MEASURING POVERTY AND CHILD MALNUTRITION WITH THEIR DETERMINANTS FROM HOUSEHOLD SURVEY DATA

by

FAUSTIN HABYARIMANA

UNIVERSITY OF KWAZULU-NATAL

Pietermaritzburg

2016

MEASURING POVERTY AND CHILD MAL-NUTRITION WITH THEIR DETERMINANTS FROM HOUSEHOLD SURVEY DATA

by

FAUSTIN HABYARIMANA

Submitted in fulfilment of the academic requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

in the

School of Mathematics, Statistics and Computer Sciences

UNIVERSITY OF KWAZULU-NATAL

Pietermaritzburg

2016

Dedication

To God Almighty for his mercy endures forever
To my wife Mrs Odette Umugiraneza
and our lovely sons: Blaise Nere Byiringiro, Bruce Neri Singizwa Byishimo and Happy Blessing Ishimwe.

Declaration

The research work described in this thesis was carried out in the School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Pietermaritzburg campus, under supervision of Professor Temesgen Zewotir and Dr Shaun Ramroop. The work represents an original work by the author. It has not been submitted in any form for any degree or diploma to any other university. Where use has been made of the work of others it is duly acknowledged.

Author: Faustin Habyarimana	Date
Supervisor: Prof. Temesgen Zewotir	Date
Co-Supervisor: Dr. Shaun Ramroop	Date

Published Papers

The following papers have been published from this thesis

1. F. Habyarimana, T. Zewotir and S. Ramroop (2014). A proportional odds model with complex sampling design to identify key determinants of malnutrition of children under five years in Rwanda, *Mediterranean Journal of Social Sciences, 5(23), 1642-1648.*

2. F. Habyarimana, T. Zewotir and S. Ramroop (2015). Analysis of demographic and health survey to measure poverty of household in Rwanda, *African Population Studies, 29(1), 1472-1482*.

3. F. Habyarimana, T. Zewotir and S. Ramroop (2015). Determinants of poverty of households in Rwanda: An application of quantile regression, *Journal of Human Ecology, 50(1): 19-30.*

4. F. Habyarimana, T. Zewotir and S. Ramroop (2015). Determinants of households: Semiparametric analysis of demographic and health survey data from Rwanda,*Journal of Economics and Behavioral Studies,7(3),47-55.*

5. F. Habyarimana,T. Zewotir and S. Ramroop (2015). Key determinants of malnutrition of children under five years in Rwanda: simultaneous measurements of three anthropometric indices, *under review.*

6. F. Habyarimana, T. Zewotir and S. Ramroop (2015). Spatial distribution of key determinants of malnutrition of children under five years in Rwanda: Simultaneous measurement of three anthropometric indices, *under review.*

7. F. Habyarimana, T. Zewotir and S. Ramroop (2015). Joint model of poverty and malnutrition of children under five years: case of Rwanda,*under review.*

Acknowledgments

I would like to express my sincerest gratitude to my supervisor, Prof. Temesgen Zewotir, and my co-supervisor Dr. Shaun Ramroop, for their guidance, encouragement, patience and their time. Every consultation with each of them was an illuminating experience. They helped me to understand statistics and its applications to other disciplines. They were always available to assist me in whatever problem I experienced in this research.

I am also grateful for the facilities made available to me by the School of Mathematics, Statistics and Computer Sciences, UKZN. I would like to thank the staff members of the School of Mathematics, Statistics and Computer Sciences, Pietermaritzburg for their conducive environment and encouragement. I also gratefully acknowledge financial support from the government of Rwanda through the Rwanda Education Board. I would like also to extend my acknowledgement to the University of Rwanda for granting me study leave for this period of research.

I am also thankful to a number of people who directly or indirectly contributed to this work. In particular, thanks are due to Dr Dawit Ayele, P. Phepa, O. Otegbeye, Abdala, Joseph, Koffi, Eliphaz ,Christel Barnard, Isaac Ntahobakurira, Bev Bonhomme, A. Munyengabe and all Rwandese community here in KZN.

My father, A. Ndababonye, my mother, J. Bantegeye, all my brothers and sisters have been always pillars of strength; this is an opportunity to thank them for their sacrifices in ensuring my earlier studies.

Last but not least, I am indebted to my wife Odette Umugiraneza and our sons Blaise Nere Byiringiro, Bruce Neri Singizwa Byishimo and Happy Blessing Ishimwe for their incessant loving support, attention and encouragement.

Contents

List of Figures

Abstract

The eradication of poverty and malnutrition is the main objective of most societies and policy makers. But in most cases, developing a perfect or accurate poverty and malnutrition assessment tool to target the poor households and malnourished people is a challenge for applied policy research. The poverty of households and malnutrition of children under five years have been measured based to money metric and this approach has a number of problems especially in developing countries. Hence, in this study we developed an asset index from Demographic and Health Survey data as an alternative method to measure poverty of households and malnutrition and thereby examine different statistical methods that are suitable to identify the associated factors. Therefore, principal component analysis was used to create an asset index for each household which in turn served as response variable in case of poverty and explanatory (known as wealth quintile) variable in the case of malnutrition. In order to account for the complexity of sampling design and the ordering of outcome variable, a generalized linear mixed model approach was used to extend ordinal survey logistic regression to include random effects and therefore to account for the variability between the primary sampling units or villages. Further, a joint model was used to simultaneously measure the malnutrition on three anthropometric indicators and to examine the possible correlation between underweight, stunting and wasting. To account for spatial variability between the villages, we used spatial multivariate joint model under generalized linear mixed model. A quantile regression model was used in order to consider a complete picture of the relationship between the outcome variable (poverty index and weight-for-age index) and predictor variables to the desired quantiles. We have also used generalized additive mixed model (semiparametric) in order to relax the assumption of normality and linearity inherent in linear regression models, where categorical covariates were modeled by parametric model, continuous covariates and interaction between the continuous and categorical variables by non-parametric models. A composite index from three anthropometric indices was created and used to identify the association of poverty and malnutrition as well as the factors associated with them.

Each of these models has inherent strengths and weaknesses. Then, the choice of one depends on what a research is trying to accomplish and the type of data being used. The findings from this study revealed that the level of education of household head, gender of household head, age of household head, size of the household, place of residence and the province are the key determinants of poverty of households in Rwanda. It also revealed that the determinants of malnutrition of children under five years in Rwanda are: child age, birth order of the child, gender of the child, birth weight of the child, fever, multiple birth, mother's level of education, mother's age at the birth, anemia, marital status of the mother, body mass index of the mother, mother's knowledge on nutrition, wealth index of the family, source of drinking water and province. Further, this study revealed a positive association between poverty of household and malnutrition of children under five years.

CHAPTER 1

Introduction

The eradication of poverty and malnutrition is the first target of the Millennium development goals (eradicate extreme hunger and poverty). But developing a gold standard for poverty and malnutrition measurement is a challenge for applied policy research. This measure is very useful not only in estimating poverty and inequality within the society but also can be used as a control variable in assessing the effect of other variables associated with wealth (Filmer and Pritchett, 2001)

Most measurements and analyses of poverty have been done based on income in developed countries, but on consumption or expenditure in developing countries (Sahn and Stifel, 2003). However, collecting data on income and expenditure in developing countries can be both time and money consuming (Vyas and Kumaranayake, 2006). In addition, in low-income countries, measurement of consumption and expenditure is fraught with difficulties such as the problem of recall and reluctance to divulge information. Additionally, prices are likely to differ substantially across time spans and areas, necessitating complex adjustment of the expenditure figures to reflect these price differences. Sahn and Stifel (2003) studied the theoretical framework underpinning household income or expenditure as a tool for classifying socio-economic status in developing countries. Their theoretical framework underscored five problems. Firstly, the quality of income and expenditure data is most likely to be poor. Secondly, these data are collected on the basis of recall memory. The recall data are prone to measurement errors. Thirdly, prices of goods, nominal interest rates and depreciation rates for semi-durable or durable goods are difficult to discern when constructing consumption aggregates. Fourthly, consumer price indices in developing countries are unavailable and unreliable, especially when inflation tends to be high or variable. In addition, regional and seasonal price indices in most developing countries are greatly variable and rare to find. Problems of sampling bias, under-reporting of income and difficulties of converting household products into money terms are also raised. For this reason, measuring poverty of household based on asset index approach is essential to determine socio-economic status as an alternative tool for classifying the households in their socio-economic status. A measure of the socio-economic status of households is an important element in most economic and demographic analyzes. All indices used to measure poverty have some strength and weaknesses, where some are used to only measure absolute poverty (refers to a set standard which is the same in all countries and which does change over time) which requires money-metric, for instance Gini index (it measures the degree of inequality in the distribution of family income in developed countries or consumption in developing countries in a country), and others used to measure relative poverty (refers to a standard which is defined in terms of society in which individuals lives and which therefore differs between countries and overtime) which does not necessarily require money-metric for instance(asset index) (Palmer, 2010). The detailed strength and some weaknesses of asset index compared to other indices are thoroughly discussed in Falkingham and Namazie (2002) and Sahn and Stifel (2003). Likewise malnutrition is a very serious problem for public health in developing countries. Children are more prone to suffer from malnutrition deficiencies than adults because they are in a physiologically less stable situation. Child malnutrition is a clinical sign of nutrient deficiency manifested as stunting, underweight and wasting. These manifestations are often measured using biomedical or anthropometric indicators. However, anthropometric indicators are mostly used for its affordability and relation availability. Commonly used anthropometric indicators of child malnutrition under the age of five years (WHO, 1995) are: height-for-age, known as stunting which is an indicator of child's long-term or chronic nutritional status and is also affected by the current or chronic illness. Wasting is weight-for-height index which measures body mass in relation to body height and describes current nutritional status of the child. Wasting represent the failure to receive adequate nutrition in the period immediately preceding the survey and may be the result of inadequate food intake or a recent episode of illness causing loss of weight and the onset of malnutrition. The third one is underweight which is a composite index given by weight-for-height and height-for-age. Depending on the purpose of the assessment and the nature of intervention, the above three indices can either be used separately or together. When anthropometric measurements are taken regularly over time, they could provide information on how the health status of the population is changing and provide a timely warning on the food supply of a given area. If the purpose is to obtain a quick picture of a community or large body of population to understand the extent of the problem, the measurement of wasting alone would provide sufficient information. However, if the purpose is to obtain information to decide what type of programs are needed in the specific area, the study involves all three indices of anthropometric measurements.

The demographic and health survey (DHS) data is the most used survey in many developing countries, and is generally done each five years. The data are collected using multistage sampling (including stratification, clustering and unequal probability of selection). Therefore fitting the DHS data without considering the survey sampling design may lead to biased estimates of parameters and incorrect variance estimates (Anthony, 2002; Liu and Koirala, 2013). Das and Rahman (2011) determined the risk factors of malnutrition using proportional odds model but they never included the complexity of sampling design. On our knowledge there is no researcher in literature used multivariate joint model to account for possible correlation of anthropometric indices. Kandala et al. (2011a) used geo-additive semi-parametric mixed model to find out whether the geographic location can affect malnutrition, however they only considered one anthropometric index and did not account for possible correlation between different anthropometric indices. All these studies considered binary or ordinal response variable but they never considered the whole distribution of the response variable and this can help to reveal the information which can be hidden by binary variable or ordinal variables (Koenker and Basett, 1978).

The main objective of this study was develop an alternative method for measuring poverty of household and malnutrition of children under five years together with their determinants from demographic and health survey.

The findings from this study will help the researchers and scholars to model the demographic and health survey data and therefore perfectly assess the determinants of poverty and malnutrition of children under five years based on demographic and health survey data (DHS). We seek to test different statistical methods which in turn can help to propose suitable techniques and to appropriately fit future work from the demographic and health survey data and any other related data. Therefore the specific objectives are:

- To computer a reliable asset index and composite index of three anthropometric indicators using principal component analysis
- To account for complexity from sampling design and to make valid statistical inference using survey logistic regression
- To deal with ordered categorical data by extending classical proportional odds model to include sampling design
- To deal with symmetric distribution of the data by fitting the quintile regression at different parts of the distribution of the response variable
- To account for variability between primary sampling units by fitting the generalized linear mixed model (GLMM)to the data
- To account for correlation between the anthropometric indices using Multivariate joint model under GLMM
- To develop a model that account for joint effect and spatial autocorrelation
- To deal with nonlinear effects of continuous covariates by fitting simiparametric generalized additive mixed model to the data

The thesis is structured as follows: in Chapter 2, we discuss poverty index and classification of households in socio-economic status.

Chapter 3 presents a review of generalized linear models (GLM) which accounts for the complexity of the survey. Chapter 4 presents a review of generalized linear mixed model as an alternative to GLM to handle survey data analysis. Chapter 5 presents a comprehensive review of joint modelling of three anthropometric indicators known as stunting, wasting and stunting. In chapter 6, we extend chapter 5 to include spatial variability and also produce the smooth maps of prevalence of malnutrition. Chapter 7 presents quantile regression model and its application to poverty of households as well as malnutrition of children under five years in Rwanda. In chapter 8 we review generalized additive mixed model and apply it to poverty of households as well as malnutrition of children under five years in Rwanda. In chapter 9 we discuss the composite index and joint modelling of poverty and malnutrition. Chapter 10 presents the discussion and conclusion as well as the possible future researches.

CHAPTER 2

Poverty index and classification of households

As previously discussed in Chapter 1, the measurement of poverty and malnutrition based on income and expenditure or consumption has a number of problems. In order to solve these problems, in this chapter we create an asset index that can be used as an alternative approach to measure poverty of households. This index can also be used in malnutrition as wealth quintile.

2.1. Data

There are many surveys used to collect data. Examples of such surveys are: the Demographic Health Survey (DHS) which is done every five years, the Census of the population which is generally done every 10 years (but is too expensive compared to the demographic and health survey) and Household Budget Surveys done in general every five years. The DHS is available in many countries. DHS has earned worldwide repute for collecting and disseminating accurate, nationally representative data on households'characteristics, fertility, family planning, early childhood mortality, maternal and child health, maternal and child nutrition, malaria and HIV/AIDS and it usually includes Global positioning system (GPS) coordinates. The data used in this study is from Rwanda Demographic Health Survey (2010). The sampling in this survey was done in two stages. In the first stage 492 villages, known also as clusters or primary sampling units or enumeration areas, were considered with probability proportional to the village size (the number of households residing in the village). Then, a complete mapping and listing of all households existing in the selected villages was conducted. The resulting lists of households served as the sampling frame for the second stage of sample selection. Households were systematically selected from those lists for participation in the survey. A total of 12,792 households were selected, of which 12,570 households were identified and occupied at the time of the survey. Among these households 12,540 completed the household questionnaire of which 2009 and 10531 households were urban and rural respectively; yielding a response rate of 99.76 %. In the 12,540 households surveyed, 13,790 women were eligible for the individual interview and 13,671 of them completed interviews; yielding a response rate of 99.13 %. A total of 6,414 men aged 15-49 were identified in subsample of households and 6,329 of them completed individual interviews, yielding a response rate of 98.67 %. interviews were completed by 13,671 The survey had various types of questionnaires developed for households, for men and for women. The man's questionnaire did not contain questions on maternal and nutrition and these questions were contained in women's questionnaire (NISR et al., 2012). Therefore, in this study, only the households and women questionnaires are considered. The missingness in this data is negligible (0.24 % for household questionnaire and 0.87 % for women questionnaire).

2.1.1. Baseline characteristics of the study.

Independent variable poverty case: The predictor variables considered in this study are from household head characteristics such as: education level, gender and age of household head, household characteristics such as the size (number of family member) of household and environmental characteristics such as place of residence and province or region. The levels and coding of the categorical variables are given in Table 2.1.

TABLE 2.1. Table of predictor variables used in poverty

Variable	**Level and coding**
Province/Region	1=Kigali, 2=South, 3=West, 4=North, 5=East
place of residence	1=urban, 2=rural
Gender of the household head	1= male, 2=female
Education level of household head	1= Higher, 2= secondary, 3= primary, 4= no education
Size	continuous
Age of household head	continuous

The characteristics of households heads, the number of household members and their proportions in percentages are presented in Table 2.2. We observe from this table that 66.8 % of households in Rwanda were headed by males while 33.2 % were headed by female. The biggest proportion of households heads (58.2 %) attained primary

TABLE 2.2. Characteristics of Head of Households

Characteristic	**Category**	Number of household heads	Percent
Gender of the household head	Male	8382	66.8
	Female	4158	33.2
Education level of household head	No education	3668	29.3
	Primary	7377	58.2
	Secondary	1182	9.5
	Tertiary	279	2.2

education only, followed by 29 % who did not have any former education, 9.5 % households heads attained secondary education and only 2.2 % managed to attain education beyond secondary school. The minimum age of household head was 13, the mean age of household head was 43.7 and the maximum age of household head was 98 years old Table 2.3. We observe from the same table that the minimum number of household member was 1, mean of household members was 4 and the maximum number of household members was 20.

TABLE 2.3. Continuous variables for household head and household

Variable	**Minimum**	Mean	Maximum
Age household head	13	43.7	98
Size of household	1	4.5	20

TABLE 2.4. Environmental Characteristic of the Household

Characteristic	**Category**	Number of household	Percentage
Place of residence	Rural	10531	84.0
	Urban	2009	16
Province	Kigali	1522	12.1
	South	3262	26.0
	West	2840	22.6
	North	2047	16.3
	East	2869	22.9

The biggest proportion of households (84.0 %) were from rural while 16 % were urban Table 2.4. In the same table we observe that the big proportion of households was from Southern province (26.0 %), followed by Eastern province (22.9 %), and a smaller proportion of households was from Kigali city (12.1 %).

Independent variables used in malnutrition: In this research, it was considered the following covariate variables: child age, gender of the child, birth order, multiple birth, birth weight, mother's level of education, mother's knowledge on nutrition, body mass index of the mother, incident of anemia, mother's age at the birth, assistance of the mother at the delivery, place of residence (urban or rural), province or region, source of drinking water, toilet facilities, wealth index, sickness such as cold, cough, diarrhea, whether the child had fever or not in two weeks before the survey, child caretaker, breastfeeding, feeding index. It is reported in Table 2.5, any covariate that is at least significantly associated to one of the three response variables (stunting, wasting and underweight) and then finally used in the analysis. The level and coding of the categorical variables are given in Table 2.5

Table 2.5. Table of predictor variables used in malnutrition

Variable	**Level and coding**
Child age in moths	1=0 -11 months, 2= 12-23 months, 3=24-59 months
Birth order	1=first order, 2=2-3rd order, 3=4-5th order, 4=6th order and more
Mother's age at the birth	1= less than 21 years old, 2= 21 years old and more
Mother's education	1= secondary and higher, 2= primary, 3= no education
Gender of child	1= male,2=female
Wealth index	1=rich, 2= middle, 3=poor
Birth weights	1= weight greater or equal to 2500g, 2= weight less than 2500g
Province/region	1=Kigali, 2=South, 3=West, 4=North, 5=East
Knowledge on nutrition	1= has knowledge on nutrition, 2= no knowledge on nutrition
Multiple birth	1=singleton, 2= first multiple, 3=second multiple and more
Anemia	1= no anemic mother, 2= anemic mother
Marital status of the mother	1=married, 2= never in union, 3=separeted, 4= widow
Body Mass Index	1= greater or equal to 18.5, 2= less than 18.5
Had fever in last two weeks	1= no fever, 2= had fever
Source of drinking water	1= piped in dwelling, 2=public tape,3=protected spring, 4= other sources

Children belonging to aged group 24-59 months were 61.8 % of all other children considered in the survey, the children belonging to infant were 19.0 % and 19.2 % were children belonging to the age group 12-23 months Table 2.6. Most of children in Rwanda were born with higher birth weight 95.8 % while the children born with lower birth weight were 4.2 %. The children who did not have fever in the last two weeks prior to the survey were 84.2 % whereas 15.8 % of children considered in the survey had fever. The proportion of male children was 50.9 % while 49.1 % of children were

TABLE 2.6. Child's Characteristic

Characteristic	**Category**	Number of children	Percentage
Child age in months	0-11	786	19.0
	12-23	794	19.2
	24-59	2553	61.8
Birth order	first	2253	25.0
	Two and third	3078	34.2
	Fourth and fifth	1940	21.6
	sixth and more	1731	19.2
Province	Kigali	991	11.0
	South	2244	24.9
	West	2167	24.1
	North	1374	15.3
	East	2227	24.7
Gender of the child	Male	4586	50.9
	Female	4416	49.1
Birth weights	$\geq 2500g$	8599	95.8
	$< 2500g$	379	4.2
Had fever in last two weeks	Yes	1332	15.8
	No	7085	84.2

female.

Most of women at birth had 21 years old or more (95.5 %) while 4.5 % of women had less than 21 years old at the birth Table 2.7. In the same table, we observe that 71.9 % of mothers had primary education, 18.9 % of mothers had no formal education and 9.4 % had higher school education or more. The proportion of incident of anemia was 4.5 % of all mothers considered in the survey. The proportion of mothers who had knowledge on nutrition was 64.0 % while 36.0 % had no knowledge on nutrition.

Table 2.7. Mother's Characteristic

Characteristic	**Category**	Number of mothers	Percent
Mother's age in years at the birth	< 21	383	4.5
	≥ 21	8101	95.5
Mother's education	No education	1702	18.9
	Primary	6451	71.9
	Secondary and higher	849	9.4
Incident of anemia	No anemic	2305	61.7
	Anemic	1428	38.7
Knowledge on nutrition	No	2497	36.0
	Yes	4443	64.0
Body mass index	≥ 18.5	4326	95.5
	< 18.5	205	4.5

2.2. Principal components analysis and computation of asset index

Introduction

The technique of principal components analysis was first described by Pearson (1901). He apparently believed that this was the correct solution to some of the problems that were of interest to biometricians at the time, although he did not propose a practical method of calculation for more than two or three variables. A description of practical computing methods came much later from Hotelling (1933). Even then, the calculations were extremely daunting for more than a few variables because they had to be done by hand. It was not until the electronic computer became generally available that the principal components technique achieved widespread use. This allowed the applications of PCA in many application, such as dimension reduction.

2.2.1. Mean and covariance matrices.

Population and sample values for single random variables are often summarized by the values of the mean and variances. Therefore, if a sample of size n yields $x_1, x_2, x_3, ..., x_n$ then its sample mean and variance are given respectively by

$$\bar{x} = \sum_{i=1}^{n} x_i/n \tag{2.1}$$

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 /(n-1) \tag{2.2}$$

where $\bar{x}$ and s^2 are the estimates of the corresponding population mean μ and population variances σ^2. In a similar way, multivariate populations and samples can be summarized by mean vector and covariance matrices. Let us consider p variables $x_1, x_2, x_3, ..., x_p$ and that a sample of n values for each of these variables is available, using the equations (2.1) and (2.2), the sample mean $\bar{x}_i$ and variance s_i^2, are given respectively by

$$c_{jk} = \sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right) \left(x_{jk} - \bar{x}_k\right) / (n-1) \tag{2.3}$$

where c_{jk} is the sample covariance between variables x_j and x_k, x_{ij} is the value of the variable x_j for the i^{th} multivariate observation. This covariance is therefore a measure of the extent to which there is a linear relationship between x_j and x_k, where a positive value indicates that the large value of x_j and x_k tend to occur together, whereas a negative value indicates that large values for one variable tend to occur with the small values of the other variable. The equation (2.3) is related to ordinary correlation coefficient between two variables and given by

$$r_{jk} = \frac{c_{jk}}{s_j s_k} \tag{2.4}$$

Moreover, the definitions suggest that $c_{kj} = c_{jk}, r_{kj} = r_{jk}, c_{jj} = s_j^2$ and $r_{jj} = 1$. Furthermore, the sample matrix of variances and covariances, or the covariance matrix, is given

$$C = \begin{bmatrix} c_{11} & c_{12} & . & . & . & C_{1p} \\ c_{21} & c_{22} & . & . & . & C_{2p} \\ . & . & & & . & . \\ . & . & & & . & . \\ . & . & & & . & . \\ c_{p1} & c_{p2} & . & . & . & c_{pp} \end{bmatrix}$$

where $c_{ii} = s_i^2$ and the population covariance matrix is given by

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & . & . & . & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & . & . & . & \sigma_{2p} \\ . & . & & & . & . \\ . & . & & & . & . \\ . & . & & & . & . \\ \sigma_{p1} & \sigma_{p2} & . & . & . & \sigma_{pp} \end{bmatrix}$$

and finally, the sample correlation matrix is given by

$$R = \begin{bmatrix} 1 & r_{12} & . & . & . & r_{1p} \\ r_{21} & 1 & . & . & . & r_{2p} \\ . & . & & & . & . \\ . & . & & & . & . \\ . & . & & & . & . \\ r_{p1} & r_{p2} & . & . & . & 1 \end{bmatrix}$$

2.2.2. Principal component analysis.

Principal Component Analysis (PCA) is a multivariate statistical technique that linearly transforms an original data set of variables into a substantially smaller set of uncorrelated variables that represents most of information in the original set of variables (Jolliffe, 1986; Stevens, 1986; Jobson, 1992; Manly, 2005). The basic idea is to present a set of variables by a smaller number of variables called *principal components*. A small set of uncorrelated variables is much easier to understand and use in further analysis than a larger set of correlated variables (Lewis-Beck, 1994).

