

Children whose height-for-age Z-score is below two standard deviations (2 SD) from the median of the reference population are considered short for their age (stunted) and are chronically malnourished. Children who are below three standard deviations (3 SD) from the median of the reference population are considered severely stunted. Stunting reflects failure to receive adequate nutrition over a long period of time and is also affected by recurrent and chronic illness. Height-for-age, therefore, represents the

long-term effects of malnutrition in a population and does not vary according to recent dietary intake.

The number of biological and medical datasets have been growing exponentially over the past several years. Analysis that systemically incorporates prior information is becoming essential to making inferences about the numerous, complex data. A Bayesian approach can help to capture such information and incorporate it seamlessly through a rigorous, probabilistic framework.

The statistical analysis in this thesis is based on modern Bayesian approaches which allow a flexible framework for realistically complex models. These models allow us to analyze usual linear effects of categorical covariates and nonlinear effects of continuous covariates within a unified semi-parametric Bayesian framework for modeling and inference. There has been much recent interest in Bayesian inference for generalized additive and related models. The increasing popularity of Bayesian methods for these and other model classes is mainly caused by the introduction of Markov Chain Monte Carlo (MCMC) simulation techniques which allow realistic modeling of complex problems.

There exists no widely used statistical software for Bayesian inference that compares to commercial products such as SAS, SPSS, or Stata (This is likely due to the fact that all of these packages were developed before MCMC methods were widely used). Over the last decade a number of promising software packages have emerged among which WinBUGS and BayesX are the most basic ones. All computations to implement the methodology discussed here are carried out with BayesX program-version 2.0 (Belitz et.al, 2009), R and STATA 10 also applied to display some of nonparametric outputs and we used SPSS 15 vastly in editing and managing the data.

## **1.2 Statement of the problem**

Many demographers and scholars believe and recommend the need to conduct in-depth studies on the various aspects of children's nutritional status in different demographic, economic and socio-cultural settings. The researcher shares the idea and the main reason behind the need to study the socio-economic, demographic, health and environmental determinants and differentials of nutritional status in Ethiopia is, so far, there are no much detailed studies conducted to explore all aspects of nutritional status in Ethiopia particularly the effects of socio-economic factors on malnutrition using Bayesian approach.

This study attempts to explore the major socio-economic, demographic, health and environmental factors that affect nutritional status of children in Ethiopia using Semiparametric Bayesian regression model.

## **1.3. Objective of the study**

The general objective of this study is to identify the most dominant predictors of children nutritional status in Ethiopia using Bayesian approach.

**The specific objectives of the study are:**

- i. To examine the effects of some variables that may influence nutritional status of a child and identify the most important determinants;
- ii. To describe the relationship between nutritional status of a child and its determinants.
- iii. To analyze the effect of continuous predictors on nutritional status; and
- iv. To check the consistency of the results in Bayesian and frequentist approach in identifying predictors of nutritional status of children.

#### **1.4. Significance of the Study**

Since children are basic for every development aspect of a country there is a need to: identify factors that affect nutritional status using alternative approach, design a policy and take intervention actions. Thus this research has a significant role for our country to identify the most serious determinants using a Bayesian approach that will help to take action on those identified determinants.

#### **1.5. Limitation of the study**

Spatial information can be analyzed in Bayesian approach and result in very important information to identify the prevalence of stunting region-wise and/or district-wise. Because of lack of having complete spatial information, in this study we have not analyzed the prevalence of child malnutrition in-depth spatially.

Although many factors affect malnutrition as indicated by various studies in different countries (including social, economic, political, cultural, demographic, physiological, biological, reproductive health rights, family planning program etc...)the study undertaken explores the most serious determinants by checking their consistency with previous studies. Since most of observations for some of the very important determinant of children malnutrition like Body mass index and Standardized *height-for-age* (Zscore) are missing, these are excluded from this study.

#### **1.6 Organizational overview of the Study**

This study contains five major chapters. Chapter one presents background of the study, objectives, statement of the problem, its significance and limitation of the study. Chapter two discusses previous literatures on nutritional status of children either in or outside Ethiopia. Chapter three discusses the data and methodology of the study such as sources of data and variables to be included in the study with their coding and

description. Methods of data analysis are also described in this chapter. Chapter four presents statistical data analysis output. Finally, discussion of each of the covariates as per the output in chapter five, conclusions and policy recommendations of the study are dealt with in chapter five.

# CHAPTER TWO

## 2. LITRATURE REVIEW

### 2.1 General

Child as well as infant mortality remains higher in the developing world despite a considerable decline in mortality levels in the last decades. According to the latest State of the World Children Report (UNICEF, 2008), approximately 26,000 children are dying each day around the world mostly from preventable causes and one third of these deaths occur in the first month of life. Nearly all of these deaths occur in 60 developing countries. Moreover, nearly half of the under-five deaths are due to undernutrition. Sub-Saharan Africa is the region worst hit by higher under-five mortality followed by South Asia. Similarly, undernutrition remains to be a grave public health problem in the developing world with South Asia having the most affected region followed by Sub-Saharan Africa (UNICEF, 2008). Chronic malnutrition has been a persistent problem for young children in sub-Saharan Africa. A high percentage of these children fail to reach the normal international standard height for their age; that is, they are "stunted." In contrast, the percentage of children stunted in Southeast Asia dropped from 52 percent to 42 percent between 1990 and 2006. The number of undernourished (low weight for age) people of all ages in sub-Saharan Africa increased from about 90 million in 1970 to 225 million in 2008, and was projected to add another 100 million by 2015, even before the current world food price hikes (Charles H. et al, 2008).

Childhood undernutrition is amongst the most serious health issues facing developing countries. It is an intrinsic indicator of well-being, but it is also associated with morbidity, mortality, impaired childhood development, and reduced labor productivity.

Much of the burden of deaths resulting from malnutrition, estimated to be over half of childhood deaths in developing countries, can be attributed to just mild and moderate undernutrition. Underweight children are particularly vulnerable to increased risk of death from infectious illnesses such as diarrhea and pneumonia (WHO, 2002). For those children that do not survive, the impact of chronic malnutrition in the first few years of life are long lasting and can lead to cognitive and physical developmental deficits, higher levels of chronic illness and disability in adult life (resulting in reduced work capacity), as well as adverse pregnancy outcomes (low birth weight). Several biological and social economic factors contribute to malnutrition (Silva, 2005).

Ethiopia has one of the highest child malnutrition rates in the world. A considerable effort to monitor child malnutrition rates over the last two decades shows that, despite some improvements, approximately half of the children under five are still malnourished (Christiaensen and Alderman, 2001).

The prevalence of stunting in children below five years in East Africa averages about 48 percent (ACC/SCN 2000), which is the highest in the world. Evidence also showed that the situation in Ethiopia is worse than in other East African countries. A review of the trends of the nutritional status of Ethiopian children from 1983-1998 showed that the national rural prevalence of stunting increases from 60 percent in 1983 to 64 percent in 1992. Another national survey undertaken in 1998 with the inclusion of urban areas and children in the age group 3-5 months showed a relative decline in the proportion of stunted children to 52 percent (Zewditu et al., 2001). A few local studies (Woldemariam and Timotiows, 2002; Genebo et al., 1999; Yimer, 2000; Dejen, 2008) on child nutrition have also shown similar results (a more than 40 percent prevalence in stunting) and confirmed that malnutrition, i.e., malnutrition, is one of the most important public health problems in this country.

Undernutrition among children is usually measured by determining the anthropometric status of the child with most research focusing on children below six years of age (e.g. WHO, 2006). Anthropometric measure is the commonly used direct methods for the assessment of nutritional status. The frequently employed anthropometric measurements are weight and height. Anthropometric measurements are economical to carry out, objective, easily understandable, give result which can be numerically graded and provides information on different degree of malnutrition. Therefore, most studies on nutritional status are performed using anthropometric measurements (Anthropometric Indicators Measurement Guide, 2003 Edition).

Since children's height and weight changes with age, the anthropometric measurements are converted into Z-scores based on National Centre for Health Statistics (NCHS) growth standard. The NCHS standard is based on a reference population made up of children who are assumed to be well nourished and is recommended by WHO as a reference to be used in the evaluation of nutritional status (Woldemariam and Timotiows, 2002; Yimer, 2000; Michael, 2000 ).

## **2.2 Factors associated with Children's Malnutrition.**

Important determinants of undernutrition include the education, income, and nutritional situation of the parents, access to clean water and sanitation, and access to primary health care and sex and age of child (MoFED, 1999; Sahn and Stifel, 2002; Christiaensen and Alderman, 2001); Woldemariam and Timotiows, 2002; Yimer, 2000; Micheal, 2000; Dejen, 2008). Factors that are contributing to malnutrition may differ among regions, communities and over time. Identifying the underlying causes of malnutrition in a particular locality is important to solve the nutritional problems.

Various studies have been made and conclusions were reached by different scholars in the past regarding predictors of health and nutritional status. Survey of available literature indicated that factors like knowledge of health practices and caring level, educational level of parents, access to or interactions of age of the child have strong effect on household and community variables in which the child grows up.

Approximately 10 percent of children born in Ethiopia die before their first birthday and 17 percent will die before their fifth birthday (CSA and ORC Macro, 2001). According to formulas developed by Pelletier et al. (1994), 57 percent of under-five mortality in Ethiopia is related to severe and mild to moderate malnutrition (ORC Macro, 2001). The consequences of malnutrition in children also include poor physical development and limited intellectual abilities that diminish their working capacity during adulthood. Some of the socioeconomic and demographic factors explaining child nutrition according to studies done in different places are reviewed below.

### **2.3.1 Household economic status**

The economic status of a household is an indicator of access to adequate food supplies, use of health services, availability of improved water sources, and sanitation facilities, these are prime determinants of child nutritional status (UNICEF, 2004). Comparative studies on child nutrition for more than 15 countries (Sommerfelt and Stewart, 1994) and some local studies in Ethiopia (Woldermariam and Timotiws, 2002; Genebo et al., 1999; Yimer, 2000) showed that the higher the level of economic status of the household, the lower the level of child stunting. The results indicated that the household wealth status had strong negative effects on both stunting and underweight, but not on wasting. The effect was stronger on stunting than on underweight. The odds of stunting decline monotonically with increase in economic status (Olalekan, 2008). As acknowledged in many studies, an increase in household

income/wealth is expected to reduce child malnutrition (Christiaensen and Alderman, 2001). Household assets have a strong significant correlation with children's heights and portray a U-shaped relationship with nutritional status. The results imply that nutrition improves at a decreasing rate with assets (Sahn and Stifel, 2003).

### **2.3.2 Education of Mother**

Education is one of the most important resources that enable women to provide appropriate care for their children. Education of women is believed to exert an impact on health and nutritional status of children since it provides the mother with the necessary skills for child care, increase awareness of nutritional needs and preference of modern health facilities as well as change of traditional beliefs about diseases causation, and use of contraceptives for birth spacing. Studies in Ethiopia (Aschalew, 2000; Micheal, 2000; Yimer, 2000; Dejen, 2008), indicate that the relationship between chronic nutritional status of children and maternal education. The likelihood of being stunted was found to be double among children of mother with no education compared with children whose mothers have some secondary or higher education (Woldemariam and Timotiows, 2002). The low educational status could be one of the reasons for the low level of employment among mothers as well as the poor food security in their households (Klasen, 2008). Various studies have concluded that parental education, especially mothers' education, is a key element in improving children's nutritional status (Christiaensen and Alderman, 2004). More specifically, they found that the effect of maternal education is about twice as important as that of partner education. Education of the household head is insignificant implying that father's education may not be an important determinant of a child's nutritional status (Sahn and Stifel, 2003).

### **2.3.3 Employment Status of Mothers**

Although women's employment enhances the household's access to income, it may also have negative effects on the nutritional status of children, as it reduces a mother's time for childcare. Some studies have revealed that mothers of the most malnourished children work outside their home (Abbi et al., 1991). Children whose mothers are employed were more likely to be stunted than children whose mothers were unemployed (Dejen, 2008; Khalid, 2006; Woldemariam & Timotiwos, 2002). Another study argued that there is no association between maternal employment and children's nutritional status (Leslie, 1988). The low educational status could be one of the reasons for the low level of employment among the mothers as well as the poor food security in their households.

### **2.3.4 Source of Water and availability of Toilet Facility**

Unfavorable health environment caused by inadequate water and sanitation can increase the probability of infectious diseases and indirectly cause certain types of malnutrition. A comparative study in some developing countries (Sommerfelt and Stewart, 1994) and in Jimma, Ethiopia (Getaneh et al., 1998) showed that unprotected water source and non-availability of latrine were associated with low child stature.

A household's access to facilities is likely to be correlated with community characteristics. Households living in wealthier communities might have a relatively healthy environment, which implies better sanitation facilities, access to clean water and healthcare facilities. Water and sanitation play a particularly important role in child nutrition due to their impact on diarrhea diseases.

Christiaensen and Alderman (2001) noted that in Ethiopia, 14 per cent of households in urban communities get water from their own tap and 3 per cent have flush toilets.

But no households in rural areas get water from pipe in their home. Their findings indicate a significant and positive impact of these facilities on child nutrition.

### **2.3.5 Body Mass Index (BMI)**

Most previous studies on children malnutrition include Body Mass Index (BMI) as a major determinant of children malnutrition. Some studies in Ethiopia tell us, the percentage of stunted children (Z-score <-2) was a bit higher among stunted mothers (<145 cm in height) than normal height mothers ( $\geq 145$  cm) (Yimer, 2000; Woldemariam and Timotiows, 2002). In the Woldemariam and Timotiows study, 64.3 percent of the children of stunted mothers were stunted. Although the level of stunting, underweight and wasting was also higher in children of malnourished mothers (BMI < 18.5) as compared to well-nourished mothers (BMI  $\geq 18.5$ ), no significant statistical association was observed. But mothers with low BMI on average giving birth to babies of low birth weight.

### **2.3.6 Age of child**

Children's nutritional status is also more sensitive to factors such as feeding/weaning practices, care, and exposure to infection at specific ages. A cumulative indicator of growth retardation (height-for-age) in children is positively associated with age (Aschalew, 2000). Local studies in Ethiopia have also shown an increase in malnutrition with increase in age of the child (Yimer, 2000; Genebo et al., 1999; Samson and Lakech, 2000). Children in the age group 0-5 months were found to be at significantly lower risk of stunting as compared with children in the age group 6-11 months (Woldemariam and Timotiows, 2002). The prevalence of diseases and stunting rises with age (WHO, 2006).

### **2.3.7 Sex of Child**

In sub-Saharan Africa, male children under five years of age are more likely to become stunted than females, which might suggest that boys are more vulnerable to health inequalities than their female counterparts in the same age groups. In several of the surveys, sex differences in stunting were more pronounced in the lowest SES groups (Henry et al, 2007). A number of studies in Africa suggest that rates of malnutrition among boys are consistently higher than among girls (MoFED, 1999; Sahn and Stifel, 2003; Christiaensen and Alderman, 2001).

### **2.3.8 Birth Order**

It is expected that parents give less attention to older children when a new child is born because the new born needs much attention and care. One study showed that stunting is rare in birth orders 2-3 (Sommerfelt et al., 1994), and higher birth order (5+) is positively associated with child malnutrition. The risk of stunting was also 1.3 times higher for children of first birth order as compared with children of birth order six or more (Woldemariam and Timotiws, 2002).

### **2.3.9 Birth interval of the child**

Closely spaced pregnancies are often associated with the mother having little time to regain lost fat and nutrient stores (ACC/SCN, 2000). Higher birth spacing is also likely to improve child nutrition, since the mother gets enough time for proper childcare and feeding. Studies in developing countries showed that children born after a short birth interval (less than 24 months) have higher levels of stunting in most countries where DHS surveys have been conducted. It was also observed that as the preceding birth interval of the child decreases, the likelihood of being stunted increases (Woldemariam and Timotiws, 2002).