Computation of principal components: The principal components can be calculated on either a sample variance-covariance matrix (with raw data) or a correlation matrix (with standardized data) (Jolliffe, 1986; Johnson and Wichern, 2002). The correlation matrix is used when the variables have different units (for instance: the number of fridges owned by a household, annual income, education level), while covariance matrix is used when the units are homogeneous. Let us consider a subset of variables $X_1^*, X_2^*, ..., X_p^*$ taken from n households. In order to avoid one or two variables having an undue influence on principal component, it is better to start by specifying each variable normalized by its mean and its standard deviation at the start of the analysis (Jolliffe, 1986; Stevens, 1986; Jobson, 1992; Manly, 2005) as follows

$$X_1 = \frac{X_1^* - \bar{X}_1^*}{S_1}, X_2 = \frac{X_2^* - \bar{X}_2^*}{S_2}, ..., X_p = \frac{X_p^* - \bar{X}_p^*}{S_p} \tag{2.5}$$

where $\bar{X}_i^*$ is the mean of X_i^* and S_i is its standard deviation. Therefore, the p^{th} principal component can be written as a linear combination of original variables given by

$$PC_p = \gamma_{p1}X_1 + \gamma_{p2}X_2 + ... + \gamma_{pp}X_p \tag{2.6}$$

where γ_{pp} represents the weight for the p^{th} principal component and the p^{th} variable. The principal components are chosen such that the first component

$$PC_1 = \gamma_{11}X_1 + \gamma_{12}X_2 + ... + \gamma_{1p}X_p \tag{2.7}$$

accounts for as much of the variation in the original data as possible subjected to the constraint that

$$\gamma_{11}^2 + \gamma_{12}^2 + ... + \gamma_{1p}^2 = 1 \tag{2.8}$$

The second component is completely uncorrelated with the first component, and explains additional but less variation than the first component, subjected to the same constraint. The subsequent components are uncorrelated with the previous components; therefore, each component captures an additional dimension in the data, while explaining smaller and smaller proportions of the variation of original variables in the data.

The number of principal components: When computing a principal component analysis, we need to determine the actual dimensionality of the space in which the data falls. Several methods have been proposed in literature for determining the number of components to retain. But the most widely used methods are the following:

Kaiser's rule (Kaiser-Guttman rule): The most used criterion in deciding the number of components to retain is that of Kaiser-Guttman rule also called eigenvalue-one criterion or simply Kaiser criterion (Kaiser, 1960). This criterion retains only those components whose eigenvalues are greater than 1.00. This is the default rule used by SPSS and BMDP packages. Although generally using this rule will result in retention of only the most important factors, blind use could lead to retaining factors which may have no practical significance (in terms of % variance accounted for). The cut off point for the number of principal components is based on the magnitude of the variances of the principal components. Any principal component whose variance is less than 1 (eigenvalue) is not selected.

Scree plot test: A graphical method called the scree test has been proposed by Cattell (1966). In this method the magnitude of the eigenvalues (vertical axis) are plotted against their ordinal number (whether it was the first eigenvalue, the second, etc.). Generally what happens is that the magnitude of successive eigenvalues drops off sharply (steep descent) and then tends to level off. The recommendation is to retain all eigenvalues (and hence components) in the sharp descent before the first one on the line where they start to level off. Several studies have investigated the accuracy of scree test. Tucker et al. (1969) found it to yield the correct number of factors in 12 of 18 cases. Linn (1968) found it to yield the correct number of factors in 7 of 10 cases, while Cattel and Jaspers (1967) found it to be correct in 6 of 8 cases. The extensive study on the number of factors problem by Hakstian et al. (1982) adds additional information. They note that for $N > 250$ and a mean communality (the proportion of each variable's variance that can be explained by the principal component) $\geq .60$, either the Kaiser or scree rules will yield an accurate estimate for the number of factors. They add that such an estimate will be that much more credible if Q/P ratio is $< .30$ (P is the number of variables and Q is the number of factors). With mean communality .30 or $Q/P > .30$, the Kaiser rule is less accurate and the scree rule much less accurate (Stevens, 1986).

Alternatively, the graphical method called scree diagram or scree plot is used. The eigenvalues are ordered from largest to smallest and then a scree plot is constructed by plotting the value of each eigenvalue against its number. The appropriate number of components is given by the elbow in the scree plot. Look for the points (components) after which the remaining eigenvalues decreases in the linear fashion and retains only those points above the elbow (Johnson and Wichern, 2002). So what criterion should be used in deciding how many factors to retain? Since Kaiser criterion has been shown to be quite accurate when the number of variables is < 30 and the communalities (amount of the variance in each variable that is accounted for) are$>.70$, or when $N > 250$ and the mean of communality is $\geq .60$, we would use it under these circumstances. For other situations use the scree test when an $N > 200$ will probably not lead us too far astray, provided that most of the communalities are reasonably large (Stevens, 1986).

Proportional of variance accounted for: A third criterion in solving the number of factors problem involves retaining a component if it accounts for a specified proportion (percentage) of variance in the data set; for instance, you may decide to retain any component that accounts for at least 6% or 10% of the total variance. The proportion of variance criterion has a number of positive features. For instance, in most cases, a researcher might not want to retain a group of components that, combined, account for only a minority of the variance in the data set (say, 25%). But this method is also sometimes criticized for its subjectivity (Kim and Mueller, 1978).

Increasing the interpretability by rotation: Although principal components are adequate for summarizing most of the variance in a large set of variables with a small number of components, often the components are not easily interpretable. The components are artificial covariates designed to maximize the variance accounted for, and are not designed for interpretability. To aid in interpreting, there are various so-called rigid rotations that are available. They are rigid in the sense that orthogonality (uncorrelatedness) of the components is maintained for the rotated factors. This can be done by:

Quartmax: Here the idea is to clean up the variables, that is, the rotation is done so that each variable loads mainly on one factor. Then that variable can be considered to be a relatively pure measure of the factor. The problem with this approach is that most of the variables tend to load on a single factor, making interpretation on the factor difficult.

Varimax: Kaiser (1960) designed a rotation to clean up the factors. That is, with his rotation each factor tends to load high on a smaller number of variables and low or very low on the other variables. This will generally make interpretation of the resulting factors easier. The Varimax rotation is the default option in the SPSS and

BMDP packages. It should be noted that when Varimax rotation is done the maximum variance property of the original components is destroyed. The rotation essentially reallocates the loadings. Thus, the first rotated factor will no longer necessarily account for the maximum amount of variance. The amount of variance accounted for by each rotated factor has to be recalculated (Stevens, 1986).

Let $B = (b_{ij})$ be the matrix of rotated factors. Therefore, the goal of Varimax is to maximize the following quantity

$$Q = \sum_{i=1}^{k} \left(\frac{p \sum_{i=1}^{p} b_{ij}^4 - \sum_{i=1}^{p} b_{ij}^2}{p} \right) \tag{2.9}$$

The equation (2.9) gives the raw varimax rotation and this has the disadvantage of not spreading the variance among the new factors. However, this is corrected by using the normalized-varimax rotation and equation (2.9) becomes

$$Q = \sum_{i=1}^{k} \left(\frac{p \sum_{i=1}^{p} \left(\frac{b_{ij}}{h_i}\right)^4 - \sum_{i=1}^{p} \left(\frac{b_{ij}}{h_i}\right)^2}{p^2} \right) \tag{2.10}$$

where h_i is the square root of the communality of the variable i.

Bartlett's sphericity test: The Bartlett's test compares the observed correlation matrix to the identity matrix. In other words, it checks if there is a certain redundancy between the variables that we can summarize with a small number of factors. If the variables are perfectly correlated, only one factor is sufficient. The Bartlett's test statistic indicates to what extent we deviate from the references situation $|R| = 1$. It uses the following formula

$$\chi^2 = \left(n - 1 - \frac{2p + 5}{6} \right) \times ln|R| \tag{2.11}$$

where p is the number of variables, n is the number of observations and $ln|R|$ is the natural logarithm of the determinant of R (correlation matrix if correlation matrix is used). Under H_0, it follows a χ^2 distribution with a $[p \times (p-1)/2]$ degree of freedom. However, the Bartlett's test has a strong drawback. It tends to be always statistically significant when the number of instances n increases. It is however advised to use it when the ratio n/p is lower than 5.

Kaiser-Meyer-Olkin measure of sampling adequacy: The Kaiser-Meyer-Olkin

(KMO) measure of sampling adequacy is an index for comparing the magnitude of the observed correlation coefficients to the magnitudes of the partial correlation coefficients. The overall KMO index is computed as follows:

$$KMO = \frac{\sum_i \sum_{j \neq i} r_{ij}^2}{\sum_i \sum_{j \neq i} r_{ij}^2 + \sum_i \sum_{j \neq i} a_{ij}^2} \tag{2.12}$$

where r_{ij} is the observed correlation coefficient of i^{th} and j^{th} and a_{ij} is the corresponding partial correlation coefficients given by

$$a_{ij} = -\frac{v_{ij}}{\sqrt{v_{ii} \times v_{jj}}} \tag{2.13}$$

where v_{ij} is the inverse of the correlation matrix.

If the partial correlation is near to zero, the PCA can perform the factorization efficiently because the variables are highly related and as results $KMO \approx 1$. KMO index per variable is given by

$$KMO_j = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} a_{ij}^2} \tag{2.14}$$

This index is used to detect those variables which are not related to the others. If the KMO index is high (≈ 1), the PCA can act efficiently; if KMO is low (≈ 0), the PCA is not relevant. The Bartlett's sphericity test and KMO index enable us to detect if we can or cannot summarize the information provided by the initial variables in a few number of factors. However, they do not provide an indication about the appropriate number of factors to retain.

Reliability test of asset index: A reliable index has to be internally coherent; this means that it has to consistently produce a clear separation across poor, middle and rich household for each asset included in the index. It has also to be robust; that means the asset index produces very similar classifications when different subsets of variables are used in its construction (Filmer and Pritchett, 2001).

2.2.3. Application to computation of poverty index.

The main objective of this section is to create an asset index of each household included in 2010 Rwanda demographic and health survey and thereafter classify the households into socio-economic status (poor or not)(Habyarimana et al., 2015a; Vyas and Kumaranayake, 2006; Filmer and Pritchett, 2001, 1998) that measures whether a household is poor or not.

2010 RDHS gathered information on households' ownership of durable goods, school attendance, source of drinking water, sanitation facilities, washing places, housing quality, etc. In this study, when computing the socio-economic index, we have only considered the ownership of durable goods, toilet facilities, quality of house (floor, roof and wall material) and source of drinking water (Filmer and Pritchett, 2001). SPSS 22 was used in the analysis and computation of asset index

2.3. Results from PCA and socio-economic index

Tables 2.8 and 2.9 report the scoring factors of 53 variables and their corresponding percentage in the wealth quintile. Generally, a variable with a positive factor score or weights contributes to higher socio-economic status (SES), and conversely a variable with a negative factor score weighs towards lower SES. Usually, the richest households (20% or fifth quintile) have the assets with higher factor scores. For instance 8.1% of richest households have flush toilet whereas poorest and middle households are 0%; 85.2% of richest households have a cement floor against 0% of poorest households and 1.7% of middle households; 81.0% of richest households have metal roof against 53.2% of middle households and 34.4% of poor households, 53.5% of fifth quintile own electricity against 0.8% of third and fourth quintile and 0% of first and second quintile have a refrigerator; 86.6% of richest households own a mobile phone against 56.6% of middle and 3.3% of poor households; 9.5% of fifth quintile own a personal computer against 0% of poor and middle households. The higher percentage of poor households (40% first and second quintile) would have assets with lower scores (negative), 98.9% of poor households own latrine toilet against 87.3% of richest; 100% of poor households own earth/sand floor against 10.0% of richest households; 7.7% of poor households own a thatch roof against 0.0% of richest households; 82.1% households of poor use wood as cooking fuel whereas 44.6% of richest households use wood for cooking; 97.7% households of poor own land usable for agriculture against 53.3% of richest households.

For instance 8.1% of richest households have flush toilets whereas poorest and middle households are 0%; 85.2% of richest households have a cement floor against 0% of poorest households and 1.7% of middle households; 81.0% of richest households have a metal roof against 53.2% of middle households and 34.4% of poor households; 53.5% of fifth quintile own electricity against 0.8% of third and fourth quintile and 0% of first and second quintile; 86.6% of richest households own a mobile phone against 56.6% of middle and 3.3% of poor households; 9.5% of fifth quintile own a personal computer against 0% of middle and 0% of poor households Table 2.8. The higher percentage of poor households (40% or first and second quintile) would have assets with lower scores. For instance 98.9% of poor households own a latrine toilet against 87.3% of richest households; 100% of poor households own earth/sand floors against 94.3% of middle households and 10.0% of richest households; 7.7% of poor households own a thatch roof against 0.0% of richest households; 82.1% of poor households use wood as cooking fuel whereas 44.6% of richest households use wood for cooking; 97.7 % of poor households own land usable for agriculture against 53.3% of the fifth quintile Table 2.8.

In our analysis we have excluded ethnicity because it is not applicable to Rwanda. We did not include religion because it is not listed in the household data set of Rwanda, even though religion seems to be more individual than household characteristics. However, religion was used in some research such as Achia and Khadioli (2010). Asset indexes derived from DHS data can be subjected to a number of tests (Filmer and Pritchett, 1998). For instance a good index has to be internally coherent, which means that it has to consistently produce a clear separation across the poor, the middle and rich household for each asset included in the index. This means that each of the variables included in the index can be compared across households that fall into the poorest 40%, middle 40% and richest 20% of the population based on the asset index. The internal coherence is tested in Tables 2.8 and 2.9. From these tables we can see a clear separation of an asset among poorest households, middle and richest households;for instance 85.2% of richest households have a cement floor against 0% of poorest households and 1.7% of middle households Table 2.8. It has also to be robust, that means produce similar classifications of households or individuals across constructions of asset index based on different subsets of variables Booysen (2002). The robustness is tested respectively in Tables 2.8 and 2.9.

TABLE 2.8. Component scores and classification into wealth quintile

Variables	Component score	Poorest40%	Middle 40%	Richest 20%
Toilet facilities				
Flush toilet	.465	0.0	0.0	8.1
Latrine	-.262	98.9	92.3	87.3
Ventilated	.075	0.0	2.9	3.7
Other	-.027	0.6	2.0	0.6
Floor material				
Earth/Sand	-.736	100	94.3	10.0
Dung	-.004	0.0	1.5	0.6
Ceramic tiles	.339	0.0	0.0	2.6
Cement	.710	0.0	1.7	85.2
Other	.005	0.0	2.5	1.6
Roof material				
Thatch/Palm leaf	-.132	7.7	3.7	0.0
Rustic/Plastic	-.038	0.8	0.9	0.1
Metal	.434	34.4	53.2	81.0
Ceramic tiles	-.383	55.7	41.2	17.6
Cement	.072	0.1	0.1	0.6
Other	.001	1.4	0.9	0.7
Wall material				
Dirt	-.084	5.6	5.1	1.0
Bamboo /stone/trunks with mud	-.235	43.6	37.0	12.7
Uncovered adobe	-.113	9.7	10.3	1.3
Reused	-.039	2.9	2.3	1.6
Cement	.378	1.6	3.9	24.2
Covered adobe	.124	33.2	38.4	54.9
Other	-.041	3.4	2.9	1.7
Cooking fuel				
Biogas	.016	0.0	0.0	0.1
Kerosene	.078	0.0	0.0	0.6
Charcoal	.763	0.7	3.7	47.0
Wood	-.512	82.1	83.3	44.6
Straw	-.107	16.7	11.4	3.3
Other	.079		0.5	1.0

The scree plot in Figure 2.1 is used to show the proportion of the variance explained by each principal component. It is observed that only 2 components suffice to explain the original variables. In the creation of the household asset index, the first factor

Table 2.9. Component scores and classification into wealth quintile (continuation)

Variables	Component score	Poorest 40%	Middle 40%	Richest 20%
Source of drinking water				
Piped into dwelling	.285	0.0	0.0	1.7
Piped to yard	.647	0.0	0.6	24.3
Public tap water	.147	12.6	33.2	37.7
Borehole	-027	1.7	3.2	2.0
Protected well	-.032	2.4	2.7	2.2
Unprotected well	-.054	2.3	1.9	0.9
Protected spring	-.288	52.4	32.3	19.1
Unprotected spring	-.157	18.3	14.4	5.0
River/dam/lake/pond water	-.085	9.6	9.7	3.1
Rain water	-.009	0.3	0.5	0.3
Bottled	.139	0.0	0.0	0.7
Other	.55	0.3	1.6	2.9
Ownership of durable goods				
Has electricity	.804	0.0	0.8	53.5
Has radio	.287	38.7	75.2	87.4
Has television	.760	0.0	0.1	30.8
Has bicycle	.065	4.8	21.8	20.9
Has motorcycle/scooter	.194	0.0	0.2	5.1
Has watch	.293	6.8	30.9	40.6
Has refrigerator	.569	0.0	0.0	7.9
Has car/truck	.471	0.0	0.0	5.4
Has mobile phone	.503	3.3	56.6	86.6
Own land usable for agriculture	-.463	97.7	77.3	53.3
Own livestock	-.196	60.4	59.6	43.7
Has computer	.562	0.0	0.0	9.5
Number of rooms for sleeping	.247	18.0	35.7	47.0
KMO	0.786			
Bartlett test	χ^2=238721.7	p-value< .0001		

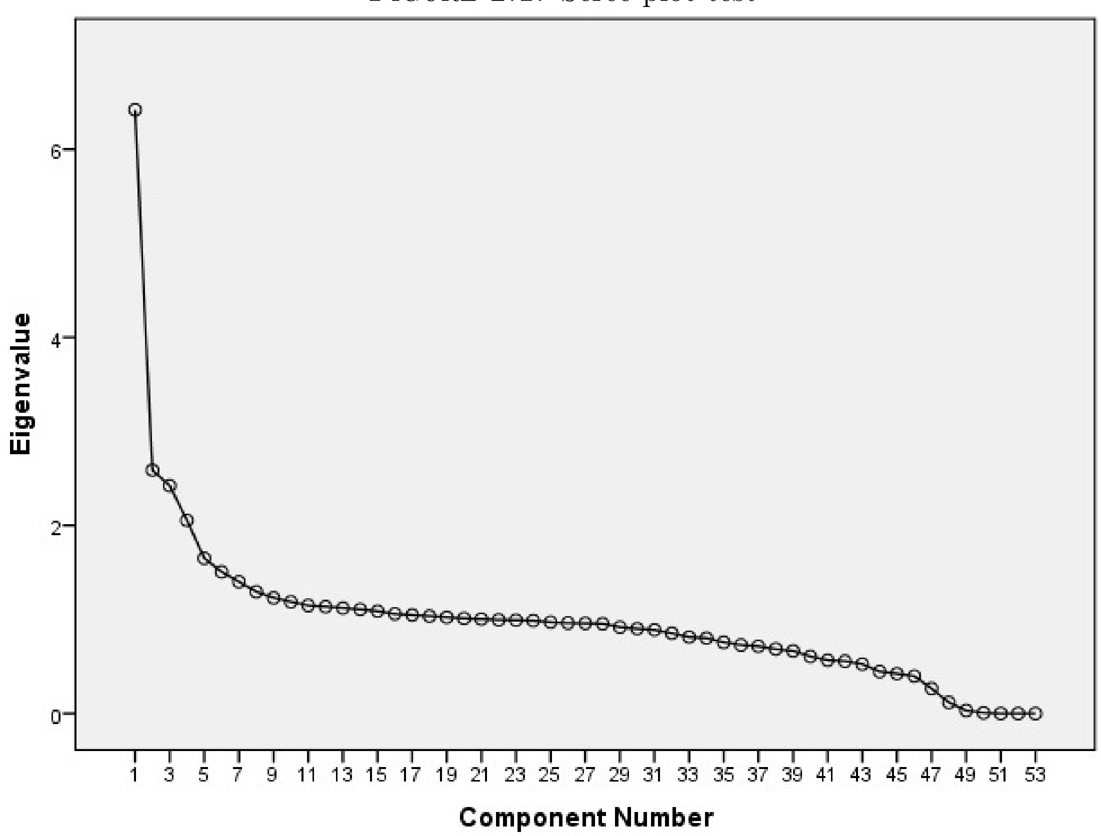
score of the first principal component is used.

The reliability test of asset index: The internal coherence is tested in Tables 2.8 and 2.9, where the last three columns compare the average ownership of each asset across the poor, middle and richest households. The robustness is tested in Table 2.10 and can be found by comparing the differences between the ranking of the poorest

TABLE 2.10. Difference in the classification of the households on the original index two assets indexes constructed from different sets of variables

Full asset index	index with 12 asset ownership variables		
Full asset index	Bottom 40 %	Middle 40 %	Richest 20 %
Bottom 40 %	83.5	16.5	0.0
Middle 40 %	11.8	74.7	13.5
Richest 20 %	4.5	25.3	70.2
Full asset index	**index with 6 housing infrastructure**		
Full asset index	Bottom 40 %	Middle 40 %	Richest 20 %
Bottom 40 %	63.8	35.7	0.5
Middle 40 %	35.9	58.1	6.1
Richest 20 %	0.8	12.4	86.9

FIGURE 2.1. Scree plot test



40% of the households of the original asset index and their ranking based on the indexes constructed using some subsets of different variables. We have used 12 variable indicators of durable goods and seven variable indicators from housing infrastructure (toilet facility, wall material, floor material, roof material, source of drinking water, source of cooking fuel) Table 2.10. The asset index produced a similar classification when different subsets of variables were used Table 2.10. Therefore, this asset index is robust.

Assessment of the demographic and spatial profiles of the poor is based on the principal component scores and household ranking into five quintiles from the poorest to the richest, where the first two quintiles are commonly classified as poorer and poor (40%), the third and fourth quintiles as middle (40%) and the fifth quintile as richest (20%). Therefore, in this study, the first two quintiles are considered as cut-off points (40%) and computed a dichotomous variable (socio-economic status or SES) indicating whether the household is poor or not (Habyarimana et al., 2015a; Achia and Khadioli, 2010; Vyas and Kumaranayake, 2006; Filmer and Pritchett, 2001, 1998). A household is classified as poor if the household poverty index is below 40% percentile, otherwise it was classified as not poor. It is given by

$$\text{SES} = \begin{cases} 1, & \text{if household is poor} \\ 0, & \text{otherwise} \end{cases} \tag{2.15}$$

2.4. Summary

In this study the 2010 Rwanda demographic and health survey data is used. The data of interest is from the household questionnaire in case of poverty study and women questionnaire in case of malnutrition.

The poverty index was created based on principal component analysis, and thereafter it was used to classify each household in socio-economic status (whether a household is poor or not).

The prevalence of poverty is higher in households headed by female 50.5% and it is also higher in rural household, where 54.9% of households are poor. The main advantage of this method over the classical methods based on income and consumption expenditure is that it avoids many of the measurement problems associated with the classical method, such as recall and seasonality. This method may be very important for countries which lack the requisite household survey data to design policies and evaluate program effectiveness, but also do not have the financial or human resources to generate such information. However, the use of asset index has some limitations such as the DHS data sets which are more reflective of longer-run household wealth or living standards (Filmer and Pritchett, 2001). Therefore, if we are interested in current resources available to households an asset based index may not be the right measure (Falkingham and Namazie, 2002)

CHAPTER 3

Ordinal survey logistic regression in the measure of poverty and malnutrition

There are some situations where the response variable has more than two categories such in nutrition status case, where nutrition status of the child can be categorized as severely malnourished, moderately malnourished and nourished. This outcome variable may be ordinal when considering ordered categorical outcomes or multinomial when non-ordered categorical outcome is considered. The data from Demographic and Health Survey are collected using multistage sampling with complex sampling design. Therefore, in order to get valid statistical inferences it is essential to account for the complexity of sampling design as failure to do so may result in biased estimates and underestimation of the variabilities. Therefore, in this chapter, we use binary and ordinal survey logistic regression models. These models offer an option for accounting for complexity of sampling design. In addition ordinal survey logistic regression also accounts for ordering level of outcome variables that are more than two.

3.1. Ordinal logistic regression

The ordinal logistic regression falls into the class of generalized linear models. This approach is used when the outcome variables are three or more and when the information from ordered categorical outcomes are for interest. The widely used ordinal logistic regression models are proportional odds models, partial proportional odds model without restriction (PPOM-UR) and with restriction (PPOM-R), continuation ration model (CRM) and stereotype model (SM) (Abreu et al., 2008; Ananth and Kleinbaum, 1997).

Proportional odds model

The proportional odds model (POM) also called ordinal logistic regression or cumulative logit model (McCullagh, 1980; Powers and Xie, 2000; Agresti, 2002; William, 2006; Freese and Long, 2006; Agresti, 2007; O'Connel, 2006; Liu, 2009) is a commonly used model for the analysis of ordinal categorical data and comes from the class of generalized linear models. It is a generalization of binary logistic regression model when the response variable has more than two ordinal categories. The proportional odds model is used to estimate the odds of being at or below a particular level of response variable. For instance, if there are j levels of ordinal outcome, the model makes $J-1$ predictions, each estimating the cumulative probabilities at or below the j^{th} level of the outcome variable. This model can also estimate the odds of being at or beyond a particular level of the response variable. The ordinal logistic regression is expressed in logit form as

$$\text{logit}\left[P_r(Y \leq j|X)\right] = ln\left\{\frac{Pr(Y \leq j|x_1, x_2, ..., x_p)}{Pr(Y > j|x_1, x_2, ..., x_p)}\right\} = \gamma_j + \beta_1 x_1 + \beta_2 x_2 ... + \beta_p x_p \quad (3.1)$$

where $Pr(Y \leq j|X) = Pr(Y \leq j|x_1, x_2, ..., x_p) = Pr_j(X)$ is the probability of being at or below category j, given a set of predictors $X = (x_1, x_2, ..., x_p)$, γ_j are the cut points (intercepts) and $\beta = \beta_1, \beta_2, ..., \beta_p$ are the logit coefficients. The cumulative logits associated with being at or below a particular category j can be exponentiated to arrive at the estimated cumulative odds and then used to find the estimated cumulative probabilities associated with being at or below category j. Equation (3.1) is POM in SAS formulation using the ascending option. We can also use descending option to get

$$\text{logit}\left[pr(Y \geq j|X)\right] = ln\left\{\frac{Pr(Y \geq j|x_1, x_2, ..., x_p)}{Pr(Y < j|x_1, x_2, ..., x_p)}\right\} = \gamma_j + \beta_1 x_1 + \beta_2 x_2 ... + \beta_p x_p \quad (3.2)$$

where $Pr(Y \geq j|X) = Pr(Y \geq j|x_1, x_2, ..., x_p)$ represent the probability that a response falls in a category equal or bigger than the j^{th} category, γ_j, X, β are the same as in equation (3.1). In this model, the effect of each predictor is assumed to be the same across the categories of the ordinal dependent variable. This means that for each predictor, the effect on the odds of being at or below any category remains the same within the model. This restriction is known as the proportional odds, or the parallel lines assumption, and is explained in the next subsection.

Partial proportional odds model

As the proportional odds assumption is difficult to achieve in practice and generalized ordered logit regression model sometimes gives more parameters than is needed, the alternative way is to fit the data with partial proportional odds model (Koch et al., 1985; Peterson and Harrel Jr, 1990; Ananth and Kleinbaum, 1997; William, 2006). This model allows some co-variables included in the model to be modeled with the proportional odds assumption, but for those variables in which this assumption is not satisfied it is increased by a coefficient(α), which is the effect associated with each j^{th} cumulative logit, adjusted by the other co-variables.

The general form of the model is the same as the proportional odds model, but now the coefficients are associated with each category of the response variable. The partial proportional odds model can be classified as unrestricted partial proportional odds (PPOM-UR) and restricted partial proportional odds model (PPOM-R).

Unrestricted partial proportional odds model

Let us consider $X = (x_1, x_2, ..., x_p)$ as a vector of p explanatory variables and assume that the first q co-variables do not satisfy the proportional odds assumption. The unrestricted partial proportional odds model is used when proportional chances assumption is not valid and the coefficients are associated with each category of the response variable (in the case of both parallel and linear assumption are not fulfilled). The PPOM-UR is given by

$$\begin{aligned} \text{logit}\left[P_r(Y \leq j|X)\right] &= ln\left[\frac{Pr(Y \leq j|x_1, x_2, ..., x_p)}{Pr(Y > j|x_1, x_2, ..., x_p)}\right] \\ &= \gamma_j + (\beta_1 + \alpha_{j1})\, x_1 + ... + (\beta_q + \alpha_{jq})\, x_q \\ &+ \beta_{q+1}x_{q+1} + ... + \beta_p x_p, j = 1, ..., J-1 \end{aligned} \tag{3.3}$$

where Y is the response variable, $X = (x_1, x_2, ..., x_p)$ is the vector of explanatory variables, $Pr\,(Y \leq j|X) = Pr\,(Y \leq j|x_1, x_2, ..., x_p)$,$\gamma_j$ are intercepts, $j = 1, 2, ..., J-1$, and $\beta_1, \beta_2, ..., \beta_p$ are logit coefficients and $\alpha = (\alpha_{j1}, \alpha_{j2}, ..., \alpha_{jq})$ are the increased coefficients to the covariate which failed the proportional odds model. The equation (3.3) is valid when the proportional odds assumption is not valid. Note that when $\alpha = 0$, the equation (3.3) reduces to proportional odds model equation (3.1).

Restricted partial proportional odds model

When the relationship between covariate and response variable is not proportional, a kind of tendency is frequently expected (Abreu et al., 2008). In this case Peterson and Harrel Jr (1990) proposed a model that is applicable when there is a linear relationship between the logit for a co-variable and the response variable (Ananth and Kleinbaum, 1997; Abreu et al., 2008; Abreu, 2009).

In this case, restrictions (represented by α parameters and which are fixed scalars) can be inserted as parameter in order to incorporate this linearity and the model is given by

$$\begin{aligned} \text{logit}\left[P_r(Y \leq j|X)\right] &= ln\left[\frac{Pr(Y \leq j|x_1, x_2, ..., x_p)}{Pr(Y > j|x_1, x_2, ..., x_p)}\right] \\ &= \gamma_j + \omega_j\left[\left(\beta_1 + \alpha_1\right)x_1 + ... + \left(\beta_q + \alpha_q\right)x_q\right] \\ &+ \beta_{q+1}x_{q+1} + ... + \beta_p x_p, j = 1, ..., J-1 \end{aligned} \tag{3.4}$$

where Y is the response variable, $X = (x_1, x_2, ..., x_p)$ is the vector of explanatory variables, $Pr\left(Y \leq j|X\right) = Pr\left(Y \leq j|x_1, x_2, ..., x_p\right)$, γ_j are intercepts, $j = 1, 2, ..., J-1$, and $\beta_1, \beta_2, ..., \beta_p$ are logit coefficients, ω_j are fixed scale parameters that take the forms of restrictions allocated to the parameters and $\alpha = (\alpha_1, \alpha_2, ..., \alpha_q)$ are the increased coefficients to the covariate which failed the proportional odds model. The equation (3.4) is used when the proportional odds assumption is not satisfied and there is a linear relationship for odds ratio between co-variates and the response variable (Ananth and Kleinbaum, 1997; Abreu et al., 2008).

Continuation odds ratio

Fienberg (1980) proposed the continuation ratio logistic regression model, that compares the probability of a response variable equal to a given category. The odds are found by considering the probability of being at or below a category relative to the probability of being beyond that category. Suppose instead of comparing each response to the next larger response we compare each response to all lower responses that is $Y = j$ versus $y < j, j = 1, 2, ..., J$. This model is called the *continuation ratio logistic model* and is defined in logit form (Hosmer et al., 2000). The CR model also estimates odds of being in a particular category j relative to being that category or beyond. In this situation, the CR model can be formulated as (Ananth and Kleinbaum, 1997; Hosmer et al., 2000)

$$ln\left[\frac{Pr(Y = j|x_1, x_2, ..., x_p)}{Pr(Y > j|x_1, x_2, ..., x_p)}\right] = \gamma_j + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p \tag{3.5}$$

where $Pr\,(Y = j|x_1, x_2, ..., x_p)$ is the conditional probability of being in category j, conditional on being that category or beyond, given a set of predictors, $\alpha_j, j = 1, 2, ..., J$ are the cut points, and $\beta_1, \beta_2, ..., \beta_p$ are logit coefficients. The CR model can also estimate the conditional probability of being beyond a category given that individual has attained that particular category, that means, $Pr\,(Y > j|Y \geq j)$, the CR model can be expressed in the form (Allison, 1999; Hosmer et al., 2000; O'Connell, 2006; O'Connel and Liu, 2011; Agresti, 2007):

$$ln\left[\frac{Pr(Y \geq j|x_1, x_2, ..., x_p)}{Pr(Y = j|x_1, x_2, ..., x_p)}\right] = \gamma_j + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p, j = 1, ..., J, \quad (3.6)$$

where $Pr\,(Y \geq j|x_1, x_2, ..., x_p)$ is the conditional probability of being beyond a category j, conditional on being in that category, given a set of predictors, $\gamma_j, j = 1, 2, ..., J-1$ are the cut points, and $\beta_1, \beta_2, ..., \beta_p$ are logit coefficients. The advantage of CRM is that the CRM can be adjusted according to k binary logistic regression models (Hosmer et al., 2000; Abreu et al., 2008). This model is more appropriate when there is intrinsic interest in a specific category of the response variable, and not merely an arbitrary grouping of continuous variables (Ananth and Kleinbaum, 1997; Abreu et al., 2008). However, the CRM is affected by the direction chosen to model the variable; this means the property of coding invariance does not hold for this model (Greenland et al., 1994) and this is its main weakness. The OR is obtained when modeling increasing severity is not equivalent to the reciprocal obtained when modelling decreasing severity (Abreu et al., 2008). Therefore, one cannot simply invert the coefficient's signal to change directions in the comparison, as happens in proportional odds models or binary logistic regression models (Scott et al., 1997; Abreu et al., 2008).

Stereotype logistic model

The stereotype logistic model (SLM) must be used when the outcome variable is intrinsically ordinal and not a discrete version of some continuous variables. It is the most flexible model for analyzing ordinal responses. SLM can also be considered as an extension of the multinomial regression model (Greenland et al., 1994) and is given by

$$ln\left[\frac{P(Y = j|X)}{P(Y = 0|X}\right] = \gamma_j + \omega_j\left(\beta_1 x_1 + \beta_p x_2 ... + \beta_p x_p\right) \quad (3.7)$$

Because of the ordinal nature of the data, a linear structure is imposed on this model. In other words, weights are assigned to the coefficients given by $\beta_{jl} = \omega_j \beta_l, j = 1, 2, ..., k$ & $l = 1, 2, ..., p$ equation (3.7). In addition to the weights (ω_k) for the response variable Y, there is a beta parameter for each explanatory variable. These weights are straightforward related to the effect of the covariates. Therefore, the OR that is obtained will have an increasing trend, as the weights are normally constructed by the ordering ($0 = \omega_1 \leq \omega_2 \leq ... \leq \omega_j$ (Mery, 2009). Then the effect of the covariates on the first OR is less than the effect on the second, and so on (Walter et al., 2001). The main challenge with this modelling is to determine these weights, but some possibilities exist (Mery, 2009). Greenland et al. (1994) suggests that the weights can be decided in advance; this means that values are appropriately chosen or estimated, based on data from a pilot study, or using generalized linear model (McCullagh and Nelder, 1983) that estimate the weights as additional parameters in the model.

Binary logistic regression

The above theory of ordinal logistic regression can be easily modified to account for binary logistic regression. In this case the outcome variable has two levels (for instance experiencing an event=1 and not experiencing the event=0). It is formulated as

$$\begin{aligned} \text{logit}\left[Pr(Y=1|X)\right] &= ln\left[\frac{Pr(Y=1|x_1, x_2, ..., x_p)}{Pr(Y=0|x_1, x_2, ..., x_p)}\right] \\ &= \gamma + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p \end{aligned} \tag{3.8}$$

3.1.1. Maximum likelihood model fitting for cumulative logit models.

Let us consider a subject i and let $y_{i1}, y_{i2}, ..., y_{ic}$ be binary indicators of the response, with $y_{ij} = 1$ for the category j in which the response falls. This means that if $Y_i = j$ then $y_{ij} = 1$ and $y_{ik} = 0$; for; $k \neq j$. Let $\pi_j(x_i)$ denote $P(Y_i = j|X = x_i)$. Therefore, for independent observations, the likelihood function is based on the product of the multinomial mass functions for n subjects (Hosmer et al., 2000; Agresti, 2010) and is given by

$$L = \prod_{i=1}^{n} \left[\prod_{j=1}^{c} \pi_j(x_i)^{y_{ij}} \right] = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{c} [P(Y_i \leq j|x_i) - P(Y_i \leq j-1|x_i]^{y_{ij}} \right\} \tag{3.9}$$

$$= \prod_{i=1}^{n} \left\{ \prod_{j=1}^{c} \left[\frac{\exp(\alpha_j + \beta' x_i)}{1+\exp(\alpha_j + \beta' x_i)} - \frac{\exp(\alpha_{j-1} + \beta' x_i)}{1+\exp(\alpha_{j-1} + \beta' x_i)} \right]^{y_{ij}} \right\}$$

We obtain each likelihod equation by differentiating L with respect to a particular parameter and equating the derivative to zero. For simplicity, let us denote

$$G(z) = \frac{exp(z)}{1+exp(z)},$$

$$g(z) = \frac{exp(z)}{[1+exp(z)]^2}$$

Therefore, the log-likelihood equation for an effect parameter β_k is given in Agresti (2010) as

$$\sum_{i=1}^{n}\sum_{j=1}^{c} y_{ij}x_{ik}\frac{g(\alpha_j + \beta' x_i) - g(\alpha_{j-1} + \beta' x_i)}{G(\alpha_j + \beta' x_i) - G(\alpha_{j-1} + \beta' x_i)} = 0 \tag{3.10}$$

where g is the derivative of G. Iteratively methods such as Fisher scoring algorithm are then used to solve equation (3.10) and obtain the ML estimates of the model parameters.

Model selection

There are several models that can describe a given data set, therefore it is very crucial to select the simplest reasonable model that satisfactorily describes such data (Lindsey, 1997). The most frequently used approaches to select the variable that enter the model are forward, backward and stepwise. Forward selection algorithm starts with the null model (no explanatory variables) and enters one explanatory variable at a time whereas backward selection starts with a saturated model (a model with all explanatory variables) and drops one explanatory variable at time (Hosmer et al., 2000). The stepwise selection procedure uses almost the same procedure as forward selection, however stepwise has the advantage over the forward selection algorithm in that the variables already in the model are also considered for exclusion each time a new variable enters the model. Therefore, if there exists a large data set under the study, the stepwise procedures are more preferred because of their advantages of minimizing the chances of keeping redundant variables and leaving out important variables in the model. However, backward elimination is commonly used when there are only a few key predictor variables and a limited number of other potentially useful predictor variables. This means that these procedures have to be used with great caution.

In all these procedures a variable that leads to a significant change in deviance when entered or dropped from the model is retained otherwise it is dropped. The contribution of each variable to the deviance reduction is given by type 1 and type 3 analysis of effects. The type one analysis of effects depends on the sequence in which variables enter the model, whereas type 3 considers the overall model and assess the contribution of each variable to the deviance reduction regardless of the sequence in which variables enter the model. The stepwise selection of the variables terminates when all variables in the model meet the criterion to stay and no variable outside the model meet the criterion to enter.

Model checking

After fitting a model to a set of data, it is very important to enquire about the extent to which the fitted values of the outcome variable under the considered model compare with the observed values. When the agreement between the observations and the corresponding fitted values is good, then the model may be acceptable, otherwise the model is not accepted and requires to be revised. The adequacy of a model is commonly referred to as goodness-of-fit (Hosmer et al., 2000; Collet, 2003).

The goodness-of-fit in generalized linear model is mainly assessed by the log-likelihood ratio (deviance) and Pearson's chi-square statistics (Fahrmeir and Tutz, 1994; Hosmer et al., 2000; Fahrmeir and Tutz, 2001; Jiang, 2001; Collet, 2003; Kutner et al., 2005). They measure the discrepancy of fit between the maximum log-likelihood achievable and the achieved log-likelihood by fitted model. The deviance is presented below to illustrate the use of these measures. It is given by

$$D(Y, \hat{\mu}) = 2\left\{\ell(y;y) - \ell(\hat{\mu}, y)\right\} \tag{3.11}$$

where $\ell(y;y)$ is the log-likelihood under the maximum achievable (also known as saturated) model and $\hat{\mu}, y)$ is the log-likelihood under the current model. The aim is to minimize D (i.e $D(y,\hat{\mu})$) by maximizing $(\hat{\mu}, y)$. The hypothesis about the goodness-of-fit of the model is given by H_o:model is adequate vs H_1 model is not adequate. H_o is rejected if $D > \chi^2_{m-p,\alpha}$ where m is the number of observations, p is the number of parameters and α is the given level of significance. In the case of sparse or ungrouped data the deviance is unreliable (Collet, 2003) to measure the goodness of fit. But the deviance can still be used to identify important predictors. In this case, the appropriate test is the Hosmer-Lemeshow goodness of fit test (Collet, 2003). For this test firstly, the predicted probabilities ($\hat{\mu_i}'s, i = 1, 2, ..., m$) obtained using current model being checked are used to form g groups with approximately m/g subjects. One grouping strategy is the percentile strategy and it is given by Hosmer et al. (2000) as

i) Group 1 subjects are approximately m/g subjects whose $\hat{\mu_i}'s$ are less or equal to the $100/g$th percentile of all $\hat{\mu_i}'s$.

ii) Group 10 subjects are approximately m/g subjects whose $\hat{\mu_i}'s$ are more than $(1 - \frac{1}{g}) \times 100^{th}$ percentile of all $\hat{\mu_i}'s$.

iii) For $J = 2, 3, ..., g - 1$ group j subjects are approximately m/g whose $\hat{\mu_i}'s$ are greater than the $\frac{j-1}{g} \times 100^{th}$ percentile and less than or equal to the $\frac{j}{g} \times 100^{th}$ percentile of all $\hat{\mu_i}'s$. In case of large m, the frequently recommended g is 10 (Hosmer et al., 2000; Dobson, 2001; Vittingoff et al., 2005) in order for the different analyses to get consistent conclusions. Thereafter, for each group, the observed and expected frequencies of the responses $y = 0$ and $y = 1$ are determined (Hosmer et al., 2000). Then, the Hosmer-Lemeshow goodness-of-fit X^2_{HL} statistic is obtained by calculating the Pearson chi-square statistic from $2 \times g$ tables of observed and expected frequencies, where g is the number of groups. Therefore, the statistics can be written as

$$X^2_{HL} = \sum_{i=1}^{g} \frac{(O_i - N_i\bar{\pi}_i)}{N_i\bar{\pi}_i(1 - \bar{\pi}_i)} \tag{3.12}$$

where N_i is the number of total frequency of subjects in the i^{th} group, O_i is the total frequency of event outcomes in the i^{th} group, and $\bar{\pi}_i$ is the average estimated probability of an event outcome for the i^{th} group and $\bar{\pi}_i = \sum_{j=1}(m_j\hat{\pi}_j)/N_i$ and m_j is the number of subject of x_j and $0_i = \sum_{j=1}^{c_i} y_j$ is the number of responses among the c_i covariates patterns. The Hosmer-Lemeshow statistic is then compared to a critical value of the χ^2 distribution chi-square distribution with $(g - n)$ degrees of freedom, where the value of n can be specified. Therefore, if the X^2_{HL} is statistically significant, then it indicates lack of fit of the model, whereas a non-significant one indicates goodness-of-fit of the model.

The appropriateness of the link function can be assessed by refitting the model with linear predictor obtained from the original model and the square of linear predictor as explanatory variables (Collet, 2003; Vittingoff et al., 2005). When the linear predictor is statistically significant and its square linear predictor term is insignificant, the link function is appropriate. This means that the prediction given by the linear predictor is not improved by adding the square linear predictor which is basically used to evaluate the null hypothesis that the model is adequate. Alternatively, the original model can be estimated with an extra constructed variable, where for an adequate model the extra variable will be statistically insignificant (William, 2006). Moreover, the appropriateness of the link function can also be checked graphically by plotting the residuals against the fitted values and for an appropriate link, the plot should not have any systematic pattern (Collet, 2003).

Other criteria besides significance tests can help to select a good model in terms of estimating quantities of interest. The best and most commonly used is the Akaike information criterion (AIC) (Agresti, 2002, 2010). It judges a model by how close its fitted values tend to be to the true values, in terms of a certain expected value. Therefore, the estimated optimal model is the model that minimizes

$$AIC = -2\left(\text{log l-p}\right) \tag{3.13}$$

where p is the number of parameters in the model. In the case of cumulative response models, $p = k + s$, where k is the total number of response levels minus one and s is the number of explanatory effects. This penalizes a model for having many parameters. It attempts to find a model that is closest to reality. A simple model that fits adequately has an advantage of model parsimony.

Schwartz criterion (SC) or Bayesian Information Criterion (BIC) is also a measure of goodness-of-model-fit. It is given by $SC = -2Logl + p\left(\sum_j f_j\right)$ where p is the number of parameters in the model and f_j is the frequency value. In the case of cumulative responses and generalized logit model, p, k and s are the same as in equation (3.13).