### **2.3.10 Environmental Factor**

Much of the burden of deaths resulting from malnutrition, estimated to be over half of childhood deaths in developing countries, can be attributed to just mild and moderate malnutrition. Several biological and social economic factors contribute to malnutrition. The impact of access to basic environmental services, such as water and sanitation on the probability children are stunted and underweight is examined in detail in Ethiopia and found as very significant (Silva, 2005).

# CHAPTER THREE

## 3. DATA AND METHODOLOGY

### 3.1 Data

This study is based on data from the 2005 Demographic and Health Survey of Ethiopia (EDHS-2005) with reference to 3140 children providing complete information for predictors of children malnutrition those examined in this study and plausible anthropometric data. The survey draws a representative sample of women of reproductive age, administers a questionnaire and makes an anthropometric assessment of women and their children that were born within the previous five years. The datasets contain information on family planning, maternal and child health, child survival, HIV-AIDS, educational attainment, and household composition and characteristics.

In this study, height and weight measurements of the children, taking age and sex into consideration, were converted into Z-scores based on the National Center for Health Statistics (NCHS) reference population recommended by the World Health Organization (WHO). Thus, those below 2 standard deviations of the NCHS median reference for *height-for-age*, *weight-for-age* and *weight-for-height* are defined as stunted, underweight, and wasted, respectively. In this study we consider stunting which is an indicator of linear growth retardation relatively uncommon in the first few months of life. However, it becomes more common as children get older. Children with *height-for-age* z-scores below two standard deviations from the median of the reference population are considered as short for their age or stunted. Furthermore, children with z-scores below three standard deviations from the median of the reference

population are considered to be severely stunted, while children with z-scores between three and two standard deviations are known to be moderately stunted; these all three indicators are used to describe the level of child malnutrition and the relationship between maternal and child nutritional status. Moreover, stunting measures linear growth retardation and cumulative growth deficit and indicates the effect of past or chronic nutritional status in the life of the child. Therefore, an in-depth analysis was performed on stunting, focusing on factors affecting chronic malnutrition.

Undernutrition among children is usually determined by assessing the anthropometric status of a child relative to a reference standard. In this study we consider undernutrition as measured by stunting or insufficient *height-for-age*, indicating chronic undernutrition. Height-for-age score for a child  $i$  is determined using a Z-score which is defined as

$$Z_i = \frac{AI_i - MAI}{\sigma},$$

where  $AI$  refers to the child's anthropometric indicator (height at a certain age in our case),  $MAI$  refers to the median of the reference population and  $\sigma$  refers to the standard deviation of the reference population. Height-for-age z-score is an indicator of the nutritional status of a child. Here the main interest is in modeling the dependence of nutritional status on covariates including the age of the child, the body mass index of the child's mother, the district the child lives in, mother and partner education, mother working status, sex of child, birth order and birth interval, household economic status, and environmental conditions.

### **3.1.1 Variables**

As discussed in the literature review socio-economic, demographic, health and environmental characteristics are to be proximate determinants of nutritional status.

#### **3.1.1.1 The Response Variable**

Anthropometric indicators were constructed using data on the children's age, height and weight. Three key anthropometric measures were calculated using the new WHO reference standard (WHO, 2006). In our application on childhood nutritional status, which is the response variable is measured by Z-score (multiplied by 100).

#### **3.1.1.2 Explanatory Variables/factors**

Since the nutritional status of children is affected by a number of factors, the predictor variables/factors to be analyzed in this study as determinants of nutritional status of children are grouped as socio-economic, demographic, health and environmental factors. These factors include the child's age, sex, birth order, mother's age at childbirth, body mass index (BMI) and mother education, residence, water supplies and toilet facilities.

##### **i. Socio-Economic Characteristics**

It is evident that socioeconomic status of mothers has a strong relationship with chronic malnutrition. As proxy indicators of socioeconomic status of mothers the following factors are included: mother education, partner education and household income.

##### **ii. Demographic Characteristics**

Demographic characteristics includes age of the child, sex of the child, birth interval and birth order.

##### **iii. Health and Environments Characteristics**

Place of residence, water supplies and toilet facilities are important health and environmental factors included in this study.

*Table 3.1 Covariates Description with their coding.*

<b>Variable</b>	<b>Level</b>	<b>Coding</b>
Place of residence(Res)	Urban	1
	Rural	0(ref)
Child's sex(Csex)	Male	1
	Female	0(ref)
Mother Working(Mwork)	Yes	1
	No	0(ref)
Mother's Education(Medu)	No (Medu1)	0(ref)
	Primary (Medu2)	1
	Secondary and higher (Medu3)	2
Birth Order(BORD)	First to third	0(ref)
	Above	1
Source of Drinking Water(Swat)	Controlled (Protected)	1
	Not controlled (Unprotected)	0(ref)
Partner's education level(Pedu)	No (Pedu1)	0(ref)
	Primary (Pedu2)	1
	Secondary and higher(Pedu3)	2
Preceding birth interval (Pbint)	<=24 months (Pbint1)	0(ref)
	24 to 36 months (Pbint2)	0(ref)
	>36 months (Pbint3)	1
HH Economic Status(Eco)	Poor	0
	Medium or higher	1
Body Mass Index(BMI)	<b>Continuous Covariates</b>	
Child age (Cage)		
Current age of respondent (Mage)		
Height-for-age(Zscore)		
		<b>Response variable</b>

## 3.2 Methodology

There has been much recent interest in Bayesian inference for generalized additive and related models. The increasing popularity of Bayesian methods for these and other model classes is mainly caused by the introduction of Markov Chain Monte Carlo (MCMC) simulation techniques which allow realistic modeling of complex problems.

In the last two decades, great advances in technology and industry have produced various high throughput measurement instruments. Large scale measurements have been collected in different areas in industry and technology, in the form of data as curves (e.g. thermal diffusivity measurements), data array and images (e.g. DNA microarray experiments), and spatial observations of space-time systems (e.g., satellite images), etc. Efficient use of these great resources raises a host of new methodological and experimental design issues in statistics, such as: how to assess and calibrate the instruments? , how to combine data from heterogeneous sources? , how to do the error propagation in high dimensional and nonlinear input-output systems? , how to combine model uncertainty and prior information (e.g., type B uncertainty) with experimental results? , how to design multi-stage experiments used in automated testing and other areas? The Bayesian approach offers a systematic and flexible approach to these problems. By adopting an objective or non-informative prior, the Bayesian approach produces estimates and uncertainty measures comparable to the classical approach.

Thus, the Bayesian approach offers the viable and rigorous solution, though there is also the added benefit of providing much-needed uncertainty and probability assessments in nonlinear and non-Gaussian situations in a valid and rigorous way.

The statistical analysis in this thesis is based on Bayesian approaches which allow a flexible framework for realistically complex models. These approaches allow us to analyze usual linear effects of categorical covariates and nonlinear effects of continuous covariates within a unified semi-parametric Bayesian framework for modeling and inference. In this work firstly, we analyze the effects of the different types of covariates on the response variable “nutritional status” by using *height-for-age* Z-score as continuous variable with the assumption that each of the covariates has a linear effect on the response variable. In this case, a Bayesian additive Gaussian regression model is used. Since we assume each covariate has a linear effect the approach we follow in this step is Bayesian linear regression parametric approach. Secondly, since some studies suggest that body mass index and age of child as having nonlinear effect, we modify the first case, the Bayesian additive Gaussian regression parametric approach to accommodate some transformation of these two covariates. Thirdly, to analyze the nutritional status of children we employ the same Bayesian approach, but in this case, since the three continuous covariates BMI (body mass index of the mother, Cage (age of the child) and Mage (Mother age at birth) are assumed to have a possibly nonlinear effect on the Z-score and therefore modeled nonparametrically (as cubic P-splines with second order random walk prior). Finally, we employ the Deviance Information Criteria to compare the three models.

Generalized Additive Models are methods and techniques developed and popularized (Hastie and Tibshirani, 1990). We examine the generalized additive model as an alternative to the common linear model in the context of analyzing childhood nutritional status in Ethiopia. Most applications are still based on generalized linear models, assuming that covariate effects can be modeled by a parametric linear predictor. In our study, however, the data contain detailed information on continuous covariates like body mass index of the mother and child

age. Their effects are often highly nonlinear, and are difficult to assess with conventional parametric models. In this study, we propose generalized additive models which can simultaneously incorporate the usual linear effects as well as nonlinear effects of continuous covariates within a semi parametric Bayesian approach. The inference we make is fully Bayesian and uses recent Markov Chain Monte Carlo (MCMC) simulation techniques for drawing random samples from the posterior.

### **3.2.1 Bayesian structured additive regressions models based on MCMC techniques**

#### **Generalized linear models**

A common way to build regression models extending the classical linear model for Gaussian responses to more general situations such as binary responses are generalized linear models, originally introduced by Nelder and Wedderburn (1972). For more overviews see Fahrmeir and Lang (2001). In this model the influence of covariates on a response variable  $y$  is assumed to satisfy the following two assumptions:

#### **i. Distributional assumption**

Conditional on covariates  $X$ , the responses  $y$  are independent and the distribution  $y$  belongs to a simple exponential family, i. e. its density can be written as

$$p(y_i | x_i) = \exp\left\{\frac{[y_i\theta_i - b(\theta_i)]}{\phi} w_i + c(y_i, \theta_i, w_i)\right\}, i = 1, \dots, n \quad (3.1)$$

where:

$\theta_i$ , is the natural parameter of the exponential family,

$\phi$ , is a scale or dispersion parameter common to all observations,

$w_i$ , is a weight for  $i$ , and  $b(\cdot)$  and  $c(\cdot)$  are functions depending on the specific exponential family.

**ii. Structural assumption**

The (conditional) expectation  $E(y/x) = \mu$  is linked to the linear predictor as

$$\begin{aligned} \eta &= X' \beta \\ \text{Vs} \\ \mu &= h(\eta) \qquad \qquad \text{Or} \qquad \eta = g(\mu) \end{aligned} \qquad (3.2)$$

Where:

The design vector  $X$  usually includes the grand mean.

$h$  is a smooth, bijective response function.

$g$  is the inverse of  $h$  called the link function and.

$\beta$  a vector of unknown regression coefficients.

Both assumptions are connected by the fact that the mean of  $y$  is also determined by the distributional assumption and can be shown to be given as

$$\mu = E(y | x), \text{ and let it be } = b'(\theta),$$

In addition,  $\text{var}(y | X)$  is the variance of  $y$  in general which is depending on the linear predictor with  $\frac{\phi v(\mu)}{w_i}$  being the variance function of the underlying exponential family.

Thus the variance of  $y$  is given by

$$\text{var}(y_i | x_i) = b''(\theta_i) / w_i, \qquad \text{where} \qquad b''(\theta) = \frac{\phi v(\mu)}{w_i}$$

Note that, the distribution of  $y$  could be normal, Poisson and binary (binomial) or any other exponential family distribution.

The classical linear model can be subsumed into the context of generalized linear models by  $h(\mu) = \mu$ ; i.e. the response function is simply the identity. For Gaussian distributed responses this also represents the natural link function. The variance function  $v(\mu)$  is constant, while the scale parameter equals the variance of the error terms of the linear regression model.

### 3.2.2 Bayesian Linear regression (Parametric) Model

Despite the fact that practical experience has shown that continuous covariates often have nonlinear effects (mostly unknown nonlinear effects); in this study, first we need to see the result by assuming each of the covariates has a linear effect on the response variable and compare with the result from the Semiparametric model. The effect of the covariates on the response is modeled by a linear predictor as:

$$\eta_i = X_i' \beta + W_i' \gamma \quad i = 1, 2, \dots, n$$

where:

$X_i = (x_{i1}, \dots, x_{ip})$ , is a vector of continuous covariates,

$\beta = (\beta_0, \beta_1, \dots, \beta_p)$ , is a vector of regression coefficients for the continuous covariates.

$W_i = (w_{i1}, \dots, w_{ik})'$ , is a vector of categorical covariates.

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

In the Bayesian parametric regression model the parameter vectors  $\beta$  and  $\gamma$  we routinely assume diffuse priors  $p(\gamma) \propto \text{const}$  and  $p(\beta) \propto \text{const}$ . A possible alternative would be to work with a multivariate Gaussian distribution  $\gamma \sim N(\gamma_0, \Sigma_{\gamma_0})$  and  $\beta \sim N(\beta_0, \Sigma_{\beta_0})$ . However, since in most cases a non-informative prior is desired,

we consider it sufficient to work with diffuse priors. Furthermore, assuming these priors emphasize the close connection of our empirical Bayes approach to (penalized) maximum likelihood estimation.

### 3.2.3. Bayesian semiparametric Model

The assumption of a parametric linear predictor for assessing the influence of covariate effects on responses seems to be rigid and restrictive in practical application situation and also in many real statistically complex situation since their forms cannot be predetermined a priori. Besides, practical experience has shown that continuous covariates often have nonlinear effects. In our study, for the continuous covariates in the data set, the assumption of a strictly linear effect on the predictor may not be appropriate, i. e. some effects may be of unknown nonlinear form (such as, mother's age and mother's BMI) as suggested in (Khalid, 2007; Mohammed, 2008). Similarly, the continuous covariate, child age, seems to have nonlinear effect. Thus these variables have a nonlinear effect on the response variable.

Hence, it is necessary to seek for a more flexible approach for estimating the continuous covariates by relaxing the parametric linear assumptions. This in turn allows continuous covariates to follow their true functional form. This can be done using an approach referred to as nonparametric regression model. To specify a nonparametric regression model, an appropriate function that contains the unknown regression function needs to be chosen. This choice is usually motivated by smoothness properties, which the regression function can be assumed to possess.

Consider regression situations, where observations  $(y_i, X_i, W_i), i = 1, \dots, n$  and parameters  $\beta$  and  $\gamma$  as defined above. In our application to childhood nutritional status the response is stunting measured as a Z-score (multiplied by 100), the

continuous covariates ( $X_i$ ) include the child's age, mother's age at birth and mother's BMI and categorical covariates ( $W_i$ ) include child's sex, residence of the child, and educational level of mother... etc. Generalized additive and semiparametric models (Hastie and Tibshirani, 1990) assume that, given  $X_i = (x_{i1}, \dots, x_{ip})'$ , and  $W_i = (w_{i1}, \dots, w_{ip})'$  the distribution of  $y_i$  belongs to an exponential family, with mean  $\mu_i = E(y_i | X_i, W_i)$  linked to an additive predictor  $\eta_i$  by an appropriate response function  $h$ . We assume a semiparametric regression model with categorical predictors as:

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

$$\text{where } \eta = f_1(x_{i1}) + \dots + f_p(x_{ip}) + W_i' \gamma \quad (3.4)$$

Here  $h$  is a known response function, and  $f_1, \dots, f_p$  are possibly nonlinear smooth functions of continuous covariates.

We assume  $y_i \sim N(\eta_i, \tau^2)$ .

### 3.2.4 Prior distribution

In Bayesian inference, the unknown function  $f_j, j=1, \dots, p$ , the fixed effects parameters  $\gamma = (\gamma_1, \dots, \gamma_r)'$  as well as the variance parameter  $\tau^2$  are considered as random variables and have to be supplemented by appropriate prior assumptions. In the absence of any prior knowledge we assume independent diffuse priors  $\gamma_j \propto \text{const}, j=1, \dots, r$  for the parameters of fixed effects. Another common choice is highly dispersed Gaussian priors. Several alternatives are available for the priors of the unknown (smooth) functions  $f_j, j=1, \dots, p$ . Among others, random walk priors (Fahrmeir and Lang,

2001), Bayesian smoothing splines (Hastie and Tibshirani, 2000) and Bayesian P-splines (Lang and Brezger, 2004), are the most known. In this study, for the continuous covariates BMI, Cage and Mage those assumed to have a possibly nonlinear effect on the Z-score are modelled nonparametrically as cubic P-spline with second order random walk priors.

### i. The General Form of the Priors

Suppose that  $f = (f(1), \dots, f(n))'$  is the vector of corresponding function evaluations at observed values of  $x$ .

Then, the general form of the prior for  $f$  is

$$f | \tau^2 \propto \exp\left(-\frac{1}{2\tau^2} f' K f\right) \quad (3.5)$$

where,  $K$  is a penalty matrix that penalizes too abrupt jumps between neighboring parameters. In most cases  $K$  will be rank deficient; therefore the prior for  $f$  would be improper. This implies that  $f / \tau^2$  follows a partially improper Gaussian prior  $f_j / \tau^2 \sim N(0, \tau^2 K^-)$  where  $K^-$  is a generalized inverse of a band-diagonal precision or penalty matrix  $K$ . In the frequentist approach the smoothing parameter is equivalent to the variance parameter  $\tau^2$  which controls the tradeoff between flexibility and smoothness. In order to estimate the smoothness parameter  $f$ , a highly dispersed but proper hyperprior is assigned to  $\tau^2$ . The proper prior for  $\tau^2$  is required to obtain a proper posterior for  $f$  (Hobert and Casella, 1996). We choose an inverse gamma distribution with hyperparameters  $a$  and  $b$ , i.e.

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

with probability density function given by

$$p(\tau^2 | a, b) = (\tau^2)^{-a-1} \exp\left(-\frac{b}{\tau^2}\right)$$

Common choices for  $a$  and  $b$  are  $a=1$  and  $b=0.005$  (or  $b=0.0005$ ). Alternatively, one may take  $a=b=0.001$ . Brezger and Lang (2006) also suggest a general structure of the priors as

$$p(\beta_j | \tau_j^2) \propto \frac{1}{(\tau_j^2)^{\text{rank}(K_j)/2}} \exp\left(-\frac{1}{2\tau_j^2} \beta_j' K_j \beta_j\right)$$

where  $K_j$  is a penalty matrix that shrinks parameters towards zero or penalizes too abrupt jumps between neighboring parameters. In most cases  $K_j$  will be rank deficient and therefore the prior for  $\beta_j$  is partially improper

## ii. Priors for Fixed Effects

As indicated above for the parameter vector  $\gamma$  of fixed effects we choose a diffuse prior

$$\gamma_j \propto \text{const}, j = 1, \dots, r$$

Another choice would be to work with a multivariate Gaussian distribution  $\gamma \sim N(\gamma_o, \Sigma_{\gamma_o})$ . In this study, diffuse priors will be used for the fixed effects.

## iii. Priors for Continuous Effects under linear setup

Priors for the unknown functions  $f_1, \dots, f_p$  depend on the type of the covariates and on prior beliefs about the smoothness of  $f_j$ . In the following we express the vector of

function evaluations  $f_j = (f_j(x_{1j}), \dots, f_j(x_{nj}))'$  of a function  $f_j$  as the matrix product of a design matrix  $X_j$  and a vector of unknown parameters  $\beta_j$ 's, i.e.

$$f_j = X_j' \beta_j$$

Then, we obtain the predictor (3.4) which takes linear form as

$$\eta = X_1 \beta_1 + \dots + X_p \beta_p + W' \gamma$$

where  $W$  corresponds to the usual design matrix for categorical covariates.

For the variance parameter  $\tau_j^2$ , the inverse gamma with hyperparameters  $a$  and  $b$  assumed as discussed above.

On the other hand if the unknown function  $f_j, j = 1, \dots, p$  assumes a smooth nonlinear function, then a random walk priors or Bayesian P-spline can be considered as discussed in the next section.

#### iv. First and second order random walk priors

Let us consider the case of a continuous covariate  $x$  with equally-spaced observations  $x_i, i = 1, \dots, m, m \leq n$ , then,  $x_{(1)} < \dots < x_{(m)}$  defines the ordered sequence of distinct covariate values. Here  $m$  denotes the number of different observations for  $x$  in the data set. A common approach in dynamic or state space models is to estimate one parameter  $f(t)$  for each distinct  $x(t)$ ; i.e., define,  $f(t) = f(x_{(t)})$  and let  $f = (f(1), \dots, f(m))'$  denote the vector of function evaluation. Then a first order random walk prior for  $f$  is defined by

$$f(t) = f(t-1) + u(t) \tag{3.6}$$



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

Where,  $K^-$  is a generalized inverse of the penalty matrix  $K$

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

$$f(t) = f(t-1) + u(t) \quad (3.10)$$

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

i.e. by adjusting from  $\tau^2$  to  $\delta_t \tau^2$ .

Similarly random walks of second order is given by

$$f(t) = (1 + \frac{\delta_t}{\delta_t - 1})f(t-1) - (\frac{\delta_t}{\delta_t - 1})f(t-2) + u(t) \quad (3.11)$$

where  $u(t) \sim N(0; w_t \tau^2)$  and  $w_t$  is an appropriate weight. Several possibilities are conceivable for weights, see (Fahrmeir and Lang, 2001) for a discussion.

## v. Bayesian P-splines

Any smoother depends heavily on the choice of smoothing parameter, and for P-spline in a mixed (fixed and continuous) framework. A closely related approach for continuous covariates is based on the P-splines approach introduced by Eilers and Marx (1996). The basic assumption of this approach is that the unknown function

$f_j$  can be approximated by a spline of degree  $l$  with equally spaced knots  $x_{\min} = \xi_0 < \xi_1 < \dots < \xi_{r-1} < \xi_r = x_{\max}$  within the domain of  $x_j$ . The domain from  $x_{\min}$  to  $x_{\max}$  can be divided into  $n'$  equal intervals by  $n'+1$  knots. Each interval will be covered by  $l+1$  B-splines of degree  $l$ . The total number of knots for construction of the B-splines will be  $n'+2l+1$ . The number of B-splines in the regression is  $n = n'+l$ . It is well known that such a spline can be written in terms of a linear combination of  $M = r+l$  B-splines basis functions  $B_j$ , i.e.

$$f_j(x_{ij}) = \sum_{p=1}^M \beta_j B_j(x) \quad (3.12)$$

The basis functions  $B_j$  are defined locally in the sense that they are nonzero only on a domain spanned by  $2 + l$  knots. The  $n * M$  design matrix  $x_j$  for P-splines is more intricate than the case of random walk priors. Each row  $i$  contain the value of the B-spline basis functions evaluated at  $x_i$ , hence  $x_j(i, p) = B_{jp}(x_{ij})$ . In accordance with the properties of B-splines, each row  $X$  has  $M + 1$  non-zero values. As for the number of knots, Eilers and Marx (1996) recommended the number of inner knots to range between 20 and 40 and introduced a penalization of the differences between regression coefficients of adjacent B-spline basis functions in order to generate a smoothing effect. In our analysis, we typically choose B-splines of degree 3 and 10 intervals, and second order random walk priors on the B-splines regression coefficients.

### 3.2.5 Posterior Inference

When performing Bayesian inference, all inferential conclusions are based on the posterior of the model. In an empirical Bayes approach to structured additive regression, no hyperprior are assigned to the hyperparamters, i. e. the variances  $\tau_j^2$  are treated as fixed. In this case, the specific form of the posterior depends only on the

parameterization of the regression terms in the model. Then, we use Markov Chain Monte Carlo (MCMC) simulations to draw samples from the posterior and statistical inference is done by means of Markov Chain Monte Carlo techniques in a full Bayesian setting. Now we restrict the presentation to models with predictor (3.4) and (3.12). Full Bayesian inference is based on the entire posterior distribution. Let  $\alpha$  be the vector of all unknown parameters, then the posterior is given by

$$\begin{aligned}
 p(\alpha | y) &\propto L(y, \beta_1, \tau_1, \beta_2, \tau_2, \dots, \beta_p, \tau_p, \gamma) \prod_{j=1}^p p(\beta_j | \tau_j^2) p(\tau_j^2) & (3.13) \\
 &\propto L(y, \beta_1, \tau_1, \beta_2, \tau_2, \dots, \beta_p, \tau_p, \gamma) \\
 &\quad \prod_{j=1}^p \frac{1}{(\tau_j^2)^{rk(K_j)/2}} \exp\left(-\frac{1}{2\tau_j^2} \beta_j' K_j \beta_j\right) \prod_{j=1}^p (\tau_j^2)^{-a_j-1} \exp\left(-\frac{b_j}{\tau_j^2}\right)
 \end{aligned}$$

Bayesian inference via MCMC is based on updating full conditionals of single parameters or blocks of parameters, given the rest and the data. For Gaussian models, Gibbs sampling with so-called multi move steps can be applied. For non-Gaussian responses Gibbs sampling is no longer feasible and Metropolis Hastings algorithms are needed. More detail can be found in Rue (2001) or Fahrmeir and Lang (2001).

In a fully Bayesian approach, parameter estimates are generated by drawing random samples from the posterior (3.13) via MCMC simulation techniques. The variance parameters  $\tau_j^2$  can be estimated simultaneously with the regression coefficients  $\gamma_j$  by assigning additional hyperprior to them. The most common assumption is, that the  $\tau_j^2$  are independently inverse gamma distributed, i.e.  $\tau_j^2 \sim IG(a_j, b_j)$ , with hyperparameters  $a_j$  and  $b_j$  specified a priori. A standard choice is to use  $a_j=b_j= 0.001$ . In some data situations (e. g. for small sample sizes), the estimated nonlinear functions  $f_j$  may depend considerably on the particular choice of hyperparameters. It is therefore good practice to estimate all models under consideration using a (small) number of different choices for  $a_j$  and  $b_j$  to assess the dependence of results on minor changes in the prior assumptions.

Suppose first that the distribution of the response variable is Gaussian, i.e.  $y_i | \eta_i, \tau^2 \sim N(\eta_i, \tau^2), i = 1, \dots, n$ . In this case an additional hyperprior for the scale parameter  $\sigma^2$  has to be specified. Similarly as for the variances of the regression coefficients, an inverse Gamma distribution  $\tau^2 \sim IG(a_o, b_o)$  is a convenient choice. For Gaussian responses the full conditionals for fixed effects as well as nonlinear functions  $f_j$  are multivariate Gaussian. Thus, a Gibbs sampler can be used where posterior samples are drawn directly from the multivariate Gaussian distributions.

### 3.3 Model Selection

A widely used statistic for comparing models in a Bayesian framework is the Deviance Information Criterion. The **deviance information criterion (DIC)** is a hierarchical modeling generalization of the AIC (Akaike information criterion) and BIC (Bayesian information criterion, also known as the Schwarz criterion). It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. Like AIC and BIC it is an asymptotic approximation as the sample size becomes large. It is only valid when the posterior distribution is approximately multivariate normal.

Define the deviance as  $D(\theta) = -2 \log(p(y|\theta)) + C$ , where  $y$  are the data,  $\theta$  are the unknown parameters of the model and  $p(y|\theta)$  is the likelihood function.  $C$  is a constant that cancels out in all calculations that compare different models, and which therefore does not need to be known.

The expectation  $\bar{D} = E[D(\theta)]$  is a measure of how well the model fits the data; the larger this is, the worse the fit.

The effective number of parameters of the model is computed as  $p_D = \bar{D} - D(\bar{\theta})$ , where  $\bar{\theta}$  is the expectation of  $\theta$ . The larger this is, the better it is for the model to fit the data.

The deviance information criterion is calculated as

$$DIC = p_D + \bar{D}$$

The idea is that models with smaller DIC should be preferred to models with larger DIC. Models are penalized both by the value of  $\bar{D}$ , which favors a good fit, but also (in common with AIC and BIC) by the effective number of parameters  $p_D$ . Since  $\bar{D}$  will decrease as the number of parameters in a model increases, the  $p_D$  term compensates for this effect by favoring models with a smaller number of parameters.

The advantage of DIC over other criteria, for Bayesian model selection, is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation. AIC and BIC require calculating the likelihood at its maximum over  $\theta$ , which is not readily available from the MCMC simulation. But to calculate DIC, simply compute  $\bar{D}$  as the average of  $D(\theta)$  over the samples of  $\theta$ , and  $D(\bar{\theta})$  as the value of  $D$  evaluated at the average of the samples of  $\theta$ . Then the DIC follows directly from these approximations.

### 3.4 Software

There exists no widely used statistical software for Bayesian inference that compares to commercial products such as SAS, SPSS, or Stata. (This is likely due to the fact that all of these packages were developed before MCMC methods were widely used.) Over the last decade a number of promising software packages have emerged among these WinBUGS and BayesX are the most basic ones.

Currently, the most widely used software for Bayesian inference is WinBUGS (Spiegelhalter et al, 2003) which has been developed by the MRC Biostatistics Unit in Cambridge. The package is available free of charge at <http://www.mrc-bsu.cam.ac.uk/bugs/>. WinBUGS may be seen as a kind of (easy to use) programming language that allows specifying and estimating almost any Bayesian model. Hence, in principle the models supported in BayesX could be estimated in WinBUGS as well. However, a price is paid for the extreme flexibility: the comparison with WinBUGS shows that BayesX is much faster and shows superior mixing properties for the resulting Markov chains.

All computations to implement the methodology discussed here are carried out with BayesX program-version 2.0 (Belitz et al, 2009) and we present the nonparametric outputs using R 2.9.0 by calling BayesX in R, we also use STATA 10 to plot relationship between response variable and the covariates and SPSS 15 mostly applied in editing and managing the data.

There has been much recent interest in Bayesian inference for generalized additive and related models. The increasing popularity of Bayesian methods for these and other model classes is mainly caused by the introduction of Markov chain Monte Carlo (MCMC) simulation techniques which allow realistic modeling of complex

problems. BayesX extends the capabilities of existing software for semiparametric regression included in S-PLUS, SAS, R or Stata. Examples are generalized additive (mixed) models, dynamic models, varying coefficient models, geoadditive models, geographically weighted regression and models for space-time regression. BayesX supports the most common distributions for the response variable. For univariate responses these are Gaussian, Binomial, Poisson, Gamma, negative Binomial, zero inflated Poisson and zero inflated negative binomials. For multicategorical responses, multinomial logit and probit models for unordered categories of the response as well as cumulative threshold models for ordered categories can be estimated. Moreover, BayesX allows the estimation of complex continuous time survival and hazard rate models.

# CHAPTER FOUR

## 4. ANALYSIS AND RESULT

### 4.1 Descriptive and Explanatory Analysis of Variables

Based on the data from the 2005 Ethiopia DHS (CSA and ORC Macro, 2006), the overall prevalence of stunting among children in Ethiopia is nearly one in two (45.4 percent) and more than one in four children (23.3%) are severely stunted. The level of stunting, underweight, and wasting are also higher for rural children than urban children. This shows that Ethiopia has a very high prevalence of stunting, underweight and wasting according to the classification established by the World Health Organization to indicate levels of child malnutrition (WHO, 2006).

In this paper, we estimated a model that mainly focuses on factors that are underlying determinants of nutritional status of children. The aim of this chapter is twofold: first, present and discuss descriptive results. Secondly, through our empirical model building to investigate linear and non-linear effects of covariates more flexibly than most previous works.

#### 4.1.1 Categorical Covariates

The major socioeconomic, demographic, health and environmental background characteristics of the respondents and children are presented in Table 4.1. The total number of children covered in the present study is 3,140. Among these, 1,424 (45.4%) were found to be stunted.

We will go through the description and explanation of the variables included in this study. The variables that will be described in this section are used to assess the most

important influential factors on child malnutrition. Summary result of the various variables included are presented in Table 4.1 and discussed as follows.

**i. Rural and urban residence**

Place of residence, whether urban or rural, is one important characteristic that determines access to services and exposure to information pertaining to reproductive health and other aspects of life. As expected, the majority of respondents reside in rural areas, with only 18 percent of women residing in urban areas (EDHS, 2005). Since most of stunted children live in rural areas, rural area is taken as the reference category.

**ii. Mother's educational attainment**

Mother's educational attainment is recorded into three categories: "no education"(reference category), "primary education" and "secondary education and higher", respectively. From EDHS 2005 we can see that most of mothers 79.4 percent have no education and correspondingly their children are most likely to be stunted. Out of 3,140 cases considered in this study, only 38 mothers have attained secondary or higher level of education. But the table shows a child whose mother attained higher education level is less likely to be stunted.