The concordance index is given by

$$C = \left[(n_c + 0.5)(t - n_c - n_d)t^{-1}\right] \tag{3.14}$$

where t is the total number of pairs with different outcomes given by $n(n-1)/2$, n_c is the number of concordance pairs, n_d is the number of discordance pairs and $t - n_c - n_d$ is the number of tied pairs. According to Agresti (2002), a value C=0.5 means that the predictions were not better than random guessing, between 0.6 and 0.7 is termed as moderate, between 0.7 and 0.8 acceptable and finally and excellent if C is greater than 0.8. But as the value of c approaches 1, the better the model predictive power. Wald test(Z-test) is used to test the statistical significance of individual estimated coefficients of the ordered logit regression or partial proportional odds logit regression. For ML estimators are distributed asymptotically. This means that as sample size increases, the sampling distribution of an ML estimator becomes approximately normal. So the hypothesis is $H_0 : \beta_m = 0$, and the z-statistic follows the standard normal distribution $N(0, 1)$ given as

$$Z = \frac{\hat{\beta}_m}{\hat{\sigma}_{\hat{\beta}_m/\sqrt{n}}} \tag{3.15}$$

where β_m is the m^{th} coefficient of the model, and $\hat{\beta}_m$ is the estimator of β_m; $\hat{\sigma}_{\hat{\beta}_m}$ is the estimator of standard deviation of the coefficient β_m; n is the number of observations. If H_0 is true, the coefficient β_m of the model is not statistically significant. If H_0 is rejected at a confidence level (usually is 0.05), the coefficient β_m is significant to the response. When the sample size is small, the distribution of $\frac{\hat{\beta}_k - \beta_k}{SE}$ need not be close to standard normal. Therefore, it is better to use likelihood ratio test and confidence intervals based on the profile likelihood function.

It is very important to find an overall test for all coefficients of the model, in other words, to test whether all coefficients are simultaneously equal to zero or not. The hypothesis may be written as $H_0 : \beta_m = 0$. The likelihood ratio test can be used to test this hypothesis. It makes the comparison between the estimates obtained after the constraints implied by the hypothesis ($\beta = 0$) have been imposed to the estimates obtained without the constraints. To define the test, let M_β be the unconstrained model that includes constant γ_M and slope coefficients β_M. Let M_γ be the constrained model that excludes all slope coefficients. To test the hypothesis, the test statistic is used:

$$G^2(M_\beta) = -2\left(lnL(M_\gamma) - lnL(M_\beta)\right) \tag{3.16}$$

where LM_β is the likelihood function of the model containing all the predictor variables and LM_γ is the likelihood function of the model containing only the intercept. When the null hypothesis is true, the test statistic is distributed as chi-square with degrees of freedom equal to the number of slope coefficients. When the test statistic falls into the rejection region, p-value is less than a confidence level (usually is 0.05), then the null hypothesis is rejected. Therefore, as conclusion not all slope coefficients are equal to 0. This means that at least one predictor variable significantly affects the model response.

Model diagnostic is very important; it helps to identify observations which may have undue influence on the model fit or that might be outliers. An outlier is a datum point that differs from the general trend of the data and is not necessarily influential (Lindsey, 1997). With an influential point, a small amount change or omitted, will change considerably the parameter estimates of the model. The magnitude of influential is measured by the leverage denoted by h_{ii}, which is the i^{th} diagonal element of the hat-matrix, with $0 \leq h_{ii} \leq 1$ (Lindsey, 1997; Kutner et al., 2005). In the case of generalized linear models, the hat-matrix is given by

$$H = V^{\frac{-1}{2}} X \left(X'VX\right) X'V^{\frac{-1}{2}} \tag{3.17}$$

where X is the design matrix of the known covariates and W is a diagonal weight matrix with i^{th} diagonal element given by

$$V = \frac{1}{var(Y_i)[g'(\mu_i)]^2}$$

The most frequently used measure for detection of influential data points is the Cook's distance given by Lindsey (1997) as follows

$$C_i = \frac{r_{p_i}^2 h_{ii}}{(1 - h_{ii})^2} = \frac{r_{p_{is}}^2 h_{ii}}{(1 - h_{ii})} \tag{3.18}$$

with $r_{pi} = (1 - h_{ii})\, y_i$-the Pearson's residual and $r_{pis} = r_{pi}/\sqrt{1 - h_{ii}}$ is the standardized Pearson's residual (Lindsey, 1997; Kutner et al., 2005). A large C_i means that the i^{th} observation has undue influence on the set of parameter estimates and most commonly used cut-off value of C_i is 1.

The score test statistic is used to test the validity of the proportional odds model. A nonsignificant test is taken as evidence that the logit surfaces are parallel and that the odds ratios can be interpreted as constant across all possible cut points of the outcome. If this assumption is violated it may lead to wrong interpretations (Ananth and Kleinbaum, 1997). This test is nonconservative (that is, it rejects the assumption very often) (Peterson and Harrel Jr, 1990; Bender and Grouven, 1998). Therefore, it is convenient to use other tests such as Brant test to find the single score test for each explanatory variable; this test can show which variable violated or did not violate the proportional odds assumption.

Brant (1990) proposed a Wald test to assess the parallel lines or proportional odds assumption of the ordinal regression model. This test allows both overall tests; that the coefficients for overall variables are equal and tests the equality of the coefficients for individual variables.

For overall test, $k-1$ binary regression are constructed as following: $z_j = 1$ if $Y > j$ and 0 otherwise with $j = 1, 2, 3, ..., k-1$. Therefore, we have

$$logit\,[Pr\,(z_j|X)] = \alpha_j + X\beta_j \tag{3.19}$$

The hypothesis of overall test is

$$H_0 : \beta_1 = \beta_2 = \beta_3 = ... = \beta_{k-1} = \beta \tag{3.20}$$

A Wald test statistic is derived as chi-square with $(k-2)m$ degrees of freedom, where m is the number of explanatory variables. For the m^{th} individual variable, the null hypothesis is

$$H_0^m : \beta_{m,1} = \beta_{m,2} = \beta_{m,3} = ... = \beta_{m,K-1} = \beta_m \tag{3.21}$$

The resulting test statistics follows χ^2 distribution with $k-2$ degrees of freedom. If the probability of these tests (p-value) is less than 0.05 (usually), the hypothesis is rejected; in other words, this indicates that there are evidences for the violation of the assumption for overall variables or individual ones.

When the proportional odds assumption is not valid the alternative way is to fit the data with partial proportional odds model (Koch et al., 1985; Peterson and Harrel Jr, 1990; Ananth and Kleinbaum, 1997). Another alternative is to dichotomize the ordinal outcome variable by means of several cut-off points and then use separate binary logistic regression model for each dichotomous outcome variable (Bender and Grouven, 1998). However, Gameroff (2005) suggested that the separate binary logistic regression model should be not used if possible because of the loss in statistical power and reduced generality of analytical solution.

3.2. Ordinal survey logistic regression

Some standard statistical methods used when analyzing the data collected under simple random sampling, where each sampling unit has the same probability of being chosen from the population, are not convenient for analyzing the data collected using complex survey sampling designs, where stratified sampling and clustered sampling are used (Anthony, 2002; Liu and Koirala, 2013). Therefore the survey logistic regression models are needed to adjust the classical logistic regression models in order to account for complexity of sampling designs. The survey sampling design may induce correlation among observations, especially when clusters samples are drawn. To appropriately estimate standard errors associated with the model parameters and estimated odds ratios, it is very crucial to account for sampling design. The survey logistic regression models have the same theory as classical logistic regression models. The only difference is the estimation of the variance. However, when these two models are used to the data collected using simple random sampling, the results are identical.

Therefore, the main objective of this section is to extend ordinal logistic regression models to ordinal survey logistic regression models that accounts for the complexity of survey design, in other words, it takes into account the effects of stratification and clustering used in the survey design.

Model overview

Let Y_{ijh} be the response variable, with $i = 1, 2, 3, ..., m_{hj}$, $j = 1, 2, 3, ..., n_h$ and $h = 1, 2, 3, ..., H$, where h is the stratum, j is the cluster and i is the household and denote the sampling weight for ijh^{th} observation as w_{ijh} and x_{ijh} the row vector of the design matrix corresponding to the i^{th} household in j^{th} PSU, nested in h^{th} stratum. Therefore, the survey logistic model is given by

$$logit(\pi_{ijh}) = x'_{ijh}\beta \tag{3.22}$$

where β is the vector of unknown parameters.

When the survey data have been collected under complex sampling design, straightforward application of classical maximum likelihood estimation (MLE) is no longer convenient, for various reasons. The first one is that the probabilities of selection for the $i = 1, 2, ..., n$ sample observations are no longer equal. Sampling weights are then required to estimate the finite population values of the logistic regression model parameters. Secondly the stratification and clustering of complex sample observations violates the assumption of independence of observations that is essential to the standard MLE method (Heeringa et al., 2010). There are two main approaches developed for estimating the logistic regression parameters and standards errors for complex samples survey data.

Grizzle et al. (1969) developed an approach based on weighted least square estimation and later Binder (1983) proposed pseudo maximum likelihood estimation (PLME) as the second general approach framework for fitting logistic regression and other generalized linear models to complex sample survey data. PLME approach was combined with linearized estimator of the variance-covariance matrix for the parameter estimates and taking complex sample design into consideration.

Generally, there are many methods in literature used to estimate the variance of the parameter estimates in survey logistic. The most used are Taylor series (known as linearization method), Jackknife method, bootstrap and balanced repeated replication(BRR) methods. The pseudo-likelihood approach to the estimation of the model parameters involves maximizing the following pseudo-likelihood function (Heeringa et al., 2010):

$$PL\left(\hat{\beta}|X\right) = \prod_{i=1}^{n}\left\{\prod_{k=1}^{K}\hat{\pi_k}(x_i)^{y_i^{(k)}}\right\}^{w_i} \tag{3.23}$$

where $y_i^{(k)} = 1$ if $y = k$ for sampled unit i and 0 otherwise,w_i is the survey weight for sampled unit i and $\hat{\pi_k}(x_i)$ is the estimated probability that $y_i = k|x_i$. The maximization involves application of the Newton-Raphson algorithm to solve the estimating equations (3.24). Assuming as before a complex design with strata indexed by h and clusters within strata indexed by j:

$$S(\beta) = \sum_h\sum_j\sum_i w_{hji}\left(y_{hji}^{(k)} - \pi_k(\beta)\right)x'_{hji} = 0 \tag{3.24}$$

where $y_{hji}^{(k)} = 1$ if$y = k$ for sampled unit i, 0 otherwise; x_{hji} is a column vector of design matrix; $\beta = \{\beta_{2,0}, ..., \beta_{2,p}, ..., \beta_{K,0}, ..., \beta_{K,p}\}$ is a vector of unknown parameters and

$$\pi_k(\beta) = \frac{exp(x'_{hji}\beta_k)}{1 + \sum_{k=1}^{K} exp(x'_{hji}\beta_k)} \tag{3.25}$$

The above theory of ordinal survey logistic regression can be modified to include the case of binary outcomes.

In the literature, there are a number of methods used to estimate the variance-covariance matrix of the estimated parameters. The most used are Taylor linearization method and replicated or resampling methods (Jackknife, bootstrap, balanced replication, random groups)(Wolter, 2007).

Variance estimation

Because of the variability of characteristics between items in the population, researchers use sample designs in the sample selection process to reduce the risk of distorted view of the population, and they make inference about the population based on the information from the sample survey data. In order to make statistically valid inferences for the population, they must incorporate the sample design in the analysis. There are a number of techniques used to estimate the variance, but they are often classified into the following two categories: model based methods and resampling methods. Model based methods include Taylor series approximation whereas resampling methods include Jackknife, balanced repeated replication (BRR) methods and numerous variant thereof (Efron and Tibshirani, 1993).

Taylor Series(linearization) method

The Taylor series approximation is the most frequently used method to estimate the covariance matrix of the regression coefficients for complex survey data. This method relies on simplicity associated with estimating the variance of linear statistics, even with a complex sample design. By applying the Taylor linearization method, non-linear statistics are approximated by linear forms of observation by taking the first-order terms in an appropriate Taylor series. Extending the Taylor series expansion could develop second-order approximations. However, in practice, the first-order approximation usually yields satisfactory results, with the exception of highly skewed population (Wolter, 1985). The estimation of variance of the general estimator is adapted from the Taylor series expansion. To use the Taylor series expansion, consider a finite population of size N. Let k-dimensional parameter vector be denoted by $Y = (Y_1, Y_2, ..., Y_k)'$ and let $\hat{Y} = \left(\hat{Y}_1, \hat{Y}_2, ..., \hat{Y}_k\right)'$ be the corresponding vector of estimators based on a sample size s of $n(s)$ (Lehtone and Pahkinen, 2004). Then, the estimators $\hat{Y}_i, i = 1, 2, ..., k$ depend on the sampling design generating the samples (Wolter, 2007). In many applications of Taylor series methods, Y_i represent population totals or means for k different survey characteristics and $\hat{Y}_i$ denote standard estimators of Y_i. Generally, $\hat{Y}_i$ are unbiased estimators for Y_i, however in some applications they might be biased but consistent estimators. Suppose that the population parameter of interest is $\theta = h(Y)$ and its consistent estimator is denoted by $\hat{\theta} = h(\hat{Y})$. Then, the main interest is to find the approximate expression for the design variance of $\hat{\theta}$ and constructing an appropriate estimator of the variance of $\hat{\theta}$ (Wolter, 2007). Let us assume that $h(Y)$ is twice continuously differentiable. Then, based on Taylor series principles, specifically the linear terms of the Taylor-series expression, the approximate linearized expression is given (Wolter, 2007) by

$$\hat{\theta} - \theta = \sum_{i=1}^{k} \frac{\partial h(Y)}{\partial Y_i} \left(\hat{Y}_i - Y_i\right) \tag{3.26}$$

where, as usual, $\frac{\partial h(Y)}{\partial Y_i}$ refer to partial derivative of $h(Y)$ with respect to y_i. Using equation (3.26), the variance approximation of $\hat{\theta}$ can be defined as

$$V(\widetilde{\theta}) = V\left(\sum_{i=1}^{s} \frac{\partial h(Y)}{\partial Y_i}\left(\hat{Y}_i - Y_i\right)\right) = \sum_{i=1}^{s}\sum_{j=1}^{s} \frac{\partial h(Y)}{\partial Y_i}\frac{\partial h(Y)}{\partial Y_j} V\left((\hat{Y}_i, \hat{Y}_j)\right) \tag{3.27}$$

where $V\left(\hat{Y}_i, \hat{Y}_j\right)$ are variances and covariances of the estimators $\hat{Y}_i$ and $\hat{Y}_j$. Therefore, the variance of a non linear estimator $\hat{\theta}$ is now reduced to a function of variances and covariance of s linear estimators $\hat{Y}_i$ (Wolter, 2007). Further, the variance estimator $\hat{V}(\hat{\theta})$ is obtained from equation (3.27) by substituting the variance and covariance estimators $\hat{v}\left(\hat{Y}_i, \hat{Y}_j\right)$ for the corresponding parameters $V\left(\hat{Y}_i, \hat{Y}_j\right)$(Skinner et al., 1989). The resulting variance is called first order approximation. Extending the Taylor series expansion could develop second or even higher order approximations. However, in practice, the first order approximation usually yields satisfactory results, with the exception of highly skewed populations (Wolter, 1985, 2007). Standards variance estimation techniques can then be applied to the linearized statistic. The Taylor linearization method is a widely applied method, quite straightforward for any case where an estimator already exists for totals. Its bias originates from its tendency to underestimate the true value and it relies on the size of the sample as well as the complexity of the estimated statistic. However, if the statistic is fairly simple, for instance like the case of the weighted sample mean, then the bias is negligible even for small samples, while it becomes nil for large samples (Sarndal et al., 1992). On the other hand for a complex estimator such as the variance, large samples are needed before the bias becomes small.

Replication method/resampling method

Replicate variance estimation is a robust and flexible method which can reflect a number of complex sampling and estimations used in practice. Replication approach can be used with a wide range of sample designs such as multi-stage, stratified and unequal probability samples. It can also reflect the effects of various type of estimation technique. The main concept of replication approach is based on the originally derived sample (full sample) from which we take a number of small samples (replicate samples). From each replicate we estimate the statistic of interest, and the variability of these replicates estimates is used in order to derive the variance of the statistic of the full sample.

Let θ be an arbitrary parameter of interest, $\hat{\theta} = f(data)$ its estimate (the statistic of interest) and $v(\hat{\theta})$ its corresponding variance given by

$$\hat{v}(\hat{\theta}) = c\sum_{k}^{H} h_k \left(\hat{\theta}_{(k)} - \hat{\theta}\right)^2 \tag{3.28}$$

where $\hat{\theta}_{(k)}$ is the k^{th} replicate sample estimate of θ, H is the total number of the replicates, c is a constant that depends on the replication method and h_k is a stratum specific constant (needed only for some sampling structures). There are various methods for drawing these replicate samples, leading then to a large number of replication methods for variance estimation. The most frequently used are Jackknife, bootstrap, balanced repeated replication and random groups.

Jackknife

The Jackknife technique originated outside the field of survey sampling. It was first developed by Quenouille (1949, 1956) as a method of reducing bias of an estimator in an infinite population setting. Durbin (1958) is one who first introduced it for finite population, and then the procedure was adopted to estimate variance and associated confidence intervals. Miller (1974) reviewed the possible uses of the Jackknife technique in a range of statistical applications. In the case of variance estimation, Jackknife technique consists of splitting the total sample into a set of equal-sized, disjoint, exhaustive subsamples, dropping out each of the samples in turn, and estimating the population parameter of interest from the remaining units each time. The variability between the estimates can therefore be used to estimate the variance of the original sample estimator (Rust, 1985). The dropped part is re-entered in the sample and the process is repeated successively until all parts have been removed once from the original sample. These replicated statistics are used in order to calculate the corresponding variance. With stratified cluster data each cluster is deleted in turn, and then the variance calculations are done inside the strata. Then, the Jackknife bias and variance estimates are given by

$$b_J = \sum_{h=1}^{H} (n_h - 1)\left(\bar{\hat{\theta}}_h - \hat{\theta}\right),$$

$$v_J = \sum_{h=1}^{H} \frac{(1 - f_h)(n_h - 1)}{n_h} \sum_{j=1}^{n_h} \left(\hat{\theta}_{hj} - \bar{\hat{\theta}}_h\right)^2 \tag{3.29}$$

where f_h is the proportion of clusters sampled in the h^{th} stratum, $\hat{\theta}_{hj}$ is the estimate recalculated without the j^{th} cluster of stratum h and $\hat{\theta}_h$ is the average of the estimates for that stratum and this needs a total of $\Sigma_h n_h$ recalculations of the statistic.

Disjoint parts mentioned above can be either single observation in a simple random sampling or clusters of units in multistage cluster sampling schemes. The choice of the way that sampling units are entered and re-entered in the sample leads to a number of different expressions of Jackknife variance. For instance in Jackknife-1 method (that is more suitable for unstratified design) one sampling unit or element or cluster is excluded at each time. But in Jackknife-2 (more suitable for stratified samples with two primary sampling units per stratum) and Jackknife-n (more appropriate for stratified samples with more than two primary sampling units per stratum) a single primary sampling unit is deleted from a single stratum in each replication. It should be noted that the Jackknife method for variance estimation is more applicable in with-replacement designs, though it can also be used in without-replacement surveys when the sampling fraction is small (Wolter, 1985, 2007).

Shao and Tu (1995) mentioned that the application of Jackknife requires a modification to account for sampling fractions only when the first stage sampling is without replacement. In any case, due to their nature, Jackknife variance estimation methods seem to be more appropriate for single or multistage cluster designs, where in each replicate a single cluster is left out of the estimation procedure (neglecting, though, the finite population correction).

If the number of disjoint parts (for example clusters) is large, the calculation of replicate estimates is time consuming, making the whole process rather time-demanding in the case of large-scale surveys (Yung and Rao, 2000). So alternative Jackknife techniques have been developed (Efron, 1982).

Jackknife linearization: The idea of this technique is to replace repeated calculation of the statistic (practically numerical differentiation) by analytic differentiation. The resulting formula is simple to calculate. In addition, in large samples it yields a good approximation compared to the standard Jackknife technique. This technique is also called nonparametric delta method and the infinitesimal Jackknife (Davison and Hinkley, 1997). In the case of unstratified sample of size n, the nonparametric delta method variance approximation is

$$v_L = n^{-2}\Sigma l_j^2.$$

The empirical influence value l_j, the infinitesimal change in the statistic because of inclusion of the j^{th} observation, is closely related to the influence function central to classical robust statistics (Hampel et al., 1986). Further, replacing n^{-2} by $(n(n-1))^{-1}$ reduces the slight downward bias of v_L.

For stratified cluster data the bias-adjusted variance formula in case of sampling without replacement is given by Canty and Davison (1999) as,

$$\hat{v} = \sum_{h=1}^{H} (1 - f_h) \frac{1}{n_h (n_h - 1)} \sum_{j=1}^{n_h} l_{hj}^2 \tag{3.30}$$

where l_{hj}^2 is the empirical influence value for the j^{th} cluster in stratum h l_{hj}^2 (Canty and Davison, 1999). The effort needed for calculating l_{hj}^2 is based on the complexity of statistic. For the linear estimator in stratified cluster sampling:

$$\hat{\theta} = \sum_{h,j} y'_{hj}$$

where

$$y'_{hj} = \sum_{k} \omega_{hjk} y_{hjk}$$

is the sum of $y's$ in every cluster j in each stratum h, and ω_{hjk} is the design weights then

$$l_{hj}^2 = n_h y'_{hj} - \sum_{j} y'_{hj}$$

For the ratio of two calibrated estimators,

$$\hat{\theta} = \frac{1^T W y}{1^T W z}$$

, the chain rule gives

$$l_{hj}^2 = \frac{l_{hj}^y - \hat{\theta} l_{hj}^z}{1^T W z}$$

where where l_{hj}^y, l_{hj}^z are the empirical influence values calculated from the data analytically, y and z are the vectors of the observations in the data set.

When the sampling fractions f_h are small the formula (3.30) may be also used for sampling without replacement. Its main advantage is that it is less computationally demanding, while it usually retains the good properties of the original Jackknife method. But, if non-linear statistics are considered, the derivation of separate formulae is needed, as is the case with all linearized estimators. Then, its usefulness for complex analysis of survey data or elaborate sample designs has some limitations (Rao, 1997; Canty and Davison, 1999).

Bootstrap estimator

The bootstrap was originally designed for use with independent observations. It was developed outside the field of survey sampling theory by Efron (1979, 1982). There are still some issues that need to be investigated such as non-independence between observations in the case of sampling without replacement as well as other complexities. Several studies have been carried by Sarndal et al. (1992) and Shao and Tu (1995) among others.

However, the bootstrap main idea consists of drawing a series of independent samples from the sampled observations, using the similar sampling design as one by which the initial sample was drawn from the population and calculating an estimate for each of the bootstrap samples. Therefore, in order to get an unbiased result the variance of the bootstrap estimator is multiplied with an appropriate constant. In the case of stratified sample designs, resampling is carried out independently in each stratum. Its main disadvantage is that it is too time consuming.

Balanced repeated replication method

Balanced repeated replication (BRR) was originally developed for stratified multistage designs where in each stratum two primary sampling units (PSUs) or clusters are drawn with replacement at the first stage (McCarthy, 1969). A replicate sample or a half-sample is obtained by deleting one PSU per stratum and doubling the original weight of the remaining PSU. The BRR variance estimation of a full sample estimator $\hat{\theta}$ is given by Wolter (2007), Rust (1985) and Shao and Tu (1995)

$$V(\hat{\theta}) = H^{-1} \sum_{h=1}^{H} \left(\hat{\theta}_h - \hat{\theta} \right) \tag{3.31}$$

where $\hat{\theta}_h$ is an estimator of θ using h^{th} balanced half sample and H is the total number of replicates. For more details we refer to Wolter (2007); Rust (1985) and Shao and Tu (1995).

In a case where the clusters have variable number of units, the division of them into two groups is required and thus modifications have been developed. For example, for the stratified designs one has to treat each stratum as if it were a cluster, and to use divisions of the elements into two groups.

In non-linear cases, one or more replicate estimators $\hat{\theta}_r$ may be undefined but the full sample estimator $\hat{\theta}$ is defined. Fay's BRR method adjusts the original weight by a coefficient ϵ, with $0 \leq \epsilon < 1$ so that the replicate estimators are defined for all replicate samples. The Fay's BRR variance estimator of $\hat{\theta}$ is given by Fay (1989); Judkins (1990) and Rao and Shao (1999) as

$$V(\hat{\theta}) = \left[H\left(1-\epsilon\right)^2\right] - 1 \sum_{h=1}^{H}\left(\hat{\theta}_h - \hat{\theta}\right) \tag{3.32}$$

If $\epsilon = 0$, then Fay's BRR method reduces to the traditional BRR method in equation (3.31).