**iii. Current employment status of mother**

Respondents were classified as those currently working and respondents who are not working (reference category). The survey report focuses on whether the mother was working at the time of the survey. Only 23.8 percent of those in the 2005 EDHS work for cash and overall 76.2 percent of women were not working or were not paid for work they do. Children of working mothers are slightly, more chance to be stunted (47.7%).

**Table 4.1 Empirical distribution of categorical variables in the study Vs stunting**

Background Characteristics		Categorical stunting status		Total
		Not Stunted	Stunted	
Type of place of residence	Rural	1,446 52.0%	1,337 48.0%	2,783 100.0%
	Urban	270 75.6%	87 24.4%	357 100.0%
Source of drinking water	Unprotected(Uncontrolled)	669 50.4%	658 49.6%	1,327 100.0%
	protected(controlled)	1,047 57.7%	766 42.3%	1,813 100.0%
Mother Education level	No education	1,304 52.3%	1,188 47.7%	2,492 100.0%
	Primary	286 59.1%	198 40.9%	484 100.0%
	Secondary or higher	126 76.8%	38 23.2%	164 100.0%
Partner's education level	No education	985 51.4%	932 48.6%	1,917 100.0%
	Primary	466 54.8%	385 45.2%	851 100.0%
	Secondary or higher	265 71.2%	107 28.8%	372 100.0%
Respondent currently working	No	1,325 55.4%	1,067 44.6%	2,392 100.0%
	Yes	391 52.3%	357 47.7%	748 100.0%
Birth order number	First to third	670 57.0%	506 43.0%	1,176 100.0%
	Above third	1,046 53.3%	918 46.7%	1,964 100.0%
Sex of child	Female	858 56.0%	675 44.0%	1,533 100.0%
	Male	858 53.4%	749 46.6%	1,607 100.0%
Preceding birth interval	less than 24 months	349 48.1%	377 51.9%	726 100.0%
	between 24 & 36	601 53.2%	528 46.8%	1,129 100.0%
	greater than 36 months	766 59.6%	519 40.4%	1,285 100.0%
Household economic status	poor	716 50.0%	716 50.0%	1,432 100.0%
	Middle or higher	1,000 58.5%	708 41.5%	1,708 100.0%

Overall	Stunted	Not Stunted	Total
	1,716 54.6%	1,424 45.4%	3,140 100.0%

#### **iv. Birth interval (Preceding birth interval)**

The period of time between two successive live births is referred to as birth interval. Some research has shown that children born soon after a previous birth are at greater risk of illness and death than those born after a long interval. A short birth interval is defined to be no longer than 24 months (reference category) and it is associated with high morbidity, exhausting the mother. As can be seen from Table 4.1 as the gap between successive birth increases the likely of a child being stunted will decrease.

#### **v. Sex of Child**

Among the 3,140 cases those examined in this study 1,533 (48.8 %) are female children and 52.2 % are male children. As we can see Table 4.1 male children are more likely to be stunted the female children.

#### **vi. Source of drinking water**

From the descriptive analysis in this study we have observed 1327(42.3%) of households use unprotected (uncontrolled) source of drinking water and protected source from tap water tap water which is scare and costly. As displayed in Table 4.1, 1,327 (42.3%) of households use unprotected (uncontrolled) source of drinking water. A child in a household which uses uncontrolled source of drinking water is more likely to be stunted. The report of UNICEF in 2002 indicates that more than half of the world's population used water from a piped connection at home. Moreover, 92% of the urban population and 70% of rural population in developing countries use improved drinking water source. Source of drinking water is recorded with respect to

the quality of water, whether it is obtained from unprotected source like public well, spring, river stream, pond, lake or rainwater is not controlled. Tanker water is assumed to contain uncontrolled water because it is scarce and costly.

#### **vii. Birth order**

In some of previous studies, the order of birth has been recorded into four categories assumed that a higher order births are associated with a high risk of mortality. In this work, birth order is recorded into two categories: first to third birth, and higher. As we can see from the descriptive output in most of the household the number of children ever born are greater than three. Since the output shows 62.5 % of children birth order is more than three. In addition to this we can easily see from Table 4.1 as the Birth order number increase the likely to be stunted will increase.

#### **viii. Household socio-economic characteristics**

The economic status of a household is an indicator of access to adequate food supplies, use of health services, availability of improved water sources, and sanitation facilities, these are prime determinants of child nutritional status (UNICEF, 2004). The five possible responses: poorest, poor, middle, rich and richest, for wealth index in EDHS 2005 are now categorized in this study into two groups as poor (poorest and poor) and middle or higher (middle, rich and richest). Among the 3,140 cases examined here 1,432(45.6%) are poor. And we can also say from Table 4.1 stunted children are more likely from poor family.

#### **ix. Partner's education**

As stated in the literature review, it is believed that when the partner (most probably the household head) is educated then the household income will increase as a result, a

child from this family most likely has well nutritional status. Accordingly, the likely of being stunted for this child decreases. As per the descriptive result in Table 4.1, 61.1 percent of the husbands (partners) have no education. In addition, as the educational level of the partner increased the likely of being stunted is decreased.

## 4.1.2 Continuous Covariates

### Response variable: Height-for-age (Zscore)

To build a regression model for nutritional status, we first need to check the distribution for the response variable. In the present example, it seems reasonable to assume that the Z-score is (at least approximately) normally distributed and, thus, model (3.4) could in principle be applied. Figure 4.1 shows a histogram and kernel density estimates of the distribution of the Z-scores, together with a normal density, with mean and variance estimated from the sample. The plots in Figure 4.1 suggest that the Zscore can be reasonably approximated by a Gaussian distribution. Thus, a Gaussian regression model is a reasonable choice for inference.

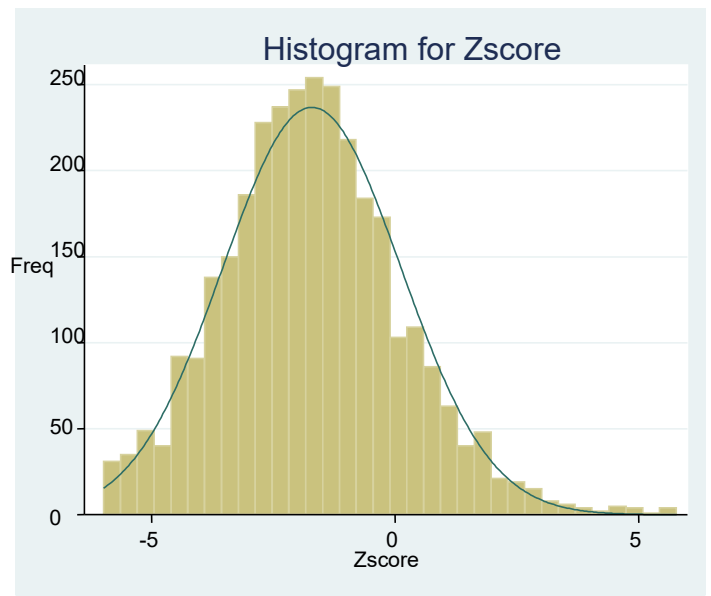
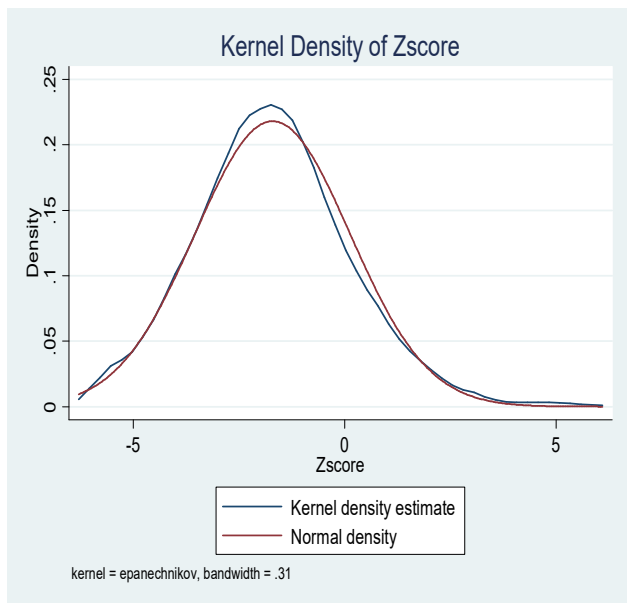


Figure 4.1 Kernel Density estimates of Zscore (left) and Histogram for Zscore (right)

**i. Age of a Child (Cage)**

Regarding the age of a child, a nonlinear, monotonically decreasing form of relationship is expected since children are usually born with almost normal anthropometric status. Afterwards, the health status of the children is expected to worsen for a certain time until it stabilizes at a low level. However, the exact shape of the influences is unknown and, hence, no simple model can be established to link the nutritional status scores to the age of the child. This effect will be explored by a nonparametric function later. The plot of mean *height-for-age* scores versus child age is given in Figure 4.2.

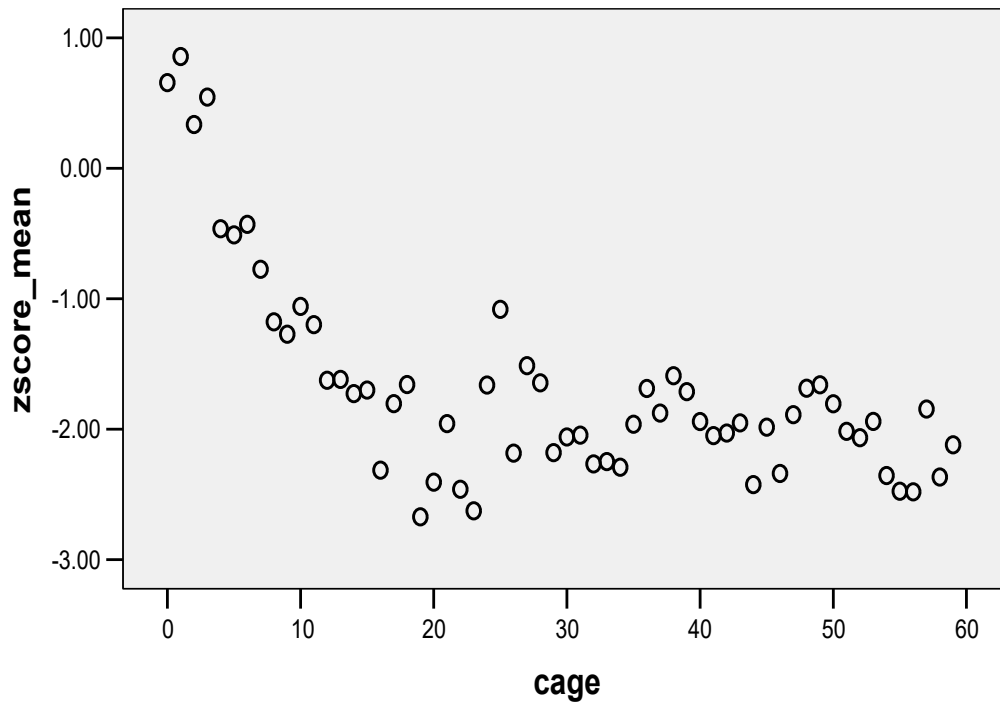


Figure 4.2 Scatter plots of mean Zscore Vs child age

## ii. Body Mass Index (BMI)

The mother's body mass index is defined as the weight in kilo-grams divided by the square of height in meters. Mothers with low BMI values are themselves malnourished and are therefore likely to have undernourished children. At the same time, very high BMI values indicate poor quality of food and, hence, may also imply malnutrition of the children. Since there is no definite pattern of relationship that can be observed from the scatter plot of body mass index (BMI) versus mean *height-for-age (zscore)* presented in Figure 4.3, this relationship can be explored by a nonparametric function.

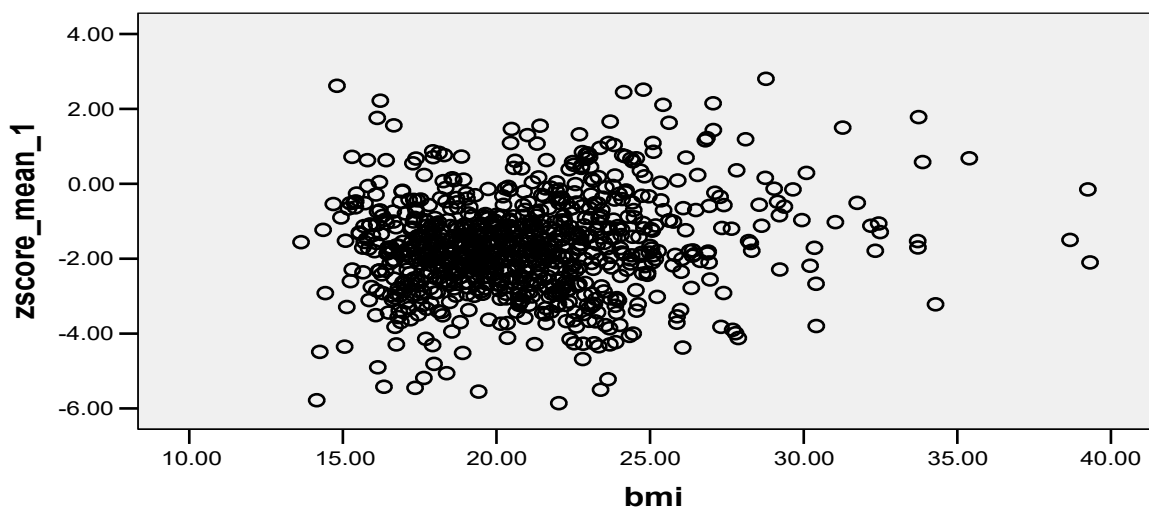
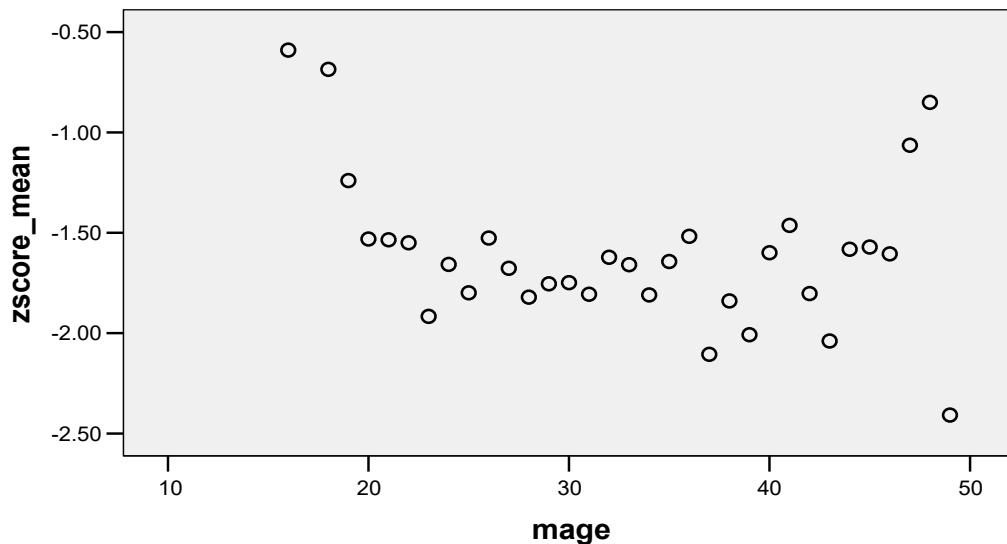


Figure 4.3 Scatter plots of mean Zscore Vs Body mass index

## iii. Mother's age at child's birth (Mage)

The effect of mother's age at child birth may be explored by categorizing the three age groups respectively as in some previous studies: young mothers (less than 22 years old),

middle-age group (between 22-35 years old), and old age group (greater than 35 years old).



*Figure 4.4 Scatter plots of mean Zscore Vs Mother age at birth*

However, in our application we include this covariate as a continuous covariate to have more reasonable results. As we can from Figure 4.4 mother age has a non linear effect on children nutritional status.