But, if there is an odd number of elements in the stratum the results are biased, and ways of reducing this bias but not eliminating it are described in Slootbeek (1998). Rao and Shao (1996) shows that only by using repeated division (repeatedly grouped balanced half samples) can an asymptotical correct estimator be obtained.

The main advantage of BRR method over the Jackknife is that it leads to asymptotically valid inferences for both smooth and non-smooth function (Rao, 1997), but it is not simply applicable for arbitrary sample sizes n_h like the bootstrap and the Jackknife techniques.

Random groups method

The main idea of random groups method of variance estimation consists of drawing a number of samples (replicates) from the population, usually using the same sampling design for each sample; estimating the parameter of interest for each replicate and assessing its variance based on the deviations of these statistics from the corresponding statistics derived from the combination of all replicates (Wolter, 1985, 2007). The random groups technique can be divided into two main fundamental variations, based on whether the replicates are mutually independent or whether there is a dependency between random groups (Sarndal et al., 1992). The independent random groups method has its origin in the work by Mahalanobis (1939, 1944, 1946) and Deming (1956). Mahalanobis called it interpenetrating samples whereas Deming called it replicated samples. This technique provides unbiased linear estimators. However, in the case of nonlinear estimator, a small technical bias may occur (WHO, 1995).

The idea of dependent random group was first described by Hansen et al. (1953). The dependent random group technique is an attempt to adapt the independent random technique to a sample that does not satisfy the requirements of independent random groups.

In the case of dependent random groups, a bias is introduced in the results, but this bias tends to be negligible for large-scale surveys with small sampling fraction. In this condition the uniformity of the underlying sampling design of each replicate is a prerequisite for safeguarding the acceptable statistical properties of the random groups variance estimator.

3.3. Application

Introduction

In this application, we have used households data in the case of poverty. The main interest was to identify the factors associated to the poverty of households, where the outcome variable was binary (a household is poor or not), and malnutrition was measured on ordinal scale (severely malnourished if z-score < -3.0, moderately malnourished if $-3.0 \leq$ z-score < -2.0 and nourished if z-score ≥ -2.0).

We considered both classical and survey binary and ordinal logistic regression in the analysis of the data and compared the results.

3.3.1. Analysis of demographic and health survey to measure poverty of household in Rwanda.

Let the response variable be 1 if the household is poor and 0 if the household is not poor. Therefore the fitted logistic regression model is given by

$$logit(\mu_i) = log\left(\frac{\mu_i}{1-\mu_i}\right) = X_i'\beta \tag{3.33}$$

where $\mu_i = E(Y_i) = Pr(Y_i = 1)$, X_i' is a vector of explanatory variables and β is a vector of unknown parameters. We considered as explanatory variables the characteristic of household head (level of education, gender, age), characteristic of household (size or number of household members) as well as spatial characteristics (province and place of residence of household) and their interaction. We used the deviance analysis for the model selection. The potential confounder was controlled by retaining all the main effects in the model. Thereafter we examined the fitting of each interaction effect one at time.

The deviance of the model with all main effects was 7256.980 and the deviance for the model with all main effects and three interactions was reduced to 7181.518. This deviance is smaller than all other nested models.

Table 3.1. Pearson chi-square statistics test for association between demographic characteristics with SES

Explanatory variable	**χ^2-value**	**df**	**P-value**
Province/Region	1115.776	4	$< .0001$
Place of residence	707.616	1	$< .0001$
Gender of the household head	283.262	1	$< .0001$
Education level of household head	1001.810	3	$< .0001$
Age of the household head	294.376	84	$< .0001$
Size of the household	243.376	17	$< .0001$

Before accepting the final model, we carried out diagnostics to see whether the model fits the data well. The goodness-of-fits was tested by Hosmer-Lemeshow test and it was 7.3263 with 8 degree of freedom with p-value=0.5019. As the value of p-value is large and nonsignificant, this shows that the model fits the data well. The observed and expected frequencies are given in Table 3.2 We tested the appropriateness of linear predictor by refitting the model with a linear predictor and its square as predictor variables. The results shown in Table 3.3 suggests that the link function is appropriate as the linear predictor was significant (p-value< .0001) whilst its square linear predictor is insignificant (p-value=0.1821) Table 3.3.

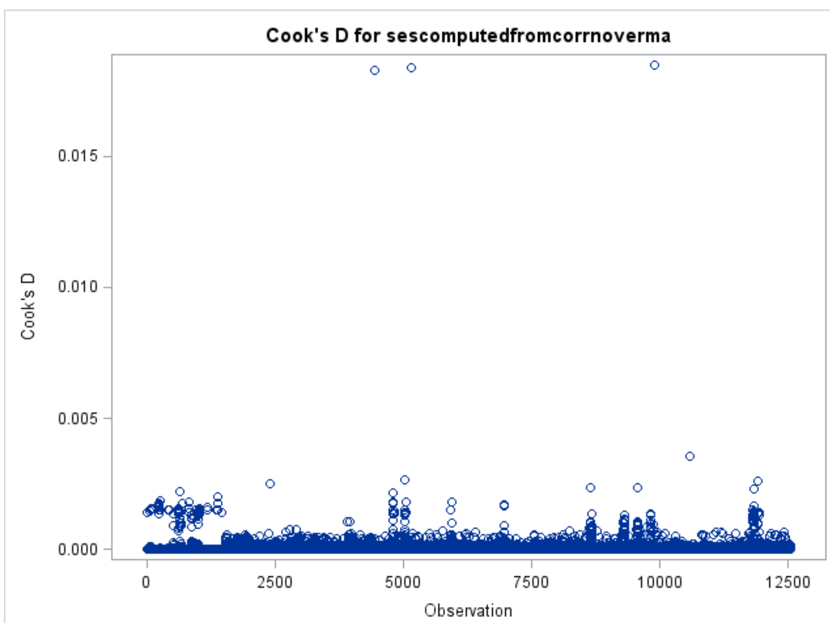
TABLE 3.2. Partition for the Hosmer and Lemeshow test

Group	Total	Event=poor		Non-event=not poor	
		observed	expected	observed	expected
1	1244	20	28.40	1224	1215.60
2	1245	192	179.52	1053	1065.48
3	1248	335	342.17	913	905.83
4	1244	431	423.73	813	820.27
5	1244	506	503.85	738	740.15
6	1244	544	547.87	700	696.13
7	1244	628	614.15	616	629.85
8	1244	653	680.07	591	563.93
9	1245	772	762.45	473	482.55
10	1242	891	889.79	351	352.21

TABLE 3.3. Criteria for assessing the link function

Effect	**Estimate**	**Standard error**	**Wald χ^2**	**df**	**P-value**
Intercept	0.0076	0.0217	0.127	1	0.7261
Linear predictor	0.9700	0.0346	788.05	1	$< .0001$
Square linear predictor	-0.0234	0.0176	1.78	1	0.1821

FIGURE 3.1. Index plot of the Cook's distance for the fitted model



From Figure 3.1, we see that none of the Cook's distance for the fitted model is bigger than 1; this suggests that there are no observations with undue influence on parameter estimate. Therefore the final fitted model is given by

$$\begin{aligned} logit\left(Pr(y_i = 1|X_i)\right) &= \beta_0 + \beta_1 Education_i + \beta_2 Province_i \\ &+ \beta_4 Place\ of\ residence_i + \beta_5 Size_i \\ &+ \beta_6 Age_i + \beta_7 Province_i * Place\ of\ residence_i \\ &+ \beta_8 Gender_i * Age_i + \beta_9 Age_i * Size_i \end{aligned} \quad (3.34)$$

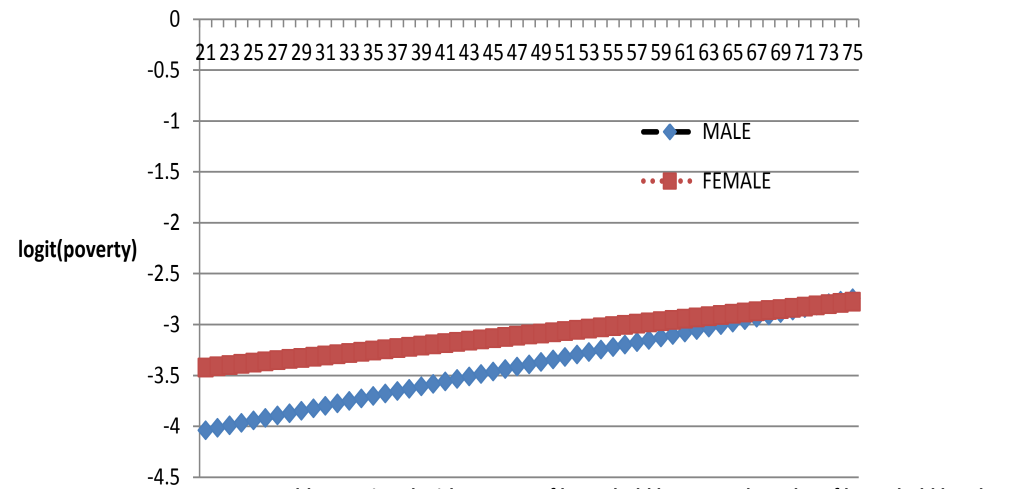
The characteristics of the household head are important to the living conditions of all household members. From Table 3.4, the logistic regression results show that the poverty increases with decreasing the level of education of the household head. A household headed by a household head with secondary education is 6.481 (p-value=0.0017) times more likely to be poor than a household headed by a household head with a higher education. A household headed by a household head with primary education is 24.416 (p-value $<$ .0001) times more likely to be poor than a household headed by a household head with a higher education, and a household headed by a household head with no education is 41.971 (p-value $<$ .0001) more likely to be poor as compared to a household headed by a household with a higher education.

Interaction effect

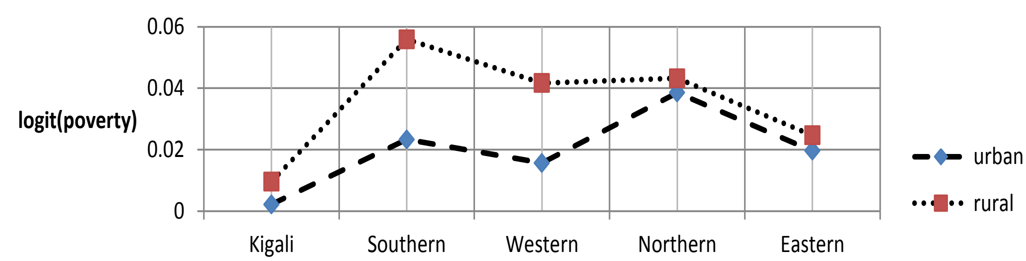
The joint effect of gender and age of the household head is presented in Figure 3.2.3.a. From Figure 3.2.3.a, we observe that a household headed by a female is more likely to be poor as compared to a household headed by a male from 21-72 years old. Furthermore, from 72 years old a household headed by a female is less likely to be poor than a household headed by a male. It is also interesting to note the relationship between age of household head and the size of the household. Figure 3.2.3.c shows that poverty decreases with the increasing age of the household head regardless of the size of the household. Furthermore, for a household headed by a young person of 21 years old, poverty increases as the size of the household increases. This result suggests that old people should not live alone and that households headed by young household head should be monitored by experienced household members.

The relationship between provinces (Kigali city, Southern, Western, Northern and Eastern) and place of residence (urban or rural) is presented in Figure 3.2.3.b. Each province of Rwanda has urban and rural places. As Figure 3.2.3.b indicates, an urban household is less likely to be poor compared to a rural household in all provinces. These results revealed that a rural household from Southern province is the poorest Figure 3.2.3.b, while rural households from Western and Northern provinces are almost the same but more likely to be poor compared to a rural household from Eastern province. A rural household from Kigali is less likely to be poor as compared to a rural household from Eastern province Figure 3.2.3.b.

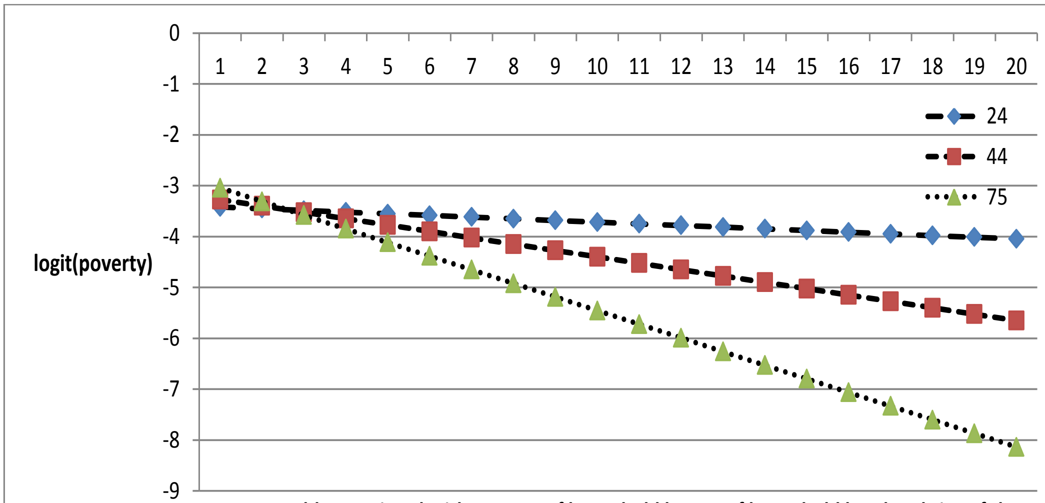
Figure 3.2. Interaction effects



3.a. Log odds associated with poverty of household by age and gender of household head



3.b. Log odds associated with poverty of household by province and place of residence of the household



3.c. Log odds associated with poverty of household by age of household head and size of the household

TABLE 3.4. Parameter estimates from binary logistic regression with main effect for poverty of households

Indicator	Estimate	S.E	P-Value	OR
Intercept	-3.6762	.6079	$< .0001$	.025
Province				
Eastern	reference			
Kigali	-.9591	.1864	$< .0001$	.383
South	.8497	.0575	$< .0001$	2.339
West	.5415	.0584	$< .0001$	1.719
North	.5796	.0636	$< .0001$	1.785
Gender of the household head				
Female	reference			
Male	-.8678	.1356	$< .0001$	.420
Education of Household head				
Higher	reference			
Secondary	1.8689	.5945	.0017	6.481
Primary	3.1952	.5870	$< .0001$	24.416
No education	3.7370	.5880	$< .0001$	41.971
Age of the household head	.012	.003224	.0002	1.012
Size of household	.0777	.0348	.0257	1.081
Place of residence				
Rural	reference			
Urban	-.2323	.2156	.2811	.793

3.3.2. Binary survey logistic regression applied to the risk factors associated to the poverty of households.

As previously stated, the main objective of this study is to identify the key determinants of poverty of households in Rwanda based on 2010 RDHS. As the data was collected under multistage sampling, this study extends Habyarimana et al. (2015a) to include the design effect.

Table 3.5. Parameter estimates from binary logistic regression with interaction effects for poverty of households

Indicator	Estimate	SE	P-Value	OR
Intercept	-3.6762	.6079	$<.0001$	.02
Province and place of residence				
Eastern and rural	reference			
Kigali and urban	-1.2489	.3284	.0001	.287
South and urban	-.6758	.2470	.0062	.509
West and urban	-7730	.3115	.0131	.462
North and urban	.1123	.3092	.7164	1.119
Gender and age of the household head				
Female	reference			
Male and age of the household	.012	.00282	$<.0001$	1.012
Size of household and age of household head	continuous variable no reference			
size and age of the household head	-.00461	.000727	$<.0001$	.995

Let the response variable be $y_{ijh} = 1$ if the i^{th}household is poor and be 0 otherwise. Then, the fitted survey logistic regression model is given by

$$logit\,(\pi_{ijh}) = log\left(\frac{\pi_{ijh}}{1-\pi_{ijh}}\right) = x'_{ijh}\beta \qquad (3.35)$$

where $\pi_{ijh} = E(y_{ijh}|x'_{ijh})$, x'_{ijh} is a vector of explanatory variables and β is a vector of unknown parameters.

Data analysis

We have used SAS 9.3 PROC SURVEYLOGISTIC procedure to analyze the data, where the deviance was used to select the best model. The model was fitted to each predictor one at time, where the significant predictor variables were used in multivariate logistic regression model. Besides the main effect, we have also included the two-way interaction effects. Afterwards, the selected model was the one of smallest changes in deviance from all nested models and it is reported in Table 3.6 and this is the full model including two-way interaction effects.

3.3.3. Results and Interpretation.

In this study we have not only considered the main effects but also the two way interaction effects.

Main effects

From Table 3.9, the logistic regression results show that poverty increases with decreasing the level of education of the household head. A household headed by a household head with secondary education is 6.859 (p-value=0.0015) times more likely to be poor as compared to a household headed by a household head with a higher education (tertiary level); a household headed by a household head with primary education is 25.175 (p-value $< .0001$) times more likely to be poor as compared to a household headed by a household head with a higher education; and a household headed by a household head with no education is 42.512 (p-value $< .0001$) more likely to be poor as compared to a household headed by a household with a higher education.

Interaction between gender and age of household head

The results of joint effect of gender and age of household head on household asset index are presented in Figure 3.5. From this figure, it is observed that a household headed by a female is 1.012 times more likely to be poor than a household headed by a male. This is in line with NISR et al. (2012); Habyarimana et al. (2015a).

Interaction between the size of household and age of the household

The joint effect of size of household and age of household head on the household asset index is presented in Figure 3.4. Figure 3.4 shows that poverty decreases with increasing age of the household head regardless of the size of the household. Furthermore, for a household headed by a young person 24 years old, poverty increases as the size of the household increases. As Figure 3.4 indicates, the poverty of household of one person increases with increasing age.

Interaction between province and place of residence

The relationship between provinces (Kigali city, Southern, Western, Northern and Eastern) and place of residence (urban or rural) of the household head is presented in Figure 3.3. Each province of Rwanda has urban and rural places; as Figure 3.3 indicates, a household from urban is less likely to be poor than a household from rural in all provinces of Rwanda, this is in line with RDHS NISR et al. (2012). This result revealed that a rural household from Southern province is the poorest, while a rural household from Western and Northern province are almost the same but more likely to be poor as compared to a rural household from Eastern province. A rural household from Kigali is less likely to be poor as compared to a rural household from Eastern province Figure 3.3. From Figure 3.3, we see that the urban household from Kigali, Western and Southern province are less likely to be poor compared to eastern province. The urban and rural households from Southern and Western province differ largely whereas the urban and rural household from Kigali, Northern and Eastern provinces the disparities are small Figure 3.3.

FIGURE 3.3. Interaction effect between province and place of residence of household head

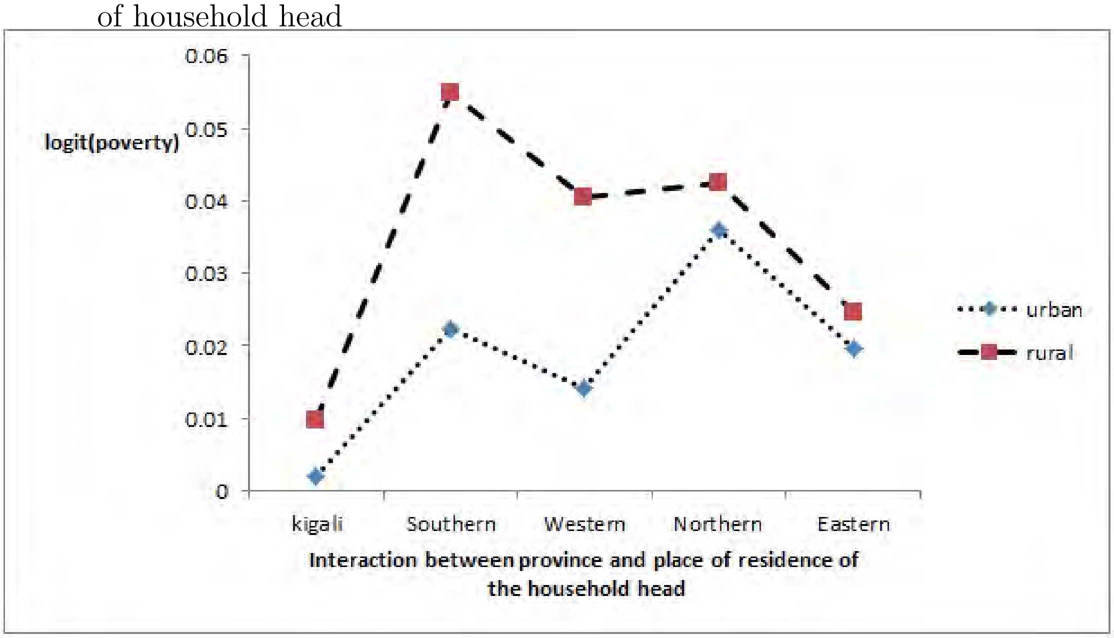


TABLE 3.6. Type 3 Analysis of effects for the survey logistic model

Effects	Waldχ^2	df	P-value
Province/Region	91.9929	4	$< .000$
Place of residence	10.0542	1	0.3045
Gender of the household head	37.9931	1	$< .000$
Highest level of education of household head	300.7408	3	$< .000$
Age of the household head	14.8441	1	$< .000$
Size of the household	3.7271	1	.0535
Region/Province*place of residence of household head	10.1946	1	0.0373
Gender*Age of the household head	16.4472	1	$< .000$
Age of the household head*size of the household	33.7280	1	$< .000$

FIGURE 3.4. Interaction effect between size of the household and age of household head

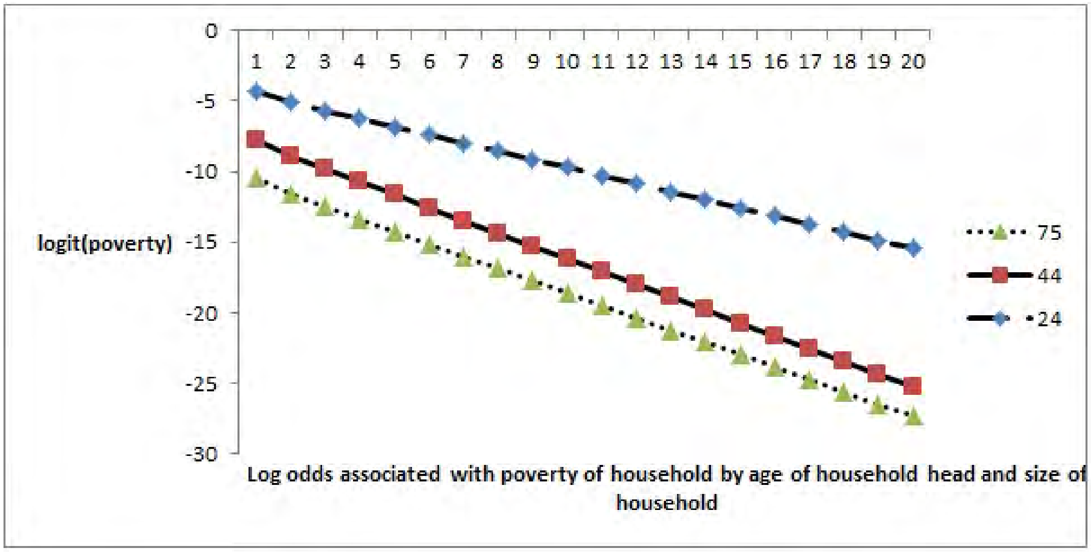
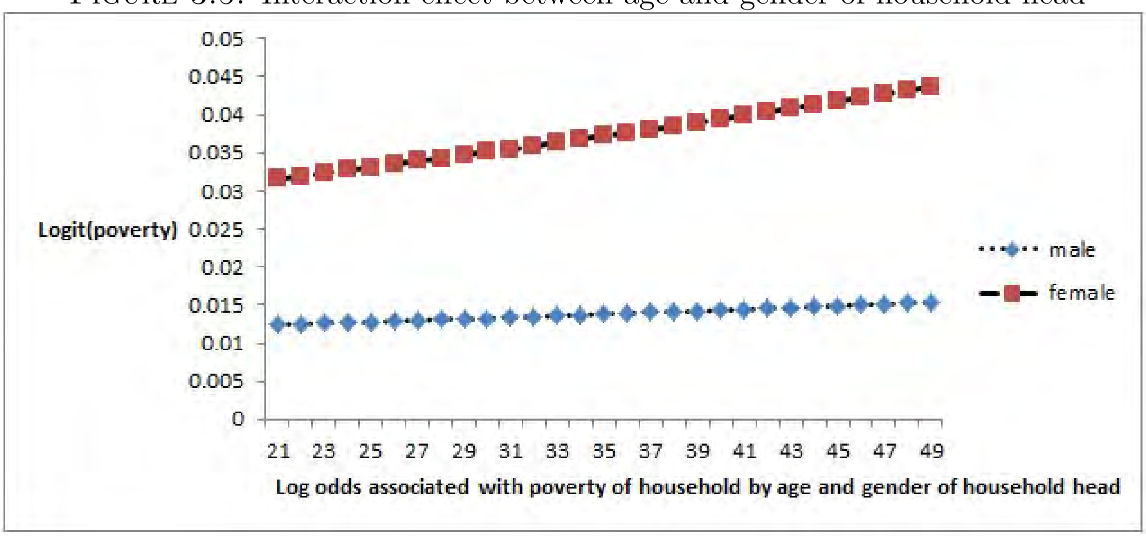


FIGURE 3.5. Interaction effect between age and gender of household head



The concordance index in Table 3.7 suggested that 73.9 % of the probability of poverty of household is predicted correctly which is very good prediction for the survey logistic model. Table 3.8 shows the likelihood ratio, the efficient score test, and the Wald test for testing the significance of the explanatory variable. As a result all these tests are highly significant.