## **4.2. Determinants of Children malnutrition using Bayesian structured Additive regression models**

In this section we analyze the predictors using Bayesian approach by employing three models discussed in the methodology section. Starting with very simple models, we increase the complexity of model building to show what can be gained by more sophisticated approaches, and then we finish with the analysis using model selection based on Deviance information criterion.

### 4.2.1 Bayesian linear regression (parametric) model.

In this model we assume that each (covariate continuous or categorical) included has linear effects on the response variable, height-for-age Z-score and hence the effect of the covariates on the response is modeled by a linear predictor as:

$$\eta_i = \beta_o + X_i' \beta + W_i' \gamma$$

In this model:

$X_i = (x_{i1}, \dots, x_{ip})'$ , is a vector of continuous covariates that includes Body Mass Index (BMI), Child age (Cage), and Mother age at birth (Mage).

$W_i = (w_{i1}, \dots, w_{ip})'$ , is a vector of categorical covariates that includes Type of residence (Res), Mother Education (Medu), Source of drinking water (Swat), Birth order (BORD), Sex of child (Csex), Previous birth interval (Pbint), Mother work (Mwork), Household economic status (HHeco) and Partner Education (Pedu).

$\beta$ 's and  $\gamma$ 's are coefficients of continuous and categorical covariates respectively and  $\beta_o$  is a constant term.

That is the linear predictor can be rewritten as:

$$Zscore_i = \beta_o + BMI \beta_1 + Cage \beta_2 + Mage \beta_3 + Res \gamma_1 + Medu2 \gamma_2 + Medu3 \gamma_3 + Swat \gamma_4 + BORD \gamma_5 + Csex \gamma_6 + Pbint2 \gamma_7 + Pbint3 \gamma_8 + Mwork \gamma_9 + HHeco \gamma_{10} + Pedu2 \gamma_{11} + Pedu3 \gamma_{12}$$

The code to analyze the above model in BayesX is:

```
>b.regress Zscore = Res + Medu + Swat + BORD + Csex + Pbint + Mwork + HHeco + Pedu  
+ BMI + Cage+ Mage, family=gaussian iterations=12000 burnin=2000 step=10 predict  
using d
```

where *b* and *d* are BayesX data and regression objects respectively, *Zscore* is the response variable and the others are covariates and simulation controller options. The options “iterations”, “burnin” and “step” define properties of the MCMC algorithm. The total number of MCMC iterations is given by “iterations” while the number of burn in iterations is given by “burnin”. Therefore, we obtain a sample of 10000 random numbers with the above specifications left the first 2000 burn in iterations. Since, in general, these numbers are correlated; we do not use all of them but thin out the Markov chain by the thinning parameter *step*. Specifying *step*=10 as above forces *BayesX* to store only every 10th sampled parameter which leads to a random sample of length 1000 for every parameter in our example. Note, that the choice of iterations of course also affects computation time.

From the above BayesX procedure using MCMC simulation provides a number of output that are referred in Annex, for example, for the semiparametric Bayesian model deviance information criteria detail is attached. However, in order to compare the three models here we only consider values of Deviance Information Criterion.

The DIC results for Model I: the Bayesian regression model are

**Table 4.2 Estimation Result for Deviance for Model I**

Deviance:	12366.369
pD:	13.958635
DIC:	12394.286

displayed in Table 4.2

#### 4.2.2 Bayesian linear regression (parametric) model with some transformation

This model is similar to the above model except it transforms one of the continuous covariates Cage (Child age). Since here as we checked the effect of Cage on the response variable Zscore seems quadratic but the other continuous covariates do not show a known pattern on the graphical displays with the respondent variable, we include the square of the covariate age as one more covariate in the above model and we named it as Bayesian linear regression (Parametric) model with some transformation. As we did in the above model, the DIC results for Model II: Bayesian linear regression with some transformation is presented in Table 4.3.

*Table 4.3 Estimation Result for Deviance for Model II*

Deviance:	12253.815
pD:	14.857979
DIC:	12283.531

#### 4.2.3 Bayesian semiparametric regression model

Now we consider the model which seems more flexible than the above two models. Here we treat continuous covariates related to the response variable nonparametrically with the help of smoothing functions  $f(\cdot)$ , whereas the categorical variables are related parametrically to the response variable. Since we have parametric and nonparametric forms of relationship in the model, it is referred to as Bayesian semiparametric regression model.

The semiparametric regression model representation is given as:

$$\mu_i = h(\eta_i), \quad \eta = f_1(x_{i1}) + \dots + f_p(x_{ip}) + w_i' \gamma$$

Here  $h$  is a known response function in this case the identity function, and  $f_1, \dots, f_p$  are possibly nonlinear smooth functions of continuous covariates.

Using the variables included in this study, this model can be rewritten as:

$$Zscore_i = \beta_o + f_1(BMI) + f_2(Cage) + f_3(Mage) + Res\gamma_1 + Medu2\gamma_2 + Medu3\gamma_3 + Swat\gamma_4 + BORD\gamma_5 + Csex\gamma_6 + Pbint2\gamma_7 + Pbint3\gamma_8 + Mwork\gamma_9 + HHeco\gamma_{10} + Pedu2\gamma_{11} + Pedu3\gamma_{12}$$

The code to analyze the above model is:

```
> b.regress Zscore = Res + Medu + Swat + BORD + Csex + Pbint +
Mwork +HHeco + Pedu + BMI(psplinerw2) + Cage(psplinerw2) +
Mage(psplinerw2), family=gaussian iterations=12000 burnin=2000
step=10 predict using d
```

Where, psplinerw2 under the parentheses in the continuous covariates represent the smoothing function Bayesian P-spline with second order random walk. The rest are as described in Model I above.

The result of DIC for model III: Bayesian semiparametric regression model is presented in Table 4.4

**Table 4.4 Estimation Result for Deviance for Model III**

Deviance:	12123.35
pD:	31.176825

DIC:	12185.703
------	-----------

### 4.3 Model Selection

As discussed in the methodology part under model selection section a widely used statistic for comparing models in a Bayesian framework is the Deviance Information Criterion. All the three models DIC is computed here put together for comparison.

*Table 4.5 Estimation Result for Deviance for the three Models*

Model	Deviance	pD	DIC
Parametric	12366.369	13.958635	12394.286
Parametric with some transformation	12253.815	14.857979	12283.531
Semiparametric	12123.35	31.176825	12185.703

#### 1. Deviance

Deviance is a measure of how well the model fits the data; the larger this is the worse the fit. In our case, from Table 4.5 we can see that semiparametric model is the one with smallest Deviance and hence this model well fits the data.

#### 2. pD (Effective number of parameters in the models)

The model complexity is measured by pD (The effective number of parameters in the model), the larger the pD is the easier to fit the data. Based on this fact, since the third model has largest value for this measure, it is selected again.

#### 3. DIC (Deviance information Criteria)

DIC is a composite measure of how well the model does (it is a compromise between fit and complexity), and small value of DIC is preferred. A difference of more than 7 to 10 units is regarded as strong evidence in favor of the model with the smaller DIC. As we can see in Table 4.5 the semiparametric model has got the smallest DIC value and also the difference in both of the other two models is much greater than 10 as a result we have a very strong evidence to choose the third model semiparametric as the best model. Thus, based on the above three strong evidences we select the Bayesian semiparametric regression model as the best model for the data under consideration.

#### 4.4 Results of Bayesian semiparametric model

The analyses were carried out using the freeware software BayesX. And all results discussed under this section are a result of the Bayesian semiparametric regression model which is the one selected in the above section. The semiparametric predictor used is of the form.

$$Zscore_i = \beta_o + f_1(BMI) + f_2(Cage) + f_3(Mage) + Res\gamma_1 + Medu2\gamma_2 + Medu3\gamma_3 + Swat\gamma_4 + BORD\gamma_5 + Csex\gamma_6 + Pbint2\gamma_7 + Pbint3\gamma_8 + Mwork\gamma_9 + HHecoy_{10} + Pedu2\gamma_{11} + Pedu3\gamma_{12}$$

The detail results of this model and BayesX codes are given in the Annex. The BayesX code to analyze the above model is found in Annex A1, MCMC simulation found in Annex A2, the estimation result for Deviance (Annex A3), the posterior mean for fixed effects (Annex A4), estimation result for smoothing parameters (Annex A5), plot of autocorrelation function (Annex B1), sampling path for Mother's Body Mass Index (Annex B2), sampling path for Child age (Annex B3) and sampling path for Mother's age at birth (Annex B4).

##### 4.4.1. Fixed (linear) effects

The result for categorical covariates is summarized in Table 4.6. The table gives posterior means along with their posterior standard deviations, posterior median and 90% credible intervals. As can be seen from Table 4.6, based on the 90% credible intervals that does not include the zero values, variables type of residence (Res), source of drinking water (Swat), birth order (BORD), (sex of child) Csex, previous birth interval (pbint) mother work (Mwork) and partner education (Pedu) are found statistically significant and variables household economic status (HHeco) and mother education (Medu) are not significant.

From the table, we can observe that children born in rural areas were more likely to be malnourished. The analysis shows that children born after a long birth interval (> 24 months) were better off than other children. Children from household who use controlled source of drinking water have got better nutritional status. An educated partner (at least primary education) contributes to better nourishment of the child in that family. It also is the case that children living with both parents (married women) were less stunted than other children, apparently benefiting from extra care of educated partner. Alternatively couples may benefit from economies of scales for child care as well as in expenditures (Kandala et al, 2001; Kandala et al. 2007).

The analysis also shows that female children were slightly less malnourished which had also been found in other studies (Mohammed, 2008; Kandala et al., 2001; Kandala et al., 2007). On the other hand, we observe that malnutrition is higher in those children whose mothers are working at the time of interview or twelve months before the interview period, children of higher birth order (other than first born). Unexpected result is found for household economic status (HHeco) it is found statistically not significant. Mother education has found not significant. There is a confirmation in previous study; the educational level of the mother has a slight impact on the level of

stunting (Khalid, 2007). In most of previous studies mother education is turned to significant when the level is secondary or higher education (Mohammed, 2008). But as we can see in the descriptive output that most of mothers have no education, only 23.2 % in secondary or higher education, as a result one of the most important variables in the literature turned out to statistically not significant.

**Table 4.6 Posterior Results of the categorical covariates**

Variable name	Pmean	Pstd	10% quantile	Pmedian	90 % quantile
const	-1.92*	0.17	-2.14	-1.93	-1.72
Type of residence(Res)	0.61*	0.12	0.47	0.61	0.75
Mother education Medu(ref)	-	-	-	-	-
Mother education (Medu1 :ref)	0.05	0.09	-0.07	0.05	0.17
Mother education (Medu2)	0.07	0.17	-0.15	0.07	0.29
Source of drinking water (swat)	0.25*	0.07	0.17	0.25	0.34
Mother work (Mwork )	-0.11*	0.07	-0.2	-0.11	-0.02
Previous birth interval (Pbint1 :ref)	-	-	-	-	-
Previous birth interval (Pbint2 )	0.08*	0.08	-0.02	0.08	0.19
Previous birth interval(Pbint3)	0.22*	0.08	0.12	0.22	0.32
Birth order (BORD)	-0.11	0.08	-0.22	-0.11	0.00
Household economy (HHeco)	-0.08	0.07	-0.16	-0.08	0.01
Partner education(Pedu1: ref)	-	-	-	-	-
Partner education(Pedu2)	0.07	0.08	-0.03	0.07	0.17
Partner education(Pedu3)	0.44*	0.12	0.28	0.44	0.59

#### 4.4.2. Continuous (non linear) effects

Nonlinear effects represented by smoothed functions, are commonly interpreted graphically. Figure 4.4 shows the smooth function of the BMI of mother versus height-for-age Zscore. In the figure the posterior means together with 80% and 95% pointwise credible intervals are shown. In looking at the mother's BMI and its impact on the level of nutritional status, the Figure shows that the influence does not have a regular form.

We observe that the mothers with BMI in between 15 and 23 kg/m<sup>2</sup> have a lower corresponding nutritional status of children, and body mass index of 23 to 27 kg/m<sup>2</sup> have slightly higher z-score of *height-for-age* (lower nutritional status) measured by stunting, and the effect stabilizes at the same level thereafter. Mothers with BMI greater than 35 kg/m<sup>2</sup> have a lower z-score of *height-for-age*. The figure also reveals that obesity of the mother is likely to pose less of a risk to the nutritional status of a child. Low BMIs of less than 18.5 suggest acute undernutrition of the mother. Furthermore, the z-score is highest at a BMI of around 27 to 30 kg/m<sup>2</sup> (and thus lowest stunting or highest nutritional status). In general, the figure shows that BMI has a slight effect on the nutritional status

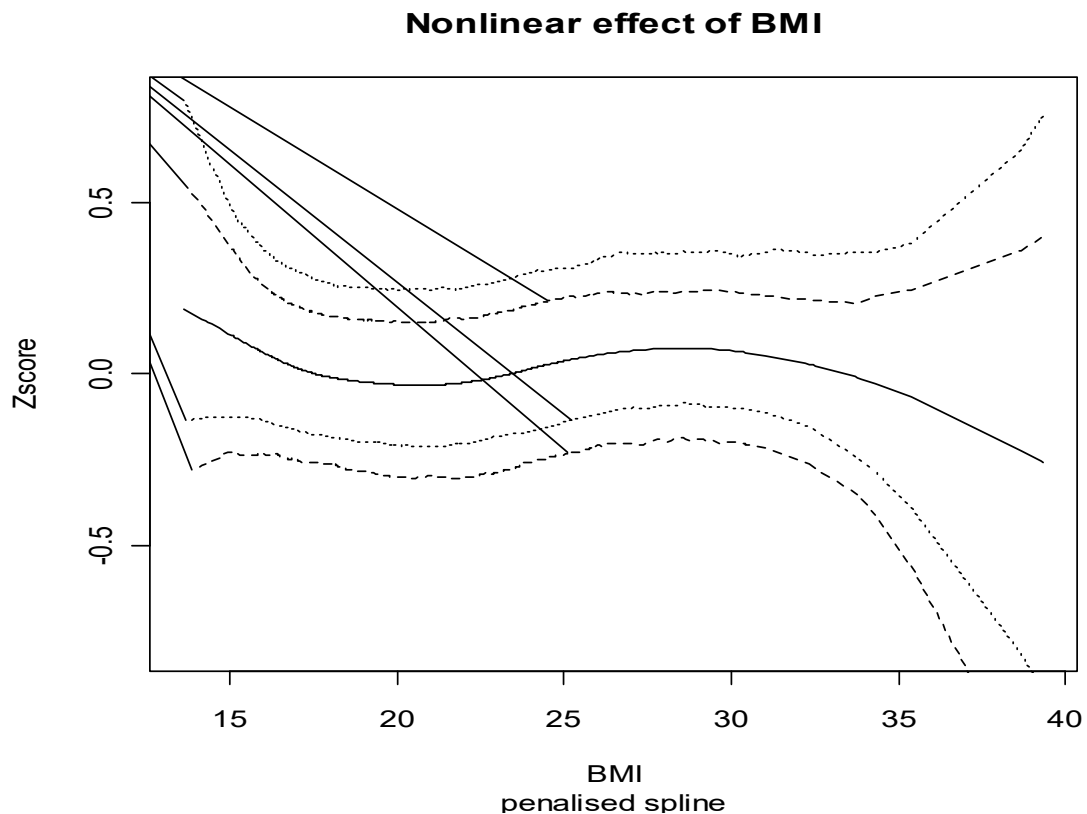


Figure 4.4 Non-linear effects of Mother Body Mass Index on nutritional status of a child

The other continuous covariate to be considered is child age. Figure 4.5 displays the nonparametric effect of the child's age in Ethiopia. Shown are the posterior means together with 80% and 95% pointwise credible intervals. We observe that the influence of a child's age on its nutritional status is considerably high in the age range between the age of 5 months and 20 months. This deterioration in nutritional status of a child begins around 1 month after birth and continues with an almost linear trend until the age of 20 months. Between the ages of 20 to 30 of months, stunting decreases, and stabilizes thereafter at a middle level with a bump till 5 years.