Table 3.7. Model fit statistics

Criterion	Intercept only	Intercept and covariates
AIC	16801.469	14559.562
SC	16808.898	14693.284
-2log L	16799.469	14523.562
C		0.739

Table 3.8. Testing global null hypothesis

Test	Wald χ^2	DF	p-value
Likelihood Ratio	2275.9068	17	$< .0001$
Score	1873.1809	17	$< .0001$
Wald	94.7871	17	$< .0001$

The results from Table 3.10 show that the estimates are the same when Taylor and Jackknife estimation for variance are used. However, the standard deviation are higher when the Jackknife is used, as a result the p-values also for some covariates are significant in Taylor case and not significant for Jackknife method. This means that Taylor method may underestimate the variance. For this reason we used Jackknife approximation technique to estimate the variance in the following analysis.

Comparison between the results from classical and survey logistic regression

Table 3.11 presents the comparison of results from classical binary logistic regression and survey logistic regression. In general the results from these two methods are different due to sampling stratification and sampling weights. The standard deviation from the classical binary logistic regression model are small compared to the standard deviation produced by survey logistic regression. This means that the classical logistic regression model tends to underestimate the variance, as consequence some explanatory variables may be statistically significant when the classical logistic regression model is fitted to the data but no significant when the survey logistic regression is used. This is the case where the variables such as household from urban Western and household from Northern province are not statistical significant in the case of survey

Table 3.9. Parameter estimate from binary survey logistic regression for poverty of household

Indicator	Estimate	S.E	P-Value	OR
Province and place of residence				
Eastern	reference			
Kigali	-0.9473	.3400	.0053	.388
South	0.8295	.0.0978	< .0001	2.292
West	0.5137	.1124	< .0001	1.671
North	.0.5635	.1068	< .0001	1.757
Gender of household head				
Female Male	reference -0.8583	 .1393	 < .0001	 .424
Education of Household head				
Higher	reference			
Secondary	1.9255	.6069	.0015	6.859
Primary	3.2258	.6082	< .0001	25.175
No education	3.7498	.6061	< .0001	42.512
Age of the household head	.0121	.00313	.0001	1.012
Size of household	.0698	.0361	.0535	1.072
Place of residence				
Rural	reference			
Urban Province and place of residence	-.2307	.2247	.3045	.794
Eastern and rural	reference			
Kigali and urban	-1.3403	.5356	.0123	.262
South and urban	-.6979	.3290	.0339	.498
West and urban	-0.8477	.5548	.1265	.428
North and urban	.0591	.5151	.9086	1.061
Gender and age of the household head				
Female Male and age of the household	reference .0119	 .00293	 < .0001	 1.012
Size of household and age of household Size and age of the household head	 -.00445	 .000767	 < .0001	 .996
Intercept	-3.6771	.6382	< .0001	

Table 3.10. Comparison between the results from binary survey logistic regression with Taylor and Jackknife variance estimation for poverty of household

	Taylor			Jackknife		
Indicator	Estimate	S.E	P-Value	Estimate	S.E.	P-value
Intercept	-3.6771	.6382	$< .0001$	-3.6771	.7566	$< .0001$
Province and place of residence(Eastern=ref)						
Kigali	-0.9473	.3400	.0053	-0.9473	.3950	.0165
South	0.8295	.0.0978	$< .0001$	.8295	.0979	$< .0001$
West	0.5137	.1124	$< .0001$	.5137	.1126	$< .0001$
North	.0.5635	.1068	$< .0001$	.5635	.1069	$< .0001$
Gender of household head(Female=ref) Male	 -0.8583	 .1393	 $< .0001$	 -0.8583	 0.1399	 $< .0001$
Education of Household head(higher=ref)						
Secondary	1.9255	.6069	.0015	1.255	0.7273	.0081
Primary	3.2258	.6082	$< .0001$	3.2258	0.7295	$< .0001$
No education	3.7498	.6061	$< .0001$	3.7498	0.7272	$< .0001$
Age of the household head	.0121	.00313	.0001	0.121	0.00315	0.0001
Size of household	.0698	.0361	.0535	0.0698	0.0364	0.0549
Place of residence(rural=ref)						
Urban	-.2307	.2247	.3045	-0.2307	0.2747	0.4009
Province and place of residence (Eastern and rural=ref)						
Kigali and urban	-1.3403	.5356	.0123	-1.3403	0.6185	0.0302
South and urban	-0.6979	0.3290	0.0339	-0.6979	0.3743	0.00623
West and urban	-0.8477	0.5548	0.1265	-0.8477	0.7165	0.2368
North and urban	0.0591	0.5151	0.9086	0.0591	0.6653	0.9292
Gender and age of the household head(female=ref)						
Male and age of the household	0.0119	0.00293	$< .0001$	0.0119	0.00294	$< .0001$
Size and age of household head	no reference					
Size and age of the household head	-0.00445	0.000767	$< .0001$	-0.00445	.000771	$< .0001$

logistic model (p-value=.7730 and p-value= 1123 respectively) whereas in logistic regression without sampling design only the household from urban Northern was not statistically significant (p-value=.716).

TABLE 3.11. Comparison between the results from binary logistic regression and binary survey logistic regression for poverty of household

	Survey Logistic			classical logistic		
Indicator	Estimate	S.E	P-Value	Estimate	S.E.	P-value
Province and place of residence						
Eastern	reference					
Kigali	-0.9473	.3400	.0053	-0.991	.14	< .0001
South	0.8295	.0.0978	< .0001	.8497	.0575	< .0001
West	0.5137	.1124	< .0001	.5415	.0584	< .0001
North	.0.5635	.1068	< .0001	.5796	.0636	< .0001
Gender of household head						
Female	reference					
Male	-0.8583	.1393	< .0001	.-.8678	.1356	< .0001
Education of Household head						
Higher	reference					
Secondary	1.9255	.6069	.0015	1.8689	.5945	.0017
Primary	3.2258	.6082	< .0001	3.7370	.5880	< .0001
No education	3.7498	.6061	< .0001	3.7370	.5880	< .0001
Age of the household head	.0121	.00313	.0001	.0122	.00324	.0002
Size of household	.0698	.0361	.0535	.0777	.0348	.0257
Place of residence						
Rural	reference					
Urban	-.2307	.2247	.3045	-.2323	.2156	0.2811
Province and place of residence						
Eastern and rural	reference					
Kigali and urban	-1.3403	.5356	.0123	-1.2489	0.3284	.0001
South and urban	-.6979	.3290	.0339	-.6758	.2470	.0062
West and urban	-0.8477	.5548	.1265	-.7730	0.3115	.0131
North and urban	.0591	.5151	.9086	.1123	.3092	.7164
Gender and age of the household head						
Female	reference					
Male and age of the household	.0119	.00293	< .0001	.0122	.00282	< .0001
Size of household and age of household head						
Size and age of the household head	-.00445	.000767	< .0001	-.00461	.000727	< .0001
Intercept	-3.6771	.6382	< .0001	-3.6762	.6079	< .0001

3.3.4. A proportional odds model with sampling design to identify the determinants of malnutrition of children under five years in Rwanda.

Introduction

The main objective of this subsection is to extend classical ordinal logistic regression to ordinal logistic regression with complex sampling weights to identify the key determinants of underweight among children under five years in Rwanda. We have considered the anthropometric indicator for underweight (weight-for-age). Stunting and wasting can be modeled in a similar way. The children's nutrition status can be categorized as nourished (z-score ≥ -2.0), moderately undernourished ($-3.0 \leq$ z-score $< -2,0$) and severely undernourished (z-score < -3.0) which made response variable to be ordinal from a continuous variable data. Therefore, nutrition status in this research is an ordinal response variable obtained from grouped continuous variables. Therefore, it is convenient to use ordinal logistic regression models (McCullagh, 1980; Ananth and Kleinbaum, 1997; Hosmer et al., 2000; Agresti, 2002; Collet, 2003; Agresti, 2007; Das and Rahman, 2011; Habyarimana et al., 2014). However, it can also be categorized as malnourished (z-score < -2.0) and nourished (z-score ≥ -2.0) and in this case the binary (survey) logistic regression model can be used.

Data analysis

The data analysis was firstly done using SAS 9.3 with PROC LOGISTIC procedure in case of proportional odds model without sampling design. We have also used Brant test command of Stata Spost package to find the single score test for each explanatory variable.

TABLE 3.12. Model fit statistic

Criterion	Intercept only	Intercept and covariates
AIC	2490.263	2359.929
Sc	2502.185	2467.221
-2LOGL	2486.263	232.929

TABLE 3.13. Testing global null hypothesis:$\beta = 0$

Test	χ^2	DF	Pr>ChSq
Likelihood Ratio	162.3344	16	$< .0001$
Score	158.6541	16	$< .0001$
Wald	137.9101	16	$< .0001$

However due to the nature of sampling technique used in Demographic Health Survey, we have extended proportional odds model without sampling design to proportional odds model with sampling design to account for complexity of sampling design (Liu

Table 3.14. Type 3 analysis of effects for POM without sampling weight for malnutrition of children under five years

Effect	DF	Wald Chi-Square	Pr>ChSq
Birth order	3	29.3769	$< .0001$
Mother's education	2	21.3041	$< .0001$
Gender of the child	1	11.3570	0.0008
Knowledge on nutrition	1	4.6247	0.0315
Birth weights	1	22.7213	$< .0001$
Multiple birth	2	13.1069	0.0014
Anemia	1	7.9663	0.0048
Marital status	3	9.9613	0.0189
BMI	1	18.5003	$< .0001$
Had fever	1	7.1224	0.0076

and Koirala, 2013) where the variance was estimated by replicated sampling methods (Jackknife). Finally SAS 9.3 with PROC SURVEYLOGISTIC procedure was used to fit ordinal logistic regression with sampling design. The final model in both models is the same and is given by

$$\begin{aligned} logitP(Y \leq j|X) = \gamma_j &+ \beta_1 BMI + \beta_2\text{Birth order} + \beta_3\text{Gender of the child} \\ &+ \beta_4\text{Birth weight} + \beta_5\text{Fever} + \beta_6\text{Multiple births} \\ &+ \beta_7\text{Mother's education level} + \beta_8\text{Mother's marital status} \\ &+ \beta_9\text{Anemia} + \beta_{10}\text{Knowledge on nutrition}, j = 1, 2, 3 \end{aligned}$$

Results and interpretations from POM with sampling weights

The score test of proportional odds assumption is found not significant at 5% level of significance(p-value=0.6421) see Table 3.19; this means that the proportional odds assumption is satisfied. The single score test for each explanatory variable is also not significant at 5% level of significance which also confirmed the validity of proportional odds model Table 3.15. The results in Table 3.19 revealed that the children born at 2-3, 4-5 and 6^+ birth order were found 2.183 ($p < .0001$), 2.235(p=0.0002) and 3.062

Table 3.15. Parameter estimate from POM without sampling design for underweight

Indicator	Estimate	S.E	P-Value	OR	Single p-value
Intercept1	-5.554	0.5391	$< .0001$		
Intercept2	-3.6564	0.5258	$< .0001$		
Birth order(first=ref)					0.133
2-3	0.7505	0.1789	$< .0001$	2.118(1.492,4.196)	
4-5	0.7300	0.1980	0.0002	2.075(1.408,3.059)	
6^+	1.0492	0.1964	$< .0001$	2.855(1.943,4.196)	
Mother's education(secondary or higher=ref)					0.413
Primary	1.9191	0.4535	$< .0001$	6.815	
No education	2.1339	0.4650	$< .0001$	8.448	
Gender of the child(male=ref)					0.996
Female	-0.4013	0.1191	0.0008	0.669	
Knowledge on nutrition(No=ref					0.4171
Yes	-0.2768	0.1287	0.0315	0.758	
Birth weights($\geq 2500g$=ref)					0.837
$< 2500g$	1.1848	0.2467	$< .0001$	3.270(2.016,5.304)	
Multiple birth(singleton=ref)					0.539
First multiple	1.3445	0.4198	0.0014	3.836	
Second multiple and more	0.7221	0.3980	0.0696	2.059	
Anemia(No=ref)					0.492
Anemic	0.3327	0.1179	0.0048	1.395	
Marital status(divorced/separated=ref)					0.757
Never in union	-0.2268	0.3258	0.4864	0.797	
Married/ partner	-0.6268	0.2186	0.0041	0.53	
Widowed	-0.4261	0.4132	0.3024	0.653	
BMI(≥ 18.5=ref)					0.538
BMI< 18.5	0.9272	0.2158	$< .0001$	2.527	
Had fever(No=ref)					0.282
Yes	-0.3842	0.1440	0.0076	0.681	
Score test for proportional odds assumption	$\chi^2 =$	14.868	Df=16	p-value=0.5343	
Goodness of fit(likelihood ratio)	$\chi^2 =$	162.334	Df=16	p-value$< .0001$	

($p < .0001$) times more likely to be in worse nutrition status respectively as compared to children born at first order. The risk of having worse nutrition status were 12.247 ($p < .0001$) and 10.555 ($p < .0001$) times higher for children born to mother without education and mother with primary education respectively as compared to children born to mother with secondary or higher education see Table 3.19. In the same Table, it was found that the female children were 0.687($p < .0001$) times less likely to be in worse nutrition status as compared to male children. The risk of having worse underweight status were 3.192 (p-value=0.0033) times higher among children born with lower weight ($< 2500g$) as compared to children born with higher weight $\geq 2500g$ Table 3.19.

A child born at first multiple (twin) is 3.574 (p=0.0020) times more likely to be in worse nutrition status than a singleton child at birth and the effect of second multiple birth was not significant (p=0.1302). The incident of anemia significantly affects the nutrition status of the child. The risk of having worse underweight was 1.403 (p-value=0.0045) times higher among the children born to anemic mother than children born to non-anemic mother. A child born to married mother or mother living with a partner was 0.577 (p=0.0166) times less likely to be in worse nutrition status as compared to child born to divorced or separated mother; however, the effect of child born to widower or mother who has never been in union was not significant as compared to child born to divorced or separated mother. A child born to thin mother ($BMI < 18.5$) was 2.601 (p-value=0.0002) times more likely to be in worse nutrition status as compared to a child born to normal or obese mother ($BMI \leq 18.5$). Children who did not have a fever during the two weeks before the survey were 0.705 (p=0.0283) times less to be in worse nutrition status than a child who was reported to have had fever in two weeks prior to the survey.

TABLE 3.16. Model fit statistic POM with sampling weights for underweight

Criterion	Intercept only	Intercept and covariates
AIC	2529.518	2394.885
Sc	2541.439	2502.177
-2LOGL	2525.518	2358.885

TABLE 3.17. Testing global null hypothesis:$\beta = 0$ for POM with sampling weights for underweight

Test	Chi-Square	DF	Pr>χ^2
Likelihood Ratio	166.6333	16	$< .0001$
Score	159.0842	16	$< .0001$
Wald	126.3620	16	$< .0001$

Table 3.18. Type 3 analysis of effects: POM with sampling weights for underweight

Effect	DF	Wald Chi-Square	Pr> χ^2
Birth order	3	31.8592	$< .0001$
Mother's education	2	19.8760	$< .0001$
Gender of the child	1	8.6546	0.0033
Knowledge on nutrition	1	3.7666	0.0523
Birth weights	1	20.5149	$< .0001$
Multiple birth	2	10.5369	0.0052
Anemia	1	8.0582	0.0045
Marital status	3	8.8754	0.0310
BMI	1	13.9095	0.0002
Had fever	1	4.8104	0.0283

Table 3.19. Comparison of the POM with and without complex survey design for underweight

		POM unweighted			POM weighted		
Indicator	Estimate	SE	P-Value	Estimate	OR	SE	P-VALUE
Intercept1	-5.554	0.5391	$< .0001$	-6.1520		0.6432	$< .0001$
Intercept2	-3.6564	0.5258	$< .0001$		-4.2422	0.6318	$< .0001$
Birth order(first=ref)							
2-3	0.7505	0.1788	$< .0001$	0.7807	2.183	0.1711	$< .0001$
4-5	0.7320	0.1980	0.0002	0.8043	2.235	0.2190	0.0002
6^+	1.0510	0.1964	$< .0001$	1.1190	3.062	0.2050	$< .0001$
Mother's education(secondary or higher=ref)							
Primary	1.9191	0.4535	$< .0001$	2.3566	10.556	0.5649	$< .0001$
No education	2.1339	0.4650	$< .0001$	2.5053	12.247	0.5637	$< .0001$
Gender of the child(male=ref							
Female	-0.4013	0.1191	0.0008	-0.3753	0.687	0.1276	0.0033
Knowledge on nutrition(No=ref)							
Yes	-0.2768	0.1287	0.0315	-0.2806	0.765	0.1380	0.0523
Birth weights($\geq 2500g$=ref)							
$< 2500g$	1.1736	0.2462	$< .0001$	1.1607	3.192	0.2563	$< .0001$
Multiple birth(singleton=ref)							
First multiple	1.3445	0.4198	0.0014	1.2737	3.574	0.4129	0.0020
Second multiple and more	0.7221	0.3980	0.0696	0.6080	1.837	0.4018	0.1302
Anemia(No=ref)							
Anemic	0.3327	0.1179	0.0048	0.3389	1.403	0.1194	0.0045
Marital status(Divorced/separated=ref)							
Never in union	-0.2268	0.3258	0.4864	-0.0568	0.945	0.3435	0.8687
Married/ partner	-0.6268	0.2186	0.0041	-0.5494	0.577	0.2293	0.0166
Widowed	-0.4261	0.4132	0.3024	-0.2960	0.744	0.5003	0.5541
BMI(< 18.5=ref)							
BMI< 18.5	0.9272	0.2156	$< .0001$	0.9559	2.601	0.2563	0.0002
Had fever(Yes=ref)							
No	-0.3842	0.1440	0.0076	-0.3497	0.705	0.1595	0.0283
Score test for proportional odds assumption	$\chi^2 = 14.8680$	Df=16	p-value=0.5343	$\chi^2 = 13.4160$	Df=16		p-value=0.6421
Goodness of fit(likelihood ratio)	$\chi^2 = 162.334$	Df=16	p-value$< .0001$	$\chi^2 = 166.633$	DF=16		p-value$< .0001$

3.4. Summary

In this chapter we have considered classical binary and ordinal logistic regression models as well as binary and ordinal survey logistic regression models to fit the households (poverty case) and women data (in malnutrition case).

In poverty case, we used classical binary and survey binary logistic regression model to identify the factors associated to the poverty of households. Taylor linearization method and Jackknife method were used to estimate the variance and the results were compared. It was found that the standard errors from Taylor linearization is smaller than the standard error produced by Jackknife. Taylor linearization method tends to underestimate the variance. For this reason we have considered the results found under Jackknife variance estimation method. Further, the results from this study revealed that the demographic and spatial profile of poor households are: education of household head, gender of household head, age of household head, place of residence (urban or rural), region(province), and size of household. In addition, this study found that the majority of poor households have low standards of education. This suggests that there is a need to improve existing access to higher education. We have also considered the two way interaction effects between place of residence (urban or rural) and province, gender of household head and age of household head, size of the household and age of household head. It was then found that a household from rural is more likely to be poor compared to a household from urban in all provinces. This supports the existing policy of grouped settlement where people are advised to build their house in a township known as *Imidugudu*. But this also suggests a special policy for targeting poverty reduction in rural households. The rural household from Southern province was found to be more likely poor compared to other households from other provinces; this suggests provincial targeting in poverty reduction. The findings from comparison of the results from classical binary logistic regression and binary survey logistic regression discouraged the use of binary logistic regression without sampling weights. Therefore, when using DHS data it is advised to account for complexity of sampling design.

The malnutrition indicator considered in this chapter was underweight. The analysis based on stunting and wasting can be done in a similar way. The anthropometric indicator for underweight (weight-for-age) was considered in this study and categorized as severely malnourished, moderately malnourished and nourished and this made the response variable to be ordinal. The proportional odds model with sampling weights and without sampling weights were used and their results were compared. This study revealed that the determinants of malnutrition (underweight case) of children under five years in Rwanda are birth order, mother's education, gender of child, birth weight, multiple birth, body mass index, anemia, marital status and whether the child had or had not fever in two weeks before the survey. The findings from the comparison of the results from proportional odds model with sampling weights and without sampling weights revealed as in binary case that it is better to include sampling weights when the data was collected using multistage sampling in order to make statistically valid inferences from the finite population.

The primary sampling units might have variability between them. In order to account for this variability, in the next chapter we use generalized linear mixed model(GLMM).

CHAPTER 4

Generalized Linear Mixed Model

The ordinal and binary survey logistic regression models discussed in chapter 3 assume all the variables effect as fixed effect. They do not have options to include the random effects. However these are situations in which the effect of the variable is random. For instance in the poverty determinants study using the DHS data, the primary sampling units (clusters) are considered to be a random effect. Therefore this chapter uses generalized linear mixed model (GLMM) which offers an option to include random effect.

4.1. Model formulation

Let Y_{ij} be the outcome measured for cluster $i, i = 1, 2, ..., N, j = 1, 2, ..., n_i$ and Y_i be the n_i-dimensional vector of all measurements available for cluster i, conditionally on random effects b_i, it is assumed that the elements of y_{ij} for y_i are independent and y_{ij} has the following density

$$\mathbf{f_i}(\mathbf{y_{ij}}|\mathbf{b_i}, \beta, \phi) = \mathbf{exp}\left[\frac{\mathbf{y_{ij}}(\theta_{\mathbf{ij}}) - \psi(\theta_{\mathbf{ij}})}{\phi} + \mathbf{c}(\mathbf{y_{ij}}, \phi)\right] \tag{4.1}$$

where μ_{ij}, the conditional mean of y_{ij} is modeled through a linear predictor containing both fixed and random factors given by $g(\mu_{ij}) = g(E[y_{ij}|b]) = x'_{ij}\beta + z'_{ij}b_i$, where g(.) is the link function, x'_i is the i^{th} row matrix for the fixed effects, z'_{ij} is the i^{th} row matrix for the random effects, β is the fixed effect parameter vector, $\mathbf{b}$ is the random effect parameter vector and $\mathbf{b} \sim N(0, D)$, ϕ is a scale parameter and θ is the natural parameter. The marginal mean, variance and co-variances are given in McCulloch (2001) as follows. The marginal mean of y is given by

$$\begin{aligned} E[y_{ij}] &= E[E[y_{ij}|b_i]] \\ &= E[\mu_i] \\ &= E[g^{-1}\left(x'_{ij}\beta + z'_{ij}b_i\right)] \end{aligned} \tag{4.2}$$

In general the equation (4.2) cannot be simplified because of nonlinearity of the function $g^{-1}(.)$. In a linear mixed model, the induced marginal mean is reduced to $E(y_{ij}) = X'_{ij}\beta$. The marginal variance of y_{ij} is given by

$$\begin{aligned} var(y_{ij}) &= var(([y_{ij}|b]) + E\left[var(y_{ij}|b)\right] \\ &= var(\mu_i) + E\left[\tau^2\nu(\mu_{ij})\right] \\ &= var\left(g^{-1}\left[X'_{ij}\beta + Z'_{ij}b_i\right]\right) + E\left[exp\{X'_{iJ}\beta + Z'_{ij}b_i\}\right] \end{aligned} \tag{4.3}$$

where the equation (4.3) cannot be simplified without making precise assumption about the form of $g(.)$ and/or conditional distribution of y_{ij}. The use of random effects introduces a correlation between the observations that have any random effect in common. Assuming conditional independence on y_i, the marginal covariance is given by

$$\begin{aligned} cov(y_{ij}, y_{ik}) &= cov\left(E\left[y_{ij}|b_i\right], E\left[y_{ik}|b_i\right]\right) + E\left[cov\left(y_{ij}, y_{ik}|b_i\right)\right] \\ &= cov\left(g^{-1}\left[X'_{ij}\beta + Z'_{ij}b_i\right], g^{-1}\left[X'_{ik}\beta + Z'_{ik}b_i\right]\right) \end{aligned} \tag{4.4}$$

4.2. Model parameter estimation

The generalized linear mixed model is fitted using either Bayesian or likelihood approaches. In the Bayesian case, there is a need to specify the prior densities and thereafter the posterior distribution can be found (Molenberghs and Verbeke, 2005). The advantage of Bayesian is the flexibility for full assessment of uncertainty in the estimated random effects and functions of models parameters; but it has also a major drawback of intensive computation which require sophisticated computer programs and questions about when the sampling process has achieved convergence (Breslow and Clayton, 1993; Agresti, 2002). The Bayesian approach is not considered in this study.

4.2.1. Maximum likelihood estimation.

The generalized linear mixed model is fitted by maximizing the marginal likelihood, obtained by integrating out the random effects. The likelihood contribution of the i^{th}

subject is given by Molenberghs and Verbeke (2005)

$$f_i = f_i(y_i|\beta, D, \phi) = \int \prod_{j=1}^{n_i} f_{ij}\left(Y_{ij}|b_i, \beta, \phi\right) f\left(b_i|D\right) db_i \tag{4.5}$$

From equation (4.5) the likelihood for β, D and ϕ can be given as

$$L_i = f_i(y_i|\beta, D, \phi) = \int \prod_{j=1}^{n_i} \left(y_{ij}|b_i, \beta, \phi\right) f\left(bi|D\right) db_i \tag{4.6}$$

Therefore, the likelihood function L is written as follows

$$L = \prod_{i=1}^{N} L_i = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}\left(y_{ij}|b_i, \beta, \phi\right) f\left(bi|D\right) db_i \tag{4.7}$$

The key problem when maximizing equation (4.7) is the presence of N integrals over the k-dimensional random effects b_i. In some cases, the equation (4.7) can be worked out analytically, for example linear mixed models for continuous outcomes, probit-normal model (Molenberghs and Verbeke, 2005). However, in general, there are no analytical solutions available for integral (4.7) and therefore numerical approximations are needed to evaluate the integral (4.7). These numerical approximations can be, in general, classified into three approaches as follows: approximation of integrand, approximation of integral and approximation of the data (Molenberghs and Verbeke, 2005).

4.2.1.1. ***Approximation of Integrand.***

The main objective of approximation of integrand is to obtain a tractable integral such that a closed-form can be obtained to make the numerical maximization of the approximated likelihood possible. Many approaches have been proposed but basically all come down to Laplace type approximations of the function to be integrated (Molenberghs and Verbeke, 2005).

Laplace approach

The Laplace approximation is the most convenient approach to approximate integrals (Tierny and Kadane, 1986) of the form

$$I = \int e^{K(b)} db \tag{4.8}$$

where $K(b)$ is a known, unimodal, and bounded function of a k-dimensional variable b. Let us consider $\hat{b}$ to be the value of b for which K is minimized. Therefore, the second order Taylor series expansion of $K(b)$ around $\hat{b}$ can be written as

$$K(b) \approx K(\hat{b}) + \frac{1}{2}(b - \hat{b})' K''(\hat{b})(b - \hat{b}) \tag{4.9}$$

where $K''(b)$ is equal to the Hessian of K, that means the matrix of the second order derivative of K, evaluated at $\hat{b}$. The integral **I** can be approximated by replacing $K(b)$ in equation (4.8) by its value from equation (4.9) and becomes

$$I \approx (2\pi)^{k/2} | - K''(\hat{b})|^{-1/2} e^{K(\hat{b})} \tag{4.10}$$

The integral (4.7) is proportional to an integral of the form (4.10), for functions $K(b)$ and is given by:

$$K(b) = (\phi)^{-1} \Sigma_{j=1}^{n_i} \left[y_{ij} \left(x'_{ij}\beta + z'_{ij}b \right) - \psi \left(x'_{ij}\beta + z'_{ij}b \right) \right] - \frac{1}{2} b' D^{-1} b \tag{4.11}$$

such that Laplace's approximation approach can be used. The Laplace approximation is exact if $K(b)$ is a quadratic function of b, that means if the integrands in (4.8) are exactly to normal kernels. Raudenbush et al. (2000) extended the Laplace method by including higher-order Taylor expansion of equation (4.9) for K up to order six, where in the simulations study they show that this considerably improves the approximation.

4.2.1.2. ***Approximation of integral***.

When the above approximation methods fail, then numerical integration proves to be very useful. Consider Gaussian and adaptive Gaussian quadrature, mainly designed for the approximation of integrals of the form

$$\int f(z)c(z)dz \tag{4.12}$$

for a known function $f(z)$ and $c(z)$ the density of univariate or multivariate standard normal distribution. Thus, the random effects have to be standardized such that they get identity covariance matrix. Let δ_i be equal to $\delta_i = D^{-1/2} b_i$. Then δ_i is normally distributed with mean 0 and covariance I, and then the linear predictor becomes $\theta_{ij} = x'_{ij}\beta + z_{ij} D^{1/2} \delta_i$. As a result, the variance components in D are now contained in the linear predictor. Therefore the likelihood contribution for subject i is given by

$$f_i\left(y_i|\beta, D, \phi\right) = \int \prod_{i=1}^{n_i} f_{ij}\left(y_{ij}|b_i, \beta, \phi\right) f(b_i, D) db_i \tag{4.13}$$

$$= \int \prod_{i=1}^{n_i} f_{ij}\left(y_{ij}|\delta_i, \beta, \phi\right) f(\delta_i, D) d\delta_i \tag{4.14}$$

where the random effects b_i are assumed to be normally distributed with mean 0 and covariance D. The expression (4.14) is of the form (4.12) as required to apply the Gaussian quadrature (Molenberghs and Verbeke, 2005; Antonio and Beirlant, 2007).

Gaussian quadrature

The classical Gaussian quadrature approximates an integral of the form (4.12) by a weighted sum, namely

$$\int f(z)c(z)dz \simeq \sum_{k=1}^{K} \omega_k f(z_k) \tag{4.15}$$

where K is the order of the approximation, the higher K, the more accurate the approximation will be. In addition, z_k are solutions of the K^{th} order Hermite polynomial and ω_k are corresponding weights. In the case of univariate integration, the approximation involves subdividing the integration region into intervals, and approximating rectangles.

Adaptive Gaussian quadrature

In the adaptive Gaussian quadrature approach, the quadrature points are centered and scaled as if $f(z)c(z)$ were a normal distribution. The mean of this normal distribution would be the model $\hat{z}$ of $\ln\left[f(z)c(z)\right]$, and the corresponding variance would be

$$\left[-\frac{\partial^2}{\partial z^2}\ln\left[f(z)c(z)\right]|_{z=\hat{z}}\right]^{-1} \tag{4.16}$$

The new quadrature points are given by

$$z_k^* = \hat{z} + \left[-\frac{\partial^2}{\partial z^2}\ln\left[f(z)c(z)\right]|_{z=\hat{z}}\right]^{-1/2} z_k \tag{4.17}$$

with corresponding weights

$$w_k^* = \left[-\frac{\partial^2}{\partial z^2}\ln\left[f(z)c(z)\right]|_{z=\hat{z}}\right]^{-1/2} \frac{c(z_k^*)}{z(z_k)} w_k \tag{4.18}$$

In this case, the integral is now approximated by

$$\int f(z)c(z)dz \approx \sum_{k=1}^{K} w_k^* f(z_k^*) \tag{4.19}$$

The adaptive Gaussian quadrature needs less quadrature points than classical Gaussian quadrature. On the other hand, adaptive Gaussian quadrature needs calculation of $\hat{z}$ for each unit in the dataset, then for the numerical maximization of N functions of the form (4.12) and makes Gaussian quadrature much more time consuming (Molenberghs and Verbeke, 2005). It has been shown that when (4.19) is applied with only one node, the result is equivalent to approximating the integral using Laplace approximation (Liu and Pierce, 1994). Some simulation results suggest that in the classical Gaussian quadrature, a large number of quadrature points (100 or more) are necessary to obtain high accuracy while the adaptive quadrature provides good accuracy with 20 or fewer quadrature points (Diggle et al., 2002). Nonetheless, the adaptive Gaussian quadrature is much more time consuming than the classical Gaussian quadrature. This is due to the fact that the adaptive Gaussian quadrature requires calculation of $\hat{z}$ for each unit in the dataset, hence the numerical maximization of N functions of the form (4.12) (Molenberghs and Verbeke, 2005). Moreover, since these functions (4.12) depend on unknown parameters β, D and ϕ, the quadrature points as well as the weights used in the adaptive Gaussian quadrature depend on those parameters, and hence need to be updated in every step of the iterative procedure (Molenberghs and Verbeke, 2005). Once the problem of intractable integral is solved, the actual maximization of the likelihood is carried out using algorithms such as Newton-Raphson and Fisher scoring. The numerical integration methods work relatively well with GLMM that have low-dimensional random effects distributions such as single random effect or two or three nested random effects (Diggle et al., 2002). However, none of the numerical methods have been made computationally practical for models with random effects distribution with $k > 5$.

4.2.2. Pseudo-likelihood approach.

Suppose Y represents the $(n \times 1)$ vector of observed data and b is a $(r \times 1)$ vector of random effects

$$E[Y|b] = g^{-1}(X\beta + Zb) = g^{-1}(\eta) = \mu \tag{4.20}$$

where $b \sim N(0, D)$ and $var[Y|b] = A^{1/2}RA^{1/2}$ following (Wolfinger and O'Connell, 1993) a first order Taylor series of μ about $\hat{\beta}$ and $\hat{b}$ yields

$$g^{-1}(\eta) = g^{-1}(\hat{\eta}) + \hat{\Delta}X(\beta - \hat{\beta}) \tag{4.21}$$

where $\hat{\Delta} = \left(\frac{\partial g^{-1}(\eta)}{\partial \eta}\right)_{\hat{\beta},\hat{b}}$ is a diagonal matrix of derivatives of the conditional mean evaluated at the expansion locus. Rearranging terms we get $\hat{\Delta}^{-1}(\mu - g^{-1}(\hat{\eta}) + X\hat{\beta} + Z\hat{b} \doteq X\beta + Zb$, the left-hand side is the expected value, conditional on b, of $\hat{\Delta}^{-1}(Y - g^{-1}(\hat{\eta}) + X\hat{\beta} + Z\hat{b} \equiv P$ and $var[P|b] = \hat{\Delta}^{-1}A^{1/2}RA^{1/2}\hat{\Delta}^{-1}$. You can thus consider the model

$$P = X\beta + Zb + \epsilon \tag{4.22}$$

equation (4.22) is a linear mixed model with pseudo-response P, fixed effects β, random effects b, and $var(\epsilon) = var[P|b]$. Therefore, the marginal variance in the linear mixed pseudo-model is given by

$$V(\theta) = ZDZ' + \hat{\Delta}^{-1}A^{1/2}RA^{1/2}\hat{\Delta}^{-1},$$

where θ is the $(k \times 1)$ parameter vector containing all unknowns in D and R. Based on this linearization model, an objective function can be defined, assuming that the distribution of P is known. The maximum log pseudo-likelihood and restricted log-pseudo-likelihood for P are therefore given by

$$l(\theta, P) = -\frac{1}{2}log|V(\theta) - \frac{1}{2}r'V(\theta)^{-1}r - \frac{f}{2}log(2\pi)$$

$$l_R(\theta, P) = -\frac{1}{2}log|V(\theta) - \frac{1}{2}r'V(\theta)^{-1}r - \frac{1}{2}log|X'V(\theta)^{-1}X - \frac{f-q}{2}log(2\pi)$$

where $r = P - X(X'V^{-1}X)^{-1}X'V^{-1}P$, q denotes the rank of X and f denotes the sum of frequencies used in the analysis. At convergence, the fixed and random effects parameters are predicted as follows

$$\hat{\beta} = \left(X'V(\hat{\theta})^{-1}X\right)^{-1}X'V(\theta)^{-1}P\hat{b}$$

$$\hat{b} = \hat{D}Z'V(\hat{\theta})^{-1}\hat{r}$$

Therefore, these parameters estimates are used to update the linearization that results in a new linear mixed model. Then the process continues until the relative change between parameter estimates at two successive iterations is sufficiently small (SAS, 2005). In general, there are two widely used approximations based on Taylor's expansion of the mean. A subject specific expansion, referred to as the penalized quasi-likelihood approximation, uses $\widetilde{\beta} = \hat{\beta}$ and $\widetilde{b} = \hat{b}$, that are the current estimates of fixed effects and predictors of random effects. The population average expansion referred to as the marginal quasi-likelihood (MQL) uses $\widetilde{\beta} = \hat{\beta}$ and $\widetilde{b} = 0$, which are the same as current estimates of fixed effects and the random effects are not incorporated in the linear predictor.

4.2.2.1. ***Approximation of the data***.

This approach is based on a decomposition of the data into the mean and an appropriate error term, with Taylor series expansion of the mean that is a non-linear function of the linear predictor. In generalized linear mixed models the penalized quasi-likelihood (PQL) estimate is obtained from the optimization of the quasi-likelihood function that only includes first- and second-order conditional moments augmented with a penalty term on the random effects (Molenberghs and Verbeke, 2005; Breslow and Clayton, 1993). There are many versions of PQL developed by different authors such as Schall (1991), Breslow and Clayton (1993) and Wolfinger et al. (1994). In this study we will discuss Breslow and Clayton (1993); Wolfinger et al. (1994) and Schall (1991). The penalized quasi-likelihood of Breslow and Clayton (1993) is, in general, similar to GLM the setting and its general form is given by

$$QL(y, \beta, b) = QL(y|b)L(b) \tag{4.23}$$

and a quasi-likelihood function based on $y_i, i = 1, 2, ..., m$ can be written (Jiang, 2007) as

$$L_Q \propto (2\pi)^{-k/2}|D|^{-1/2} \int \exp\left(-\frac{1}{2\phi}\Sigma_{i=1}^{M} d_i(y_i, \mu_i) - \frac{1}{2}b'D^{-1}b\right) db \tag{4.24}$$

where the subscript Q indicates quasi-likelihood, and

$$d_i(y_i, \mu_i) = -2\int_{y_i}^{\mu_i} \frac{y_i - u}{a_i(\phi)v(u)} du \tag{4.25}$$

and equation (4.25) is called (quasi-) deviance. In the case where y is Gaussian and g^{-1} is the identity the equation (4.25) can be solved in closed form otherwise numerical methods are needed. Equation (4.24) is Laplace's form of integral, therefore Laplace approximation can be used. Using the results from (4.10)

$$\iota_Q \approx c - \frac{1}{2}\log|D| - \frac{1}{2}\log|k''(\hat{b})| - k(\hat{b}) \tag{4.26}$$

with $k(b) = \frac{1}{2}\left(\Sigma_{i=1}^{m} d_i(y_i, \mu_i) + b'D^{-1}b\right)$ and $\hat{b}$ minimizes $k(b)$. Typically, $\hat{b} = \hat{b}(\beta, \theta)$ is the solution to the following first derivative of $k(b)$

$$D^{-1}b - \Sigma_{i=1}^{n} \frac{y_i - \mu_i}{a_i(\phi)v(\mu_i)g'(\mu_i)} z_i = 0 \tag{4.27}$$

and second derivative of $k(b)$ is given by

$$k''(b) = D^{-1} + \frac{z_i z_i'}{a_i(\phi)v(\mu_i)[g'(\mu_i)]^2} + r \tag{4.28}$$

where the remainder r has expectation 0. If the denominator of equation (4.28) is w_i^{-1}, and the term r is ignored then equation (4.28) reduces to

$$k''(b) \approx D^{-1} + Z'WZ \tag{4.29}$$

where Z is the matrix whose i^{th} row is z_i', and $W = \text{diag}(w_1, w_2, ..., w_m)$, where the quantity w_i is well known in GLM as iterated weights (McCullagh and Nelder, 1989). Thus the log-quasi-likelihood is written (Jiang, 2007) as

$$\iota_Q \approx c - \frac{1}{2}|D| - \frac{1}{2}\log|I + Z'WZD| - \frac{1}{2\phi}\Sigma_{i=1}^{m} d_i(y_i, \hat{\mu}_i) - \frac{1}{2}\hat{b}'D^{-1}\hat{b} \tag{4.30}$$

where $\hat{b}$ is chosen to maximize the sum of the last two terms and I is the identity matrix (Breslow and Clayton, 1993).

Thus $(\hat{\beta}, \hat{b}) = \left(\hat{\alpha}(\theta), \hat{b}(\theta)\right)$, with $\hat{b}(\theta) = b\left(\hat{\alpha}(\theta)\right))$, jointly maximize the Green (1987) PQL

$$-\frac{1}{2\phi}\Sigma_{i=1}^{m} d_i(y_i, \hat{\mu}_i) - \frac{1}{2}\hat{b}'D^{-1}\hat{b} \tag{4.31}$$

Differentiating the expression (4.31) with respect to β and b leads to the following score equations for the mean parameters:

$$\Sigma_{i=1}^{m} \frac{(y_i - \hat{\mu}_i)x_i}{\phi a_i v(\hat{\mu}_i) i g'(\hat{\mu}_i)} = 0 \tag{4.32}$$

and

$$\Sigma_{i=1}^{m}\frac{(y_i-\hat{\mu}_i)z_i}{\phi a_i v(\hat{\mu}_i)ig'(\hat{\mu}_i)}=D^{-1}b \tag{4.33}$$

Green (1987) developed the Fisher scoring algorithm for solving equations (4.32) and (4.33) as iterated weighted least squares but later on Breslow and Clayton (1993) modified it to include the close correspondence with the normal theory calculations of Harville (1977). Defining the working vector Y to have the components $\widetilde{y}_i = \hat{\eta}_i + (y_i - \hat{\mu}_i)g'(\hat{\mu}_i)$, then solution to equations (4.32) and (4.33) based on Fisher scoring can be given as iterative solution to the following system:

$$\begin{pmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ+D^{-1} \end{pmatrix}\begin{pmatrix} \beta \\ b \end{pmatrix}=\begin{pmatrix} X'W\widetilde{y} \\ Z'W\widetilde{y} \end{pmatrix}, \tag{4.34}$$

Harville (1977) derived the expression (4.34) for the best linear unbiased estimator (BLUE) of β and b in the associated normal theory model $\widetilde{y} = X\beta + Zb + \epsilon$, with $\epsilon \sim N(0, W^{-1})$ and $b \sim N(0, D)$, ϵ and b are independent. Equivalently, one may first solve for β in

$$\left(X'V^{-1}X\right)\beta = X'V^{-1}\widetilde{y},$$

where $V = W^{-1} + ZDZ'$ and therefore set

$$\beta = \left(X'V^{-1}X\right)^{-1}X'V^{-1}\widetilde{y}$$

$$\hat{b} = DZ'V^{-1}(Y - X\hat{\beta})$$

and this suggests that one takes as an approximate covariance matrix for $\hat{\beta}$ the matrix $(X'V^{-1}X)^{-1}$.

Schall (1991) developed the other version of the PQL algorithm but it is based on the longitudinal setting. This method is mainly based on the decomposition of the data into the conditional mean and appropriate error term with Taylor series expansion of the mean which is a non linear function of the linear predictor (Molenberghs and Verbeke, 2005). More specifically, one considers the following decomposition

$$Y_{ij} = \mu_{ij} + \epsilon_{ij} = h\left(x'_{ij}\beta + z'_{ij}b_i\right) + \epsilon_{ij} \tag{4.35}$$

in which $h(.)$ equals the inverse link function, and where the error terms have the appropriate distribution with the variance equals to $\text{var}(Y_{ij}|b_i) = \phi v(\mu_{ij})$. If the natural link function is used then

$$\text{var}(\mu_{ij}) = h'\left(x'_{ij}\beta + z'_{ij}b_i\right).$$

Let us consider the Taylor expression of $y_{ij} = \mu_{ij} + \epsilon_{ij} = h(x'_{ij}\beta + z_{ij}b_i) + \epsilon_{ij}$ (McCulloch, 2001; Molenberghs and Verbeke, 2005) around current estimates $\hat{\beta}$ and $\hat{b_i}$ of the fixed and random effects. This yields

$$\begin{aligned} Y_{ij} &\approx h\left(x'_{ij}\hat{\beta} + z'_{ij}\hat{b_i}\right) \\ &+ h'\left(x'_{ij}\hat{\beta} + z'_{ij}\hat{b_i}\right)(\beta - \hat{\beta}) \\ &+ h'\left(x'_{ij}\hat{\beta} + z'_{ij}\hat{b_i}\right)(b_i - \hat{b_i}) + \epsilon_{ij} \\ &= \hat{\mu_{ij}} + v(\hat{\mu_{ij}})x'_{ij}(\beta - \hat{\beta}) \\ &+ v(\hat{\mu_{ij}})z'_{ij}(b_i - \hat{b_i}) + \epsilon_{ij} \end{aligned} \tag{4.36}$$

where $\hat{\mu_{ij}}$ is the current predictor $h\left(x'_{ij}\hat{\beta} + z'_{ij}\hat{b_i}\right)$ for the conditional mean $E(Y_{ij}|b_i)$. In vector notation, the equation (4.36) reduces to

$$Y_i \approx \hat{\mu}_i + \hat{V}X_i(\beta - \hat{\beta}) + \hat{V}z_i(b_i - \hat{b_i}) + \epsilon_i \tag{4.37}$$

with suitable design matrix X_i and Z_i, and with $\hat{V_i}$ equal to the diagonal entries of $v(\hat{\mu_{ij}})$. Rearranging the terms the expression (4.37) becomes

$$\mathbf{Y_i^*} \equiv \hat{\mathbf{V_i^{-1}}}(\mathbf{Y_i} - \hat{\mu_{\mathbf{ij}}}) + \mathbf{X_i}\hat{\beta} + \mathbf{Z_i}\hat{\mathbf{b_i}} \approx \mathbf{X_i}\beta + \mathbf{Z_i}\mathbf{b_i} + \epsilon_{\mathbf{i}}^* \tag{4.38}$$

where $\epsilon_i^* = \hat{V}_i^{-1}\epsilon_i$, and has zero mean. Equation (4.38) can be viewed as a linear mixed model for the pseudo data Y_i^* with fixed and random effects as β and b_i and error terms ϵ_i^*. This, therefore, produces an algorithm for fitting generalized linear mixed model for these pseudo-data. Given starting values for the parameters β, D and ϕ in the marginal likelihood, then empirical Bayes estimates are calculated for b_i, and pseudo data Y_i^* are computed. Therefore, the approximate linear model equation (4.38) is fitted, yielding updated estimates for β, D and ϕ. These are then used to update the pseudo data and this whole scheme is iterated until convergence is reached.