As suggested by the nutritional literature, we were able to discern the continuous worsening of the nutritional status up until about 20 months of age. This deterioration set in right after birth and continues, more or less linearly, until 20 months. After 20 months the effect of age on stunting stabilizes at a low level. Through reduced growth and the waning impact of infections, children were apparently able to reach a low-level equilibrium that allows their nutritional status to stabilize (Kalid, 2007; Mohammed, 2008; Kandala, 2001; Belitz et.al, 2007). This is picking up the effect of a change in the data set that makes up the reference standard. Until 24 months, the currently used international reference standard is based on white children in the US of high socioeconomic status, while after 24 months it is based on a representative sample of all US children (WHO, 1995).

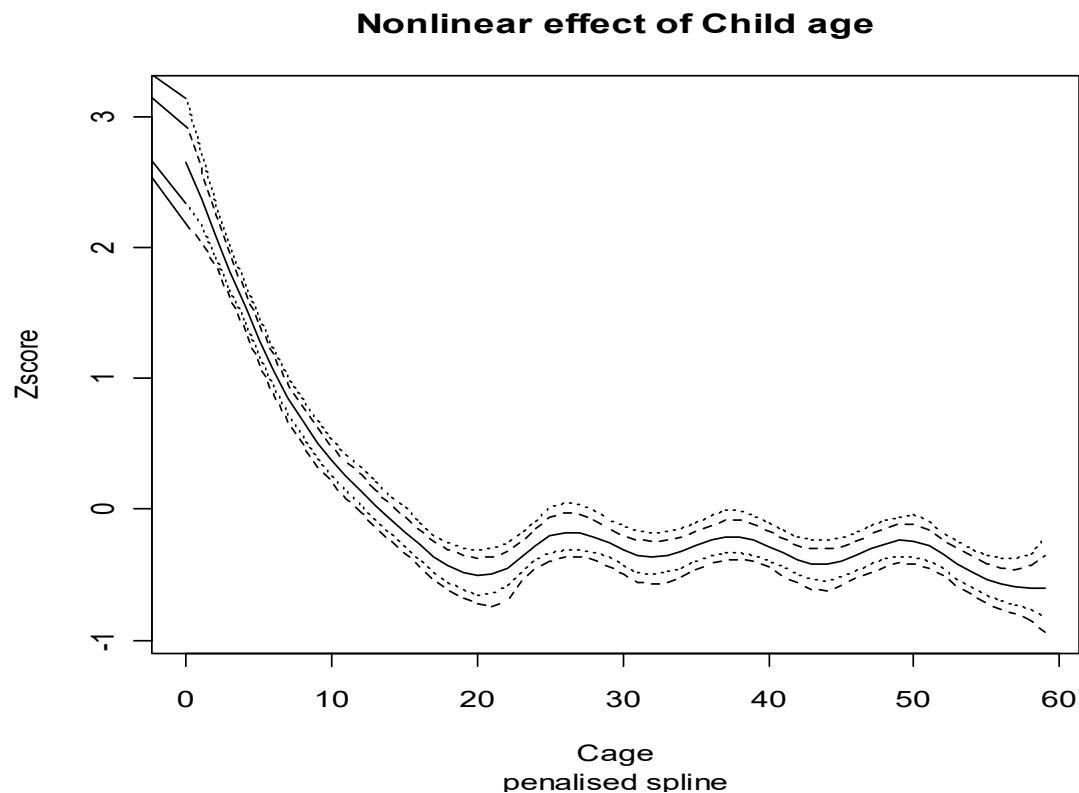


Figure 4.5 Non-linear effects of Child age on nutritional status of a child

The effect of mother's age on stunting is quite slight (figure 4.6). It shows that the *height-for-age* z-score is low for mother between the ages 15 to 30 years. The z-score of *height-for-age* increases (and nutritional status increases) after age of 30 years, the effect of the mother's age increases with an almost linear trend. This effect is also witnessed in previous studies, children whose mothers are older than 30 years of age are better in their nutritional status as compare to children whose mothers are in the younger age group. This finding is consistent with most of previous studies (Mohammed, 2007, Kandala, 2001).

### Nonlinear effect of Mother age at birth

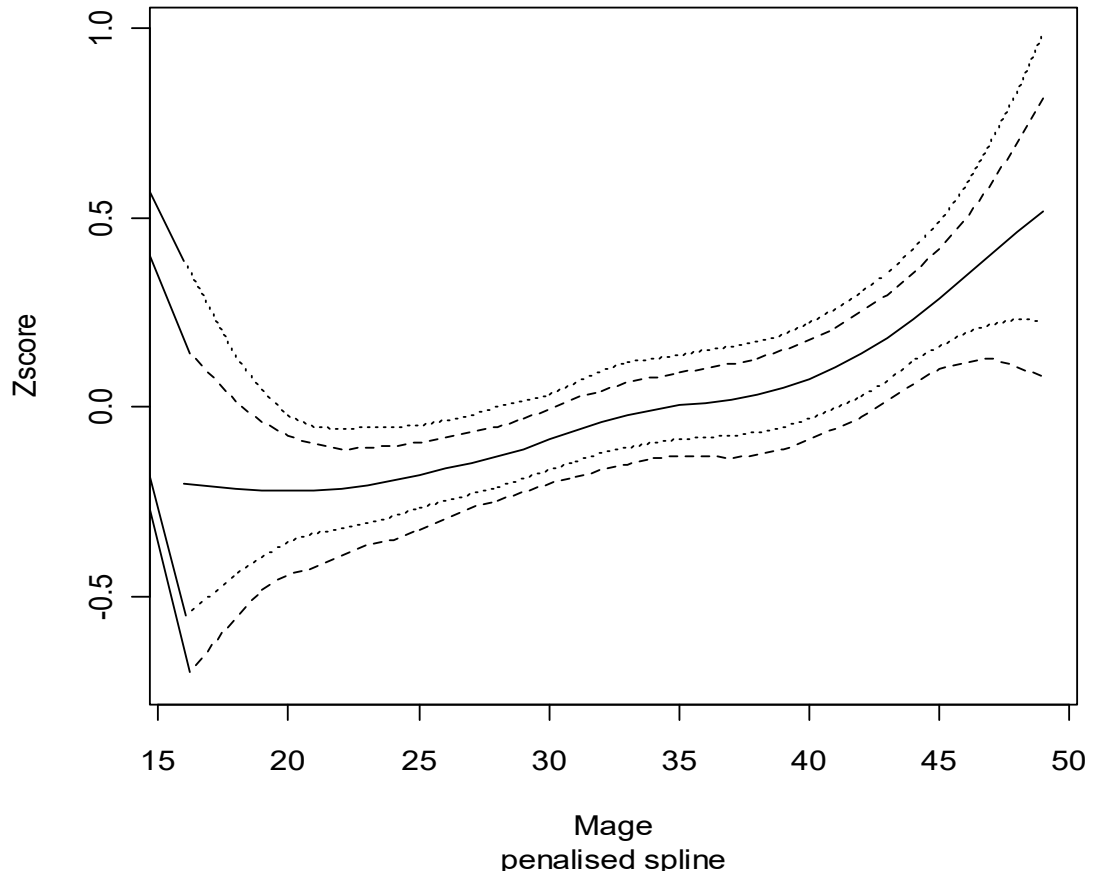


Figure 4.6 Non-linear effects of Mother age on nutritional status of a child

## CHAPTER FIVE

### 5. DISCUSSION, CONCLUSION AND RECOMMENDATION

#### 5.1 Discussion

In this section we discuss the findings of the present analysis for each covariate. In general, almost all of the findings in this study are consistent with most previous studies on children malnutrition, with few exceptional covariates like mother education and household economic status.

As shown in the analysis urban children are less likely to be malnourished than their rural counterparts because the quality of health environment and sanitation is better in urban areas, whereas, the living condition in rural areas are associated with poor health condition, use of unprotected water supplies, and lack of personal hygiene, which were the risk factors in determining malnutrition. This is consistent with some studies, where the mother lives, has a statistical significant effect on children nutritional status (Kandala et al, 2001; Woldemariam & Timotiows, 2002).

Maternal education, which is related to household wealth, is a determinant of good child-care knowledge and practices. In our analysis, the impact of education of mothers for primary as well as secondary and higher attainment is not significant. Most of mothers have no education and a very small number of mothers, 38 out of 3140 cases considered in this study, have secondary and higher education. Previous studies confirm that, at low levels of education, effects on malnutrition are small or negligible, and effect increases only at secondary or higher levels (Khalid, 2006; Mohammed, 2008). Hence, in this study for those few number of mothers under secondary or higher education group category we found the unexpected result.

Mother Work has a positive and significant effect on the malnutrition status (wasting and underweight) of children in Nigeria. However, it turned into a negatively significant effect on the children's undernutrition status in this study. This suggests that insufficient food intake of the child may be not influenced by the current working status of mothers. The results are consistent with some previous studies and not consistent with others. Some studies reported that when mothers are working, the household income is increased and access for better food will increase, as well as the access to a quality level of medical care. On the other hand, when mothers are employed outside the home, this has the effect that it curtails the duration of full breastfeeding and necessitates supplementary feeding. This result confirmed by some local studies (Dejen, 2008; Woldemariam & Timotiws, 2002) and some studies elsewhere arrived similar result (Khalid, 2006; Mohammed, 2008).

Preceding birth interval is also an important demographic variable that affects nutritional status of children. Our analysis shows that both birth interval groups (24 to 36 months and greater than 36 months) are found statistically significant. Thus as the preceding birth interval increases the nutritional status of a child increases. This finding is confirmed by most of previous studies (Kandela, 2001; Khalid, 2008; Mohammed, 2007; Dejen, 2008). The significant and higher risk of stunting among children of lower preceding birth interval could be due to uninterrupted pregnancy and breastfeeding, since this drains women's nutritional resources (Sommerfelt and Stewart, 1994). Close-spacing may also have a health effect on the previous child who may be prematurely weaned if the mother becomes pregnant too early again. In this study, rural children were found to be the most affected in their nutritional status due to close spacing and this may be due to the low contraceptive prevalence rate in the rural areas.

As described in the literature one of the very important environmental factors affecting children nutritional status is source of drinking water. The results indicate that the protected source of water has a positive statistical significance effect on child nutritional status. In other words, the source of water is associated with the nutritional status of a child through its impact on the risk of childhood diseases such as diarrhea, and is affected indirectly as a measure of wealth and availability of water. Use of improved water and sanitation has a lot of benefits: reduction of diseases (particularly diarrhea), avoids illness health-related costs; saves time associated with getting water, and that the sanitation facilities are located closer to home. As is well known presently there is a wide spectrum of waterborne disease such as cholera, trachoma, typhoid and paratyphoid. The most common disease is diarrhea, which can lead to morbidity and in many cases mortality. This finding is consistent in almost all of the literatures.

In the analysis, it was discovered that the stunting of children increases gradually from 5-20 months of age, where the minimum Z-scores of stunting is attained, then rises again through the remainder of the third year. The deterioration in nutritional status is set between 5-20 months of age, as reported in much of the literatures, due to supplementation. However, it reaches its minimum level between age 20 and 22 months, then rises again and reaches its minimum level between age 26 and 28 months and stabilizes thereafter at a middle level with a bump till 5 years. Previous studies also confirm this (Kandala, et. al. 2007; Khalid, 2006; Mohammed, 2008).

In looking at the mother's BMI and its impact on the level of stunting, we observe that the mothers with BMI between 23 and 27 have a slightly higher z-score of height-for-age (lower stunting) measured by stunting, and the effect stabilizes at the same level thereafter. Mothers with BMI less than 20 have a lower z-score of height-for-age. It shows that BMI has a slight effect on the nutritional status. The nonparametric effect of

BMI reveals that obesity of the mother is likely to pose less of a risk to the nutritional status of a child. Low BMIs of less than 18.5 suggest acute undernutrition of the mother. Furthermore, the z-score is highest at a BMI of around 27 to 30 (and thus lowest stunting or highest nutritional status. This finding also consistent in some of studies in Africa (MoFED, 1999; Sahn and Stifel, 2002; Christiaensen and Alderman, 2004) a mother's nutritional status affects her ability to successfully carry, deliver, and care for her children and is of great concern in its own right. Women who are malnourished (thinness or obesity) may have difficulty during childbirth and may deliver a child who can be malnourished. The results indicate that there is an association between the thinness condition of the mother and the nutritional status of the child. The same finding is also found in a number of studies. Mothers with low BMI on average giving birth to babies of low birth weight (Teller & Yimer, 2000; Woldemariam & Timotiows, 2002; Khalid, 2006; Mohammed, 2008)

The influence of age of mothers who are younger than 20 years is higher on the nutritional status of children in Ethiopia. It also shows that children whose mothers are older than 30 years of age are better in their nutritional status compared to children whose mothers are in the younger age group. Possible causes for this are due to childbirth among very young girls, whose bodies are not physically ready to endure the processes of childbirth. The problem is compounded by the fact that some African countries have poor obstetric care (Khalid, 2006). Furthermore, these mothers could not reach health facilities, or when they do, it is too late. Effective ways must be devised to delay age at first marriage and first birth. In addition, one study obtained in Nigeria reported that younger mothers (teenagers) are less likely in comparison to older mothers to breastfeed their children after birth, which means that the age of the mother at birth of a child influences whether the child will receive colostrums or not, which might affect the nutritional status of children (Adebayo, 2004).

## 5.2 Conclusions

The aim of the current study was to identify the most dominant predictors of children nutritional status in Ethiopia. As an indicator of nutritional status, we used the anthropometric indicator of *height-for-age* Zscore. Using Bayesian semiparametric regression model, we conclude that overall, children are more likely to be better nourished (less stunted) if the preceding birth interval is large (more than 24 months), age of the mother at birth is higher, the household uses protected source of drinking water and having secondary school level or high level educated partner. On the other hand, higher birth order (not being first-born), a working mother, residing in rural area contribute to worsened nutritional status. We also observe that male children are more undernourished than female children.

It is found that children are at high risk during the first 1-20 months of their life and then stabilize moderately with bump. The effect of BMI on the child's nutritional status is slight. It shows a child is in a good nutritional status when his/her mother body mass index is in between 20 to 30 kg/m<sup>2</sup>, which means that there is an association between the thinness condition of mothers and nutritional status. According to the mother's age at birth, it shows that younger mothers are less likely to affect their children's nutritional status positively.

### **5.3 Recommendation**

We found that the dominant predictors of children nutritional status are associated with household living style, type of residence and awareness about health and family planning programs. Thus, our recommendations given below are along these predictors.

Integrate nutrition and family planning. This can help by delaying birth to ensure optimal growth (physical, psychological and emotional), and to allow mothers attain their highest level of education and therefore improve the socioeconomic status of mothers and children. Give more attention to rural areas which have high rates of poverty, and problematic area in accessing health facilities. These areas are more likely to have a higher proportion of undernutrition compared to urban areas, due to poor health facilities and complications during childbirth or even careless and misdiagnosis during hospital care. Therefore, the most important issues to address in these areas are health care, proper food, and raising the educational level of parents.

Governments should improve socioeconomic conditions. Because, if living standards are improved, there will be better health care and a reduction in infant and, child diseases, child malnutrition and child mortality. Research must be done on ways to improve the nutritional status of households in these countries using indigenous inexpensive foods that are locally available. Although female children are less likely stunted in most of the studies, researches on female child nutrition are still very important. There is still a need for research and studies about nutrition and the important components of healthy eating to avoid the increase of illness caused by poor eating habits.

In general, most of the findings are related to household living style (family-based). We recommend focusing on creating awareness at household level about health and family planning.

## REFERENCES

- Abbi, R., P. Christian, S. Gujral, and T. Gopaldas. (1991). *The impact of maternal work on the nutrition and health status of children*. Food and Nutrition Bulletin 13(1):20-24.
- ACC/SCN (2000). *Fourth report in the world nutrition situation*. Geneva: ACC/SCN in collaboration with the International Food Policy Research Institute.
- Adebayo. B (2004). *Spatial Modelling of Initiation and Duration of Breastfeeding: Analysis of Breastfeeding Behaviour in Malawi*. World Health and Population Vol. 10 NO.3 2009.
- Aschalew, G. (2000). *Determinants of nutritional status of children in Amhara Region: A case study of Misrak Gojjam and Semen Wello Zones*. M.Sc Thesis, DTRC/IDR, AAU.
- Belitz, C., Hübner, J., Klasen, S. and Lang, S. (2007). *Determinants of the Socioeconomic and spatial pattern of undernutrition by sex in India: A geo-additive semiparametric regression approach*. Working paper, Department of Statistics, University of Munich.
- Belitz.C,Brezger.A ,Kneib.T,Stefan.L (2009) *BayesX - Software for Bayesian Inference in Structured Additive Regression*. Department of Statistics, Ludwig-Maximilians-University Munich.
- Brezger, A. and Lang, S. (2006). *Generalized Structured Additive Regression based on Bayesian P-Splines*. Computational Statistics and Data Analysis, 50, 967-991.
- Brezger, A., Kneib, T., Lang, S. (2003). *BayesX—Software for Bayesian Inference Based on Mixed Model Methodology*. Available under <http://www.stat.uniuenchen.de/~lang/BayesX>.
- Christiaensen, L., & Alderman, H. (2001). *Child Malnutrition in Ethiopia: Can Maternal Knowledge Augment the role of Income?*, World Bank, 2001.
- Central Statistical Authority (CSA) [Ethiopia] and ORC Macro. (2001). *Ethiopian*

- Demographic and Health Survey*. Ababa, Ethiopia and ORC Macro Calverton, Maryland, USA
- Central Statistical Authority (CSA) [Ethiopia] and ORC Macro. (2006). *Ethiopian Demographic and Health Survey*. Addis Ababa, Ethiopia and ORC Macro Calverton, Maryland, USA.
- Charles H , Dorosty R, Reilly J (2008). *Coexistence of social inequalities in undernutrition and obesity in preschool children: population based cross sectional study*. Caledonian University. UK.
- Dejen. A (2008). *The determinants of nutritional status of children in Ethiopia using Multivariate analysis*. AAU, MS.c thesis in applied statistics.
- Eilers, P.H.C. and Marx, B.D.(1996). *Flexible smoothing with B-splines and penalties (with comments)*. Statistical Science 11, 89-121.
- Fahrmeir, L and Lang, S. (2001). *Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors*. Applied Statistics (JRSS C), 50, 201-220.
- Genebo, T., W. Girma, J. Hadir, and T. Demmissie (1999). *The association of children's nutritional status to maternal education in Ziggbaboto, Guragie Zone South Ethiopia*. Ethiopian Journal of Health Development 13(1):55-61.
- Getaneh, T., A. Assefa, and, Z. Taddesse.(1998). *Protein energy malnutrition in urban children: Prevalence and determinants*. Eth. Med. J. 36(3).
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Kammann, E. E. and Wand, M. P. 2003. *Geoadditive Models*. Journal of the Royal Statistical Society C, 52, 1-18.
- Helen Keller International (2001). *The Nutritional Surveillance Project in Bangladesh in 1999. Towards the Goals of the 1990 World Summit for Children*. Dhaka, Bangladesh: Helen Keller International.

- Henry Wamani, Anne Nordrehaug Åstrøm, Stefan Peterson, James K Tumwine, and Thorkild Tylleskär (2007). *Boys are more stunted than girls in Sub-Saharan Africa: a meta-analysis of 16 demographic and health surveys*.
- Hobert, J. P. and Casella, G. (1996). *The effect of improper priors on Gibbs sampling in hierarchical linear mixed models*. Journal of the American Statistical Association 91, 1461-1473.
- Kandala, Fahrmeir, Klasen, Priebe (2007) *Geo-additive models of Childhood Undernutrition in three Sub-Saharan African Countries*, Clinical Sciences Research Institute, Warwick Medical School, University of Warwick, CV2 2DX, and Coventry, UK.
- Kandala, N.B., S. Lang, Klasen, S. and Fahrmeir, L. (2001). *Semiparametric Analysis of the Socio-Demographic Determinants of Undernutrition in Two African Countries*. Research in Official Statistics, EUROSTAT, Vol. 4 No.1:81-100
- Klasen, S. (2008). *Poverty, undernutrition, and child mortality: Some inter-regional puzzles and their implications for research and policy*. Journal of Economic Inequality, (1): 89-115.
- Khaled K. (2007). *Analysis of Childhood Diseases and Malnutrition in Developing Countries of Africa*. Munchen.
- Lang, S. and Brezger, A. (2004). *Bayesian P-splines*. Journal of Computational and Graphical Statistics 13, 183-212.
- Leslie, J. (1988). *Women's work and child nutrition in the third world*. World Development 16(1).
- Maleta, K. (2006). *Epidemiology of Undernutrition in Malawi*, chapter 8 in The Epidemiology of Malawi, edited by Eveline Geubbles and Cameron Bowie, available on <http://www.medcol.mw/commhealth/publications/epi%20book/Undernutrition%20chapter%208.pdf> .
- Micheal Adam (2000). *The relationship between nutritional status of children and KAP of mothers' feeding their children in a Rural Community: Hawa-Michael, Bugna Woreda*,

- and Southern Wollo*. MSc. Thesis, AAU.
- Mishra, V., S. Lahiri, and N. Luther (1999). *Child Nutrition in India*. National family health survey subject reports number 14, International Institute for Population Sciences. Mumbai, India. East-West Center, Population and Health Studies, Honolulu, Hawaii, U.S.A.
- Ministry of Planning Economic Development (1999). *Poverty situation in Ethiopia*. Welfare Monitoring Unit, Addis Ababa, Ethiopia.
- Mohammed (2008). *Gender differentials in mortality and undernutrition in Pakistan*. Peshawar (Pakistan).
- Nelder, J. A. and Wedderburn R. W. M. (1972). *Generalized linear models*. Journal of the Royal Statistical Society Series A 135, 370-384.
- Olalekan A. Uthman (2008). *The Internet Journal of Nutrition and Wellness*. Volume 6 Number 2
- Rue, H. (2000). *Fast Sampling of Gaussian Markov Random Fields with Applications*. Technical report No. 1, University of Trondheim.
- Silva .P (2005). *Environmental factors and children's malnutrition in Ethiopia*. Policy Research Working Paper Series from The World Bank No 3489.
- Pelletier, D.L., E.A. Frongillo, Jr., D.G. Schroeder, and J.P. Habicht. (1994). *A methodology for estimating the contribution of malnutrition to child mortality in developing countries*. Journal of Nutrition 124 (124 suppl.): 2106S-2122S.
- Samson, T. and G. Lakech. 2000. *Malnutrition and enteric parasites among under five children in Aynalem village, Tigray*. Ethiopian Journal of Health Development 14(1):67-75.
- Sahn D.E., & Stifel, D.C. (2003). *Urban-rural inequality in living standards in Africa*. Journal of African Economies, 12, 4, 564-597.
- Spiegelhalter D., Best N., Carlin B., and Van der Line A. (2002). *Bayesian measures of models complexity and fit*. Journal of the Royal Stat. Soc. B(64):1-34.
- Smith,L.and L.Haddad.(2001). *The Importance of Women's Status for Child Nutrition in Developing Countries*. Mimeographed, IFPRI. Wash- ington DC: IFPRI.

- Sommerfelt and Stewart (1994). *Children's Nutritional Status*. DHS Comparative Studies No. 12 Calverton, Maryland; Macro international Demographic and Health Surveys (DHS).
- Teller, H., and. Yimar. G. (2000). *Levels and determinants of malnutrition in adolescent and adult women in southern Ethiopia*. Ethiopian Journal of Health Development 14(1):57-66.
- United Nations (2004). *Food Present Report*. United Nations, March 29, 2004, p. 6.
- United Nations Children's Fund (2008). *Latest state of world children report*. New York: UNICEF.
- UNICEF (2006). *UNICEF Annual Report 2006*. UNICEF.
- UNICEF (2004) *The State of the World's Children 2004 .Annual flagship report*. New York: UNICEF.
- World Health Organization (1986). *Use and Interpretation and anthropometric indicators of nutritional status*. Bull. WHO.
- World Health Organization (1995). *Physical Status: The Use and Interpretation of Anthropometry*. WHO Technical Report Series No. 854. Geneva: WHO.
- World Health Organization (2002). *Comparative risk assessment of nutritional status of children*. WHO.
- World Health Organization (2006). *The world health report 2006 - working together for health*. WHO.
- Woldemariam Girma and Timotiows Genbo (2002). *Determinants of the Nutritional Status of Mothers, and Children in Ethiopia*. Calverton, Maryland, USA: ORC Macro.

Yimer, G. (2000). *Malnutrition among children in southern Ethiopia: Levels and risk factors*. Ethiopian Journal of Health Development 14(3):283-292.

Zewditu, G., U. Kelbessa, G. Timotiows, and N. Ayele (2001). *Review the status of malnutrition and trend in Ethiopia*. Ethiopian Journal of Health and Development 15(2):55-74

## ANNEX

### Annex A1: BayesX code for Bayesian Semi Parametric regression model

#### BayesX - Software for Bayesian Inference in Structured Additive Regression Models

Version 2.00 (8.05.2009)

```
> dataset d
> d.infile using E:\tes transfer\Personal\thes\Mythesis\data\New
data\Nv2data.dat
NOTE: 14 variables with 3140 observations read from file
E:\tes transfer\Personal\thes\Mythesis\data\New data\Nv2data.dat

> bayesreg n
> n.outfile = c:\data\nonp\n
> logopen, replace using c:\data\nonp\logmcmc.txt
> n.regress zscore = Res + Medu2 + Medu3 + BORD + Swat + Mwork +
Pbint2 + Pbint3 + HHecon + Pedu2 + Pedu3 + BMI(psplinerw2)+
Cage(psplinerw2) + Mage(psplinerw2),family=gaussian
iterations=12000 burnin=2000 step=10 predict using d
```

#### Where:

Options iterations, burnin and step define properties of the MCMC algorithm. The total number of MCMC iterations is given by iterations while the number of burn in iterations is given by burnin. Therefore we obtain a sample of 10000 random numbers with the above specifications.

Since, in general, these random numbers are correlated; we do not use all of them but thin out the Markov chain by the thinning parameter step. Specifying step=10 as above forces BayesX to store only every 10th sampled parameter which leads to a random sample of length 1000 for every parameter in our example. Note, that the choice of iterations of course also affects computation time.

**BAYESREG OBJECT n: regression procedure**

**GENERAL OPTIONS:**

Number of iterations: 12000  
Burn-in period: 2000  
Thinning parameter: 10

**RESPONSE DISTRIBUTION:**

Family: Gaussian  
Number of observations: 3140  
Number of observations with positive weights: 3140  
Response function: identity  
Hyperparameter a: 0.001  
Hyperparameter b: 0.001

**OPTIONS FOR ESTIMATION:**

**OPTIONS FOR FIXED EFFECTS:**

Priors:

diffuse priors

**OPTIONS FOR P-SPLINE TERM: f\_BMI\_pspline**

Prior: second order random walk  
Number of knots: 20  
Knot choice: equidistant  
Degree of Splines: 3

Hyperprior a for variance parameter: 0.001  
Hyperprior b for variance parameter: 0.001

**OPTIONS FOR P-SPLINE TERM: f\_Cage\_pspline**

Prior: second order random walk  
Number of knots: 20  
Knot choice: equidistant  
Degree of Splines: 3

Hyperprior a for variance parameter: 0.001  
Hyperprior b for variance parameter: 0.001

**OPTIONS FOR P-SPLINE TERM: f\_Mage\_pspline**

Prior: second order random walk  
Number of knots: 20  
Knot choice: equidistant  
Degree of Splines: 3

Hyperprior a for variance parameter: 0.001  
Hyperprior b for variance parameter: 0.001

**Annex: A2 MCMC SIMULATION**

**Computing starting values (may take some time)**

ITERATION: 1

APPROXIMATE RUN TIME: 34 seconds

ITERATION: 1000  
ITERATION: 2000  
ITERATION: 3000  
ITERATION: 4000  
ITERATION: 5000

FixedEffects1

Acceptance rate: 99.98 %

Relative Changes in

Mean: 0.20565  
Variance: 1.79769e+308  
Minimum: 0.280946  
Maximum: 0.468366

f\_BMI\_pspline

Acceptance rate: 100 %

Relative Changes in

Mean: 1.0333  
Variance: 1.79769e+308

Minimum: 2.7546  
Maximum: 1.66043

f\_BMI\_pspline\_variance

Acceptance rate: 100 %

Relative Changes in

Mean: 1.07874  
Variance: 1.79769e+308  
Minimum: 0.85705  
Maximum: 47.3243

f\_Cage\_pspline

Acceptance rate: 100 %

Relative Changes in

Mean: 0.298957  
Variance: 1.79769e+308  
Minimum: 0.737175  
Maximum: 1.0803

f\_Cage\_pspline\_variance

Acceptance rate: 100 %

Relative Changes in

Mean: 3.80436  
Variance: 1.79769e+308  
Minimum: 0.834184  
Maximum: 29.0114

f\_Mage\_pspline

Acceptance rate: 100 %

Relative Changes in

Mean: 0.407394  
Variance: 1.79769e+308

Minimum: 2.30374  
Maximum: 2.04477

f\_Mage\_pspline\_variance

Acceptance rate: 100 %

Relative Changes in

Mean: 0.0957355  
Variance: 1.79769e+308  
Minimum: 0.921704  
Maximum: 7.84287

ITERATION: 9000  
ITERATION: 10000  
ITERATION: 11000

FixedEffects1

Acceptance rate: 99.99 %

Relative Changes in

Mean: 0.00312049  
Variance: 0.0181611  
Minimum: 0.030459  
Maximum: 0.00968562

f\_BMI\_pspline

Acceptance rate: 100 %

Relative Changes in

Mean: 0.0249366  
Variance: 0.0747988  
Minimum: 0.0575758  
Maximum: 0.0699277

f\_BMI\_pspline\_variance

Acceptance rate: 100 %

Relative Changes in

Mean: 0.00530695  
Variance: 0.126401  
Minimum: 0  
Maximum: 0

f\_Cage\_pspline

Acceptance rate: 100 %

Relative Changes in

Mean: 0.00736425  
Variance: 0.0251729  
Minimum: 0.0882887  
Maximum: 0.0304518

f\_Cage\_pspline\_variance

Acceptance rate: 100 %

Relative Changes in

Mean: 0.0155263  
Variance: 0.0630309  
Minimum: 0  
Maximum: 0

f\_Mage\_pspline

Acceptance rate: 100 %

Relative Changes in

Mean: 0.011551  
Variance: 0.03803  
Minimum: 0.0304331  
Maximum: 0.113417

f\_Mage\_pspline\_variance

Acceptance rate: 100 %

### **Annex A3: Estimation results for the deviance:**

Unstandardized Deviance  $(-2 * \text{Loglikelihood}(y|\mu))$

Mean:	12151.201
Std. Dev:	9.8362621
2.5% Quantile:	12132.289
10% Quantile:	12138.63
50% Quantile:	12151.13
90% Quantile:	12163.997
97.5% Quantile:	12170.428

Saturated Deviance  $(-2 * \text{Loglikelihood}(y|\mu) + 2 * \text{Loglikelihood}(y|\mu=y))$

Mean:	3140.2291
Std. Dev:	79.02322
2.5% Quantile:	2982.9989
10% Quantile:	3036.5701
50% Quantile:	3139.9777
90% Quantile:	3245.8352
97.5% Quantile:	3293.6861

Samples of the deviance are stored in file  
c:\data\nonp\n\_deviance\_sample.raw

#### **Estimation results for the DIC:**

Estimation results for the DIC:

DIC based on the unstandardized deviance

Deviance( $\bar{\mu}$ ):	12123.35
pD:	31.176825
DIC:	12185.703

#### **Estimation results for the scale parameter:**

Acceptance rate: 100 %

Mean:	2.80719
Std. dev.:	0.0714615
2.5% Quantile:	2.66721

10% Quantile: 2.71556  
50% Quantile: 2.804  
90% Quantile: 2.89917  
97.5% Quantile: 2.95465

### **f\_BMI\_pspline**

Acceptance rate: 100 %

Results are stored in file  
c:\data\nonp\n\_f\_BMI\_pspline.res

Postscript file is stored in file  
c:\data\nonp\n\_f\_BMI\_pspline.ps

Results may be visualized using method 'plotnonp'  
Type for example: objectname.plotnonp 1

### **f\_BMI\_pspline\_variance**

Acceptance rate: 100 %

Estimation results for the variance component:

Mean: 0.00918458  
Std. dev.: 0.0146656  
2.5% Quantile: 0.00111574  
10% Quantile: 0.00172409  
50% Quantile: 0.00496679  
90% Quantile: 0.0182402  
97.5% Quantile: 0.0435727

## Annex A4: Posterior mean for fixed effects

### FixedEffects1

Acceptance rate: 100 %

Variable name	Pmean	Pstd	pqu2p5	10% quantile	Pmedian	90 % quantile	pqu97p5	pcat95	pcat80
const	-1.92	0.17	-2.27	-2.14	-1.93	-1.72	-1.61	-1	-1
Type of residence(Res)	0.61	0.12	0.38	0.47	0.61	0.75	0.83	1	1
Mother education Medu(ref)	-	-	-	-	-	-	-	-	-
Mother education (Medu1 :ref)	0.05	0.09	-0.12	-0.07	0.05	0.17	0.23	0	0
Mother education (Medu2)	0.07	0.17	-0.25	-0.15	0.07	0.29	0.42	0	0
Source of drinking water (Swat)	0.25	0.07	0.12	0.17	0.25	0.34	0.38	1	1
Mother work (Mwork )	-0.11	0.07	-0.26	-0.2	-0.11	-0.02	0.02	0	-1
Previous birth interval (Pbint1 :ref)	-	-	-	-	-	-	-	-	-
Previous birth interval (Pbint2 )	0.08	0.08	-0.08	-0.02	0.08	0.19	0.24	0	0
Previous birth interval(Pbint3)	0.22	0.08	0.06	0.12	0.22	0.32	0.37	1	1
Birth order (BORD)	-0.11	0.08	-0.27	-0.22	-0.11	0.00	0.06	0	1
Household economy HHeco	-0.08	0.07	-0.22	-0.16	-0.08	0.01	0.06	0	0
Partner education(Pedu1: ref)	-	-	-	-	-	-	-	-	-
Partner education(Pedu2)	0.07	0.08	-0.08	-0.03	0.07	0.17	0.22	0	0
Partner education(Pedu3)	0.44	0.12	0.21	0.28	0.44	0.59	0.68	1	1

## Annex A5: Estimation results for the smoothing parameter:

Mean:	757.28	(df: 3.01733)
Std. dev.:	682.035	
2.5% Quantile:	64.2094	(df: 5.53546)
10% Quantile:	153.128	(df: 4.52192)
50% Quantile:	557.526	(df: 3.27016)
90% Quantile:	1609.74	(df: 2.46284)
97.5% Quantile:	2469.08	(df: 2.18977)

Results for the smoothing parameter are also stored in file  
c:\data\nonp\n\_f\_BMI\_pspline\_lambda.res

### **f\_Cage\_pspline**

Acceptance rate: 100 %

Results are stored in file  
c:\data\nonp\n\_f\_Cage\_pspline.res

Postscript file is stored in file  
c:\data\nonp\n\_f\_Cage\_pspline.ps

Results may be visualized using method 'plotnonp'  
Type for example: objectname.plotnonp 3

### **f\_Cage\_pspline\_variance**

Acceptance rate: 100 %

Estimation results for the variance component:

Mean:	0.182253
Std. dev.:	0.172466
2.5% Quantile:	0.0185001
10% Quantile:	0.0343187
50% Quantile:	0.130634
90% Quantile:	0.373069
97.5% Quantile:	0.648506

Results for the variance component are also stored in file

c:\data\nonp\n\_f\_Cage\_pspline\_var.res

Estimation results for the smoothing parameter:

Mean:	35.8223	(df: 8.50715)
Std. dev.:	41.7972	
2.5% Quantile:	4.35454	(df: 12.6226)
10% Quantile:	7.37326	(df: 11.5245)
50% Quantile:	21.0683	(df: 9.46252)
90% Quantile:	81.1845	(df: 7.16552)
97.5% Quantile:	155.663	(df: 6.21558)

Results for the smoothing parameter are also stored in file  
c:\data\nonp\n\_f\_Cage\_pspline\_lambda.res

### **f\_Mage\_pspline**

Acceptance rate: 100 %

Results are stored in file  
c:\data\nonp\n\_f\_Mage\_pspline.res

Postscript file is stored in file  
c:\data\nonp\n\_f\_Mage\_pspline.ps

Results may be visualized using method 'plotnonp'  
Type for example: objectname.plotnonp 5

### **f\_Mage\_pspline\_variance**

Acceptance rate: 100 %

Estimation results for the variance component:

Mean:	0.00588158
Std. dev.:	0.00765519
2.5% Quantile:	0.000918636
10% Quantile:	0.00142878
50% Quantile:	0.00359666
90% Quantile:	0.0121607
97.5% Quantile:	0.0248737

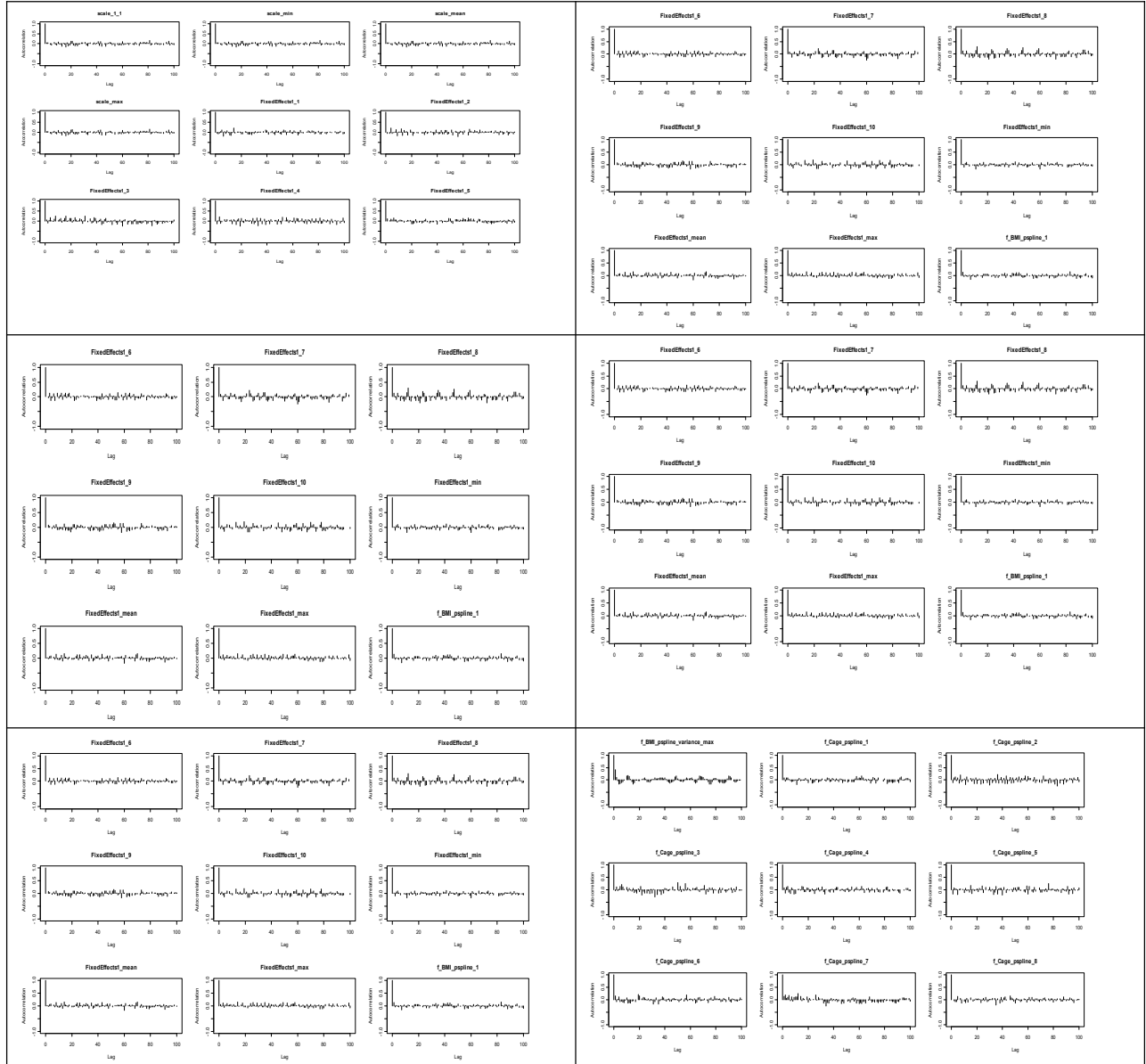
Results for the variance component are also stored in file

c:\data\nonp\n\_f\_Mage\_pspline\_var.res

Estimation results for the smoothing parameter:

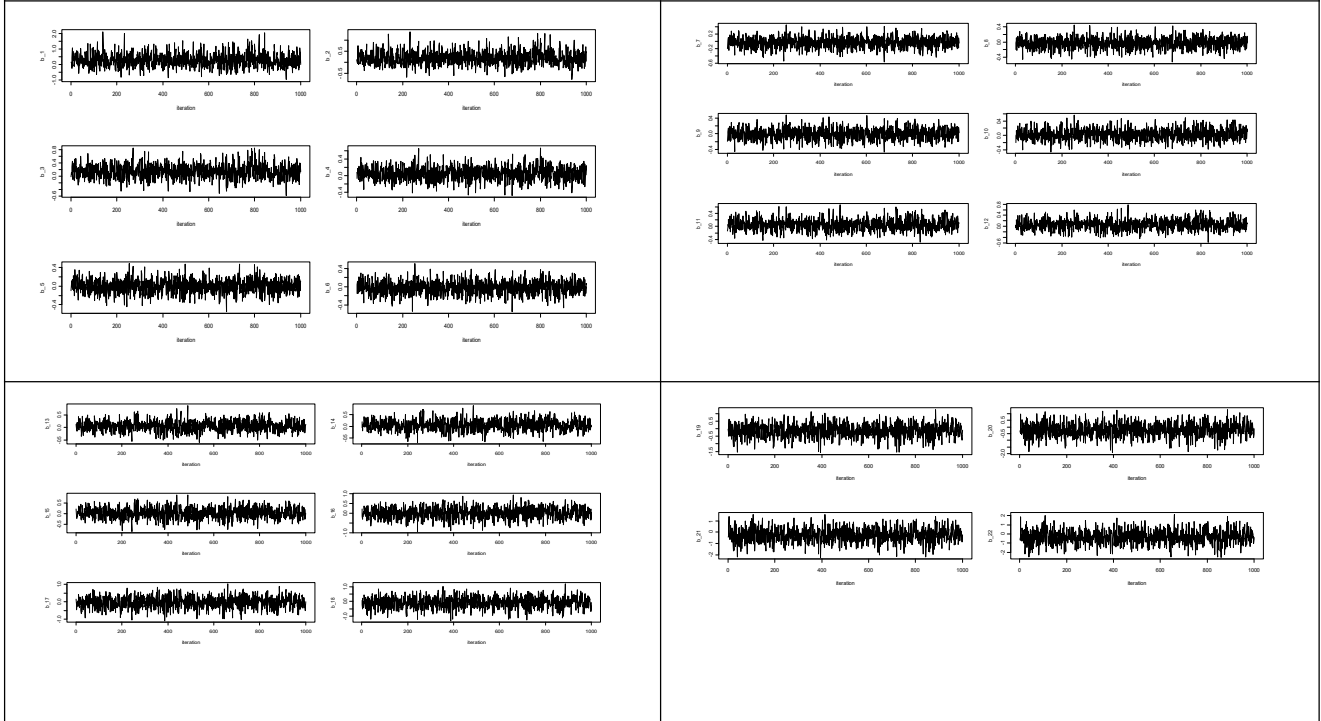
Mean:	977.666	(df: 3.61106)
Std. dev.:	774.576	
2.5% Quantile:	112.82	(df: 6.02408)
10% Quantile:	232.996	(df: 5.09887)
50% Quantile:	777.682	(df: 3.82036)
90% Quantile:	1967.45	(df: 3.03133)
97.5% Quantile:	3082.09	(df: 2.70332)

# Annex B1: Plots of autocorrelation functions for samples obtained with MCMC in BayesX

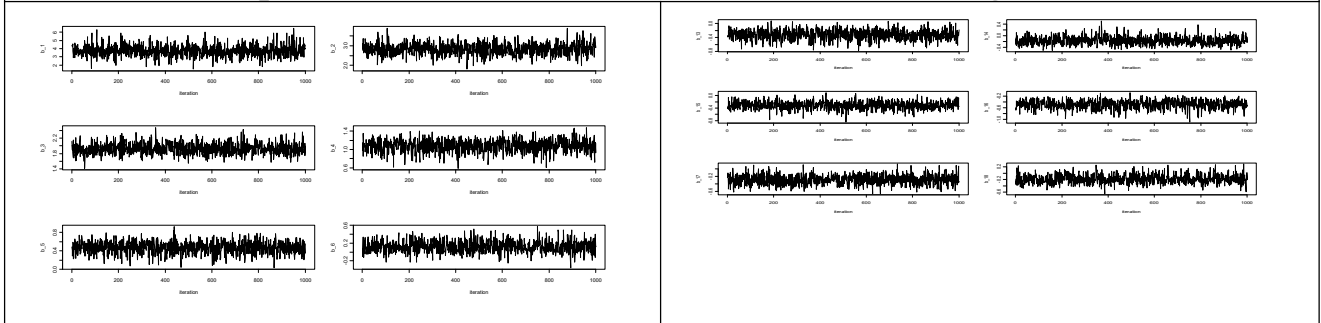


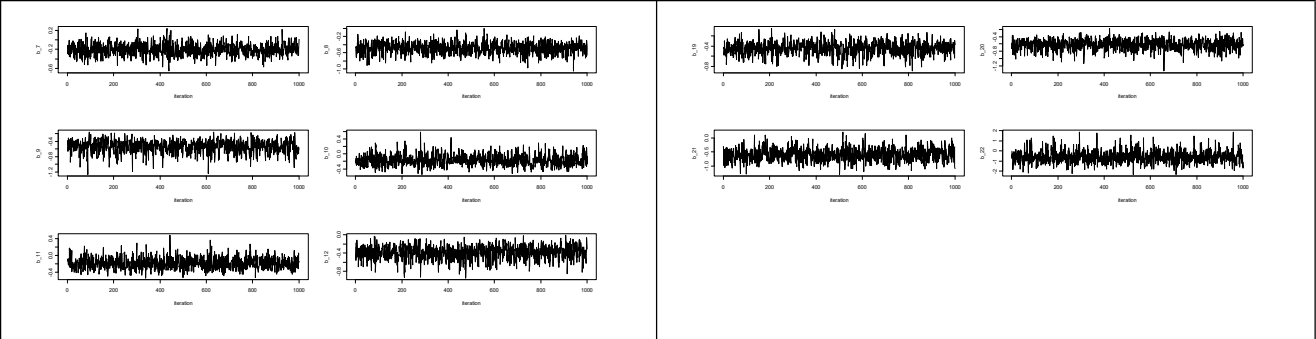
# Sample Path in the MCMC simulation

## Annex B2: Sample Path in the MCMC simulation for Mother's Body Mass Index



## Annex B3: Sample Path in the MCMC simulation for Child Age





### Annex B4: Sample Path in the MCMC simulation for Child Age