Marginal quasi-likelihood(MQL)

The alternative approximation is very similar to the PQL method, but is based on a linear Taylor expression of the mean μ_{ij} given in (Molenberghs and Verbeke, 2005) as

$$Y_{ij} = \mu_{ij} + \epsilon_{ij} = h\left(x'_{ij}\beta + Z'_{ij}b_i\right) + \epsilon_{ij} \tag{4.39}$$

around the current estimates $\hat{\beta}$ for the fixed effects and around $b_i = 0$ for the random effects; where $h(.)$ is the inverse of link function. This produces a similar result as above and the current predictor of the mean $\hat{\mu_{ij}}$ is now of the form $h(x'_{ij}\hat{\beta})$ instead of $h(x'_{ij}\hat{\beta} + z'_{ij}\hat{b_i})$. Therefore the pseudo-data is now of the form

$$Y_i^* \equiv \hat{V_i^{-1}}(Y_i - \hat{\mu_{ij}}) + X_i\hat{\beta}$$

and satisfies the approximate linear mixed model

$$Y_i^* \approx X_i\beta_i + Z_ib_i + \epsilon_i^* \tag{4.40}$$

The model is fitted by iterating between the calculation of the pseudo-data and the fitting of the approximate linear mixed model. The resulting estimates are known as marginal quasi-likelihood estimates (MQL)(Breslow and Clayton, 1993; Molenberghs and Verbeke, 2005). As the PQL estimates are obtained by optimizing a quasi-likelihood function that only involves the first and second conditional moments, in MQL the estimates are evaluated in the marginal linear predictor $x'_{ij}\hat{\beta}$ instead of the conditional linear predictor $x'_{ij}\hat{\beta} + z'_{ij}\hat{b_i}$ see (Breslow and Clayton, 1993) and even though MQL and PQL are similar in underlying key ideas, they also have some differences. MQL completely ignores the random effects variability in the linearization of the mean; as a result it provides the reasonable approximation when the variance of the random effects is very small, even when the number of measurements per cluster is increased. In contrast, PQL is consistent when both number of subjects as well as the number of measurements per subject approach infinity, even for binary outcomes (Molenberghs and Verbeke, 2005). Littel et al. (2006) argue that Breslow and Clayton (1993) and Wolfinger and O'Connell (1993) approaches are similar in that they both use generalized mixed model equations (4.38) for solutions of β and b_i. The main difference between Wolfinger and O'Connell (1993) and Breslow and Clayton (1993) approaches is that Breslow and Clayton motivate their procedure from a quasi-likelihood viewpoint using approximations based on Laplace's method whereas Wolfinger and

O'Connel's approach pseudo-likelihood(PL) or restricted pseudo-likelihood (REPL) is based on a Gaussian approximation and Taylor's expansion.

In addition, Littel et al. (2006) indicate that the other difference which comes from what Breslow and Clayton (1993) call PQL and what Wolfinger and O'Connell (1993) term PL/REPL lies in the estimation of the scale parameter ϕ, where $\phi = 1$ in the case of Breslow and Clayton (1993) method and ϕ has to be estimated if Wolfinger and O'Connell (1993) approach is considered.

4.3. Inference

Since fitting of GLMM is mainly based on maximum likelihood principles, therefore inferences for the parameters are obtained from standard maximum likelihood theory (Molenberghs and Verbeke, 2005). If the fitted model is appropriate, then the obtained estimators are asymptotically normally distributed with the correct values as means, and with the inverse Fisher information matrix as covariance matrix. Therefore, Wald-type test, comparing standardized estimates to the standard normal distribution can be used. Alternatively Likelihood ratio test and score tests can also be used.

The inference on the fixed effects can be done using Wald-type test (also called Z-test), the approximate t-tests and F-tests (Verbeke and Molenberghs, 2000). The approximate Wald test is obtained from approximating the distribution of $\left(\hat{\beta}_j - \beta_j\right)/s.e(\hat{\beta}_j)$ by a standard univariate normal distribution of each parameter β_j in β, $j = 1, 2, ..., p$. Generally, it may be of interest to construct confidence intervals and tests of hypotheses about certain linear combinations of the component β. For instance, given any unknown matrix L, a test for hypothesis

$$H_0 : L\beta = 0 \text{ and } H_A : L\beta \neq 0 \tag{4.41}$$

follows from the fact that the distribution of

$$\left(\hat{\beta} - \beta\right)' L' \left[L \left(\sum_{i=1}^{N} X_i' V_i^{-1}(\hat{\alpha}) X_i \right)^{-1} L' \right]^{-1} L \left(\hat{\beta} - \beta\right) \tag{4.42}$$

follows asymptotically a chi-square distribution with rank (L) degrees of freedom. Dempster et al. (1981) pointed out that the Wald test is based on estimated standard errors that underestimate the true variability in $\hat{\beta}$; since the variability introduced by the variance parameter is not considered. To circumvent this downward bias of standard errors it is advised to use t-and F-tests approximates for testing hypothesis about β. For each parameter β_j in β, $j = 1, 2, ...p$, an approximate t-test and associated confidence interval can be obtained by approximating the distribution of $\left(\hat{\beta}_j - \beta_j\right)/s.e(\hat{\beta}_j)$ by a suitable t-distribution (Verbeke and Molenberghs, 2000), where the degrees of freedom needed are estimated from the data. Testing general linear hypotheses of the expression (4.41) is mainly based on an F-approximation to the distribution of

$$F = \frac{\left(\hat{\beta} - \beta\right) L' \left[L\left(\sum_{i=1}^{N} X_i' V_i^{-1}(\hat{\alpha}) X_i\right)^{-1} L'\right]^{-1} L\left(\hat{\beta} - \beta\right)}{rank(L)} \tag{4.43}$$

where rank (L) is the numerator degree of freedom and the denominator degrees of freedom have to be estimated from the data. There are a number of methods in literature used to estimate the appropriate degrees of freedom for t-or F-test, among others. Satterthwaite approximation is commonly used in SAS (Verbeke and Molenberghs, 2000).

The likelihood ratio test is also used for comparison of nested models with different mean structures. Suppose that null hypothesis of interest is given by

$$H_0 : \beta \in \Theta_{\beta_0},$$

for some subspace Θ_{β_0} of the parameter space Θ_β of fixed effects β. Let L_{ML} denote the ML likelihood function and let $-2ln\lambda_N$ be the likelihood ratio test statistic defined as

$$-2ln\lambda_N = -2ln\left[\frac{L_{ML}(\hat{\theta}_{ML,O})}{L_{ML}(\hat{\theta}_{ML)}}\right] \tag{4.44}$$

where $\hat{\theta}_{ML,O}$ and $\hat{\theta}_{ML}$ are maximum estimates obtained from maximizing L_{ML} over Θ_{β_0} and Θ_β, respectively (Verbeke and Molenberghs, 2000). It then follows from classical likelihood theory that, under some regularity conditions, $-2ln\lambda_N$ follows asymptotically under H_0 a χ^2 distribution with degrees of freedom equal to the difference between the dimensions k and of Θ_β and the dimension of $\hat{\theta}_{ML,O}$. The likelihood ratio tests result is valid if the model is fitted using ML and not valid when REML is used.

This is because REML log-likelihood functions are based on different observations, which makes them no longer comparable (Verbeke and Molenberghs, 2000).

When the interest is also in inference for some of the variance components in D, standard asymptotic Wald, likelihood ratio, and score tests can be used, as long as the hypotheses to be tested are not on the boundary of the parameter space (Molenberghs and Verbeke, 2005). The classical Wald, likelihood ratio or score test are not suitable for testing whether the variance τ^2 of the single random effect in GLMM equal to zero; this means that $H_0 : \tau^2 = 0$ versus $H_A : \tau^2 > 0$. In this case, the null hypothesis is on the boundary of the parameter space where $\tau^2 \geq 0$. Therefore, under H_0, the Z-statistic cannot be normally distributed with mean zero because the estimation of τ^2 is strictly positive normal distribution in 50% of the cases, and will be equal to zero in the other 50% of the cases. Therefore the null distribution is given by a mixture of chi-squared distributions. Equivalent properties can be derived from the one-sided likelihood ratio test (Self and Liang, 1987; Stram and Lee, 1994; Verbeke and Molenberghs, 2000) and one-sided score test (Silvapulle and Silvapulle, 1995; Verbeke and Molenberghs, 2003). Nevertheless, the general theory on test of hypotheses on the boundary of the parameter space is much more general, and can be applied equally well to GLMM settings (Self and Liang, 1987; Stram and Lee, 1994; Silvapulle and Silvapulle, 1995). Furthermore, even with the information criteria, there are still concerns about the boundary effects and estimation of the degrees of freedom for random effects (Vaida and Blanchard, 2005).

4.4. Generalized linear models applied to binary outcomes

The mixed-effects logistic regression model is a common choice for the analysis of multilevel dichotomous data and is the most used in GLMM. In GLMM setting, this model uses the logit link and is given by

$$g\left(\mu_{ijk}\right) = \text{logit}(\mu_{ijk}) = \log\left[\frac{\mu_{ijk}}{1-\mu_{ijk}}\right] = \eta_{ijk} \tag{4.45}$$

The conditional expectation $\mu_{ijk} = E\left(Y_{ijk}|b_i, x_i\right)$ equals $P\left(Y_{ijk}|b_i, x_{ijk}\right)$, namely, the conditional probability of a response given the random effects and the covariates values, where Y_{ijk} is the ith response in the jth household with kth primary sampling unit. This model can also be written as

$$P\left(Y_{ijk}|b_i, x_{ijk}, z_{ijk}\right) = g^{-1}\left(\eta_{ijk}\right) \tag{4.46}$$

where $g^{-1}\left(\eta_{ijk}\right)$ is commonly known as logistic cumulative distribution function (cdf) and is given by

$$g^{-1}\left(\eta_{ijk}\right) = \left[1 + \exp\left(-\eta_{ijk}\right)\right]^{-1}$$

The logistic distribution simplifies parameter estimation because its probability density function is related to its cdf in a simple way (Agresti, 2002). The probit model that is based on the standard normal distribution is frequently proposed as an alternative to the logistic model. For the probit model, the normal cdf and pdf replace their logistic counterparts.

4.4.1. Application to the determinants of poverty of household in Rwanda.

In the previous chapter, the survey logistic regression model was used but this model is survey based. The data was collected using multi-stage sampling where the primary sampling units or villages were selected at random and this may result in some variability among these primary sampling units. Therefore, in order to account for the possible variability between the primary sampling units, we used GLMM that includes the random effects.

Data analysis

The data was analyzed using SAS 9.3, where various approaches of estimation such as pseudo-likelihood, maximum likelihood with classical Gaussian and Adaptive quadrature and maximum likelihood with Laplace approximations were considered. The GLIMMIX procedure distinguishes two types of random effects. Depending on whether the variance of the random effect is contained in D matrix (commonly known as G in SAS notation) or in R matrix. They are referred to as "G-side" and "R-side "random effects. R-side effects are also called "residual" effects.

The models without G-side effects are known as marginal or population-averaged models. Models fit with the GLIMMIX procedure can have none, one or more of each type of effects. Note that R-side effect in GLIMMIX procedure is equivalent to a repeated effect in the MIXED procedure.

The analysis was done based on classical Gaussian and adaptive Gaussian quadrature and Laplace approximations. In order to find the effect of difference, a number of quadrature points (Q=2,6,8,9,10,15,25) were considered. However, the use of differente quadrature point did not lead to considerable difference for parameter estimation. But for quadrature less than 9, there were slightly difference for estimation of parameters. But from 10 and above no difference between parameter estimates was found. The Log pseudo-likelihood and chi-square test were used to assess the model goodness-of-fit. The statistical inferences for the covariance parameters were performed based on the likelihood ratio test. Model selection was achieved by first including into the model all predictor variables and then evaluating whether or not interaction terms needed to be incorporated. This was achieved by fitting model effects one at time, each of the interaction terms formed from the predictor variables. Finally, only three significant two way interactions were retained and the final model is given by.

$$\begin{aligned} \log\left(\frac{\mu_j}{1-\mu_j}\right) &= \beta_0 + \beta_1 Education_j + \beta_2 Gender_j \\ &+ \beta_3 \text{Place of residence}_j + \beta_4 Province_j + \beta_5 Size_j \\ &+ \beta_6 Province_j * \text{Place of residence}_j + \beta_7 Age_j * Gender_j + b_{0j} \end{aligned} \tag{4.47}$$

where $\beta_1, \beta_, ..., \beta_7$ are the unknown parameter coefficients of fixed effects and b_{0j} is the random intercept.

4.4.2. Results and Interpretations.

The ratio of generalized chi-square statistic and its degrees of freedom is 0.94 and is close to 1. This is the measure of the residual variability in the marginal distribution of the data. Since the value is close to 1, this indicates that the variability in the data has been properly modeled and then there is no residual over-dispersion. Age, level of education, and gender of household head, size of the household, province and place of residence of household head were found to significantly affect the household socio-economic status. The results of the main effect are in Table 4.2 and the results of interaction effects are in Table 4.3. From Table 4.2 it is observed that the level of education of the household head significantly affects the socio-economic status of the household, where the poverty of the household increases by decreasing the level of education of the household head. Furthermore, it is observed that a household with a household head with secondary education, primary education or no formal education is 4.7242, 15.830 or 26.7651 times more likely to be poor respectively than a household headed by a household head with tertiary education.

Interaction effects

The relationship between provinces (Kigali city, Southern, Western, Northern and Eastern) and place of residence (urban or rural) is presented in Figure 4.1. Each province of Rwanda has urban and rural places. As Figure 4.1 indicates, an urban household is less likely to be poor compared to a rural household in all provinces. These results revealed that a rural household from Southern province is the poorest Figure 4.1, while rural households from Western and Northern provinces are almost the same but more likely to be poor compared to a rural household from Eastern province. A rural household from Kigali is less likely to be poor as compared to a rural household from Eastern province Figure 4.1. The joint effect of gender and age of the household head is presented in Figure 4.2. From Figure 4.2 we observe that a household headed by a female is more likely to be poor than a household headed by a male. From Figure 4.3, it is observed that poverty decreases with the increasing age of household head regardless of the size of the household.

TABLE 4.1. Type III analysis of effects for GLMM

Effect	Numb Df	F value	$Pr > F$
Size	1	4.50	< 0.0339
Age of household head	1	37.03	$< .0001$
Education of household head	3	99.87	$< .0001$
Province of household head	4	28.74	$< .0001$
Place of residence of household head	1	27.84	$< .0001$
Gender of household head	1	39.42	$< .0001$
Size*Age of household head	1	38.41	$< .0001$
Province *place of residence of household head	4	2.37	0.0512
Gender * age of household head	1	15.47	$< .0001$

Table 4.2. Parameter estimates of main effect for poverty of household

Indicator	Estimate	S.E	P-Value	OR
Intercept	-3.2855	0.6367	< .0001	
Province (reference=Eastern)				
Kigali	-1.0610	0.3010	0.0005	0.3461
South	0.9026	0.1101	< .0001	2.4660
West	0.5613	0.1115	< .0001	1.7529
North	0.6232	.06119	< .0001	1.8649
Gender of the household head (reference=female				
Male	-.8897	0.1417	< .0001	0.41078
Education of Household head(reference=higher)				
Secondary	1.5527	0.6178	0.0120	4.7242
Primary	2.7619	0.6109	< .0001	15.830
No education	3.2871	0.6120	< .0001	26.7651
Age of the household head	0.0132	0.0034	< .0001	1.0133
Size	0.0775	0.0365	.0339	1.0806
Place of residence (reference= rural)				
Urban	-0.2405	0.3877	0.5354	0.7862

Table 4.3. Parameter estimates with two way interaction effects for poverty of household

Indicator	Estimate	S.E	P-Value	OR
Province * place of residence (ref.=East and rural)				
Kigali * urban	-1.3518	0.5375	0.0122	0.2588
South * urban	-0.7866	0.4504	0.0814	0.4554
West * urban	-0.9796	0.5576	0.0796	0.3755
North * urban	-0.0294	0.5716	0.9590	0.9710
Gender * age of the household head (ref.=Female)				
Male * age of the household	0.0116	0.0029	< .0001	1.0117
Size * age of the household head	-0.0047	0.0008	< .0001	0.9953

Figure 4.1. Province and place of residence

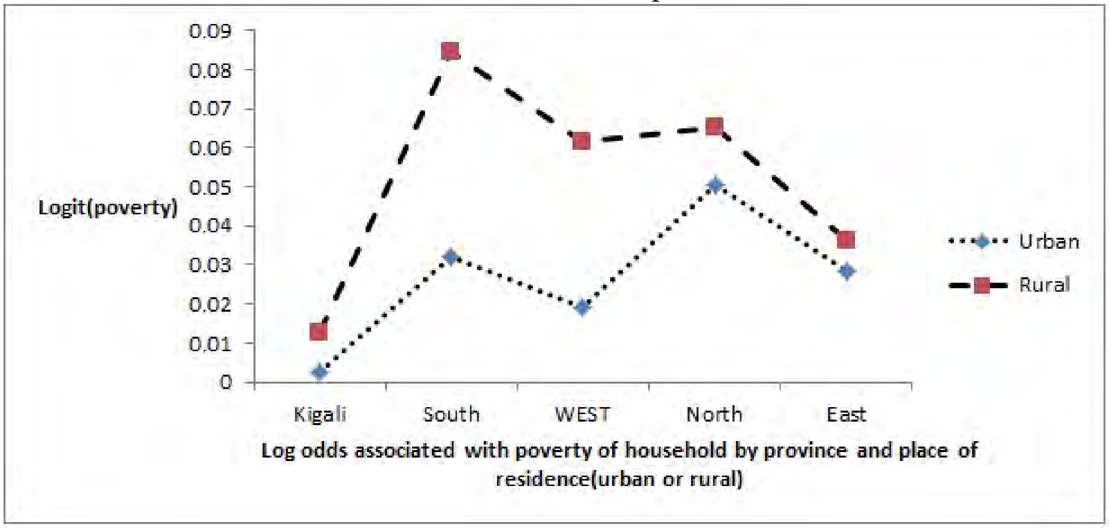


Figure 4.2. Age and gender of household head

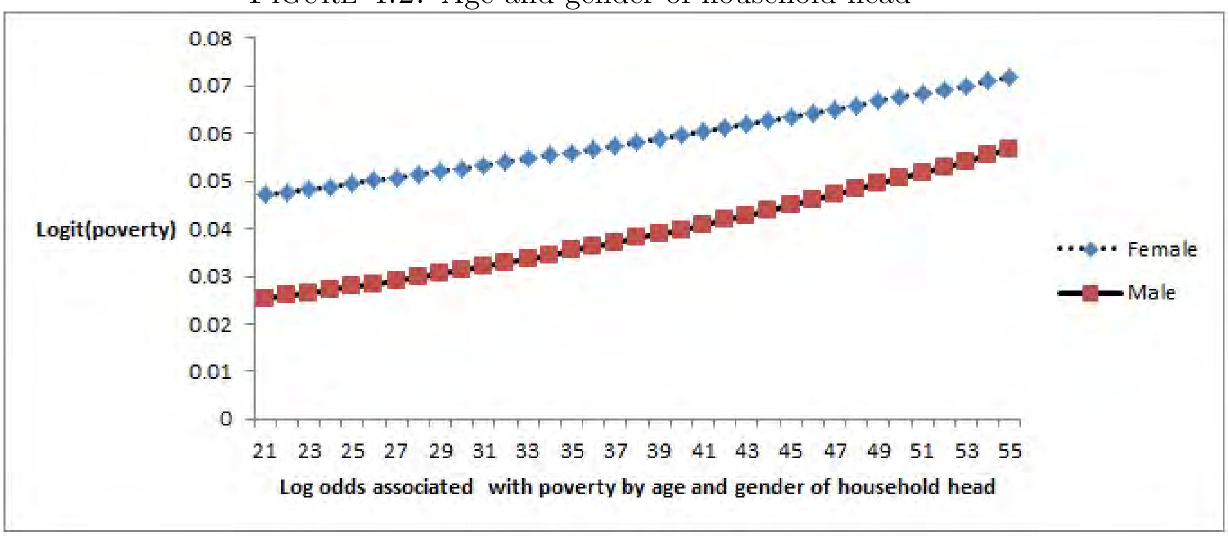
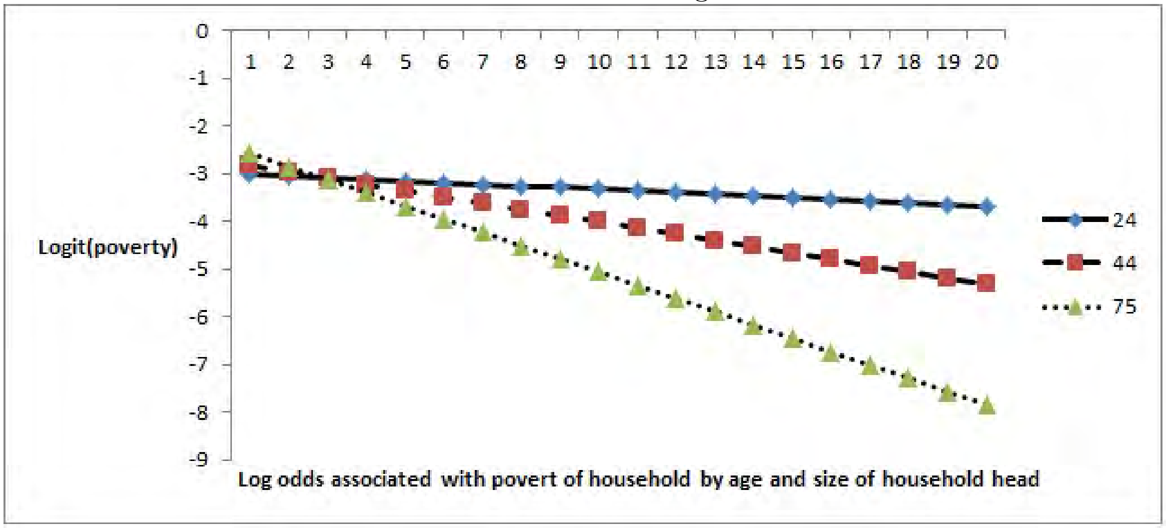


Figure 4.3. Size of household and age of household head



4.5. Summary

In this chapter, based on 2010 Rwanda Demographic and Health Survey data, we used generalized linear mixed model to identify the key determinants of poverty of households in Rwanda. The chapter extended the binary survey logistic to include the random effect. However, the findings of this study supported the findings of Habyarimana et al. (2015a) and the findings from binary survey logistic regression, where all these studies revealed that the key determinants of poverty are age of the household head, level of education of the household head, gender of household head, place of residence (urban or rural) of the household, province of residence of household and the size of the household (number of members of household). The current study also investigated the variability between the villages by including the random effects. Further, the magnitude of the effects of the above determinants is reduced in this study. In the next chapter we use multivariate joint model under GLMM in order to simultaneously measure the malnutrition on three anthropometric indictors and to examine the possible correlation between them.

CHAPTER 5

Multivariate joint modelling of the measures of malnutrition

Introduction

In chapter 3, we used proportional odds model without and with complex sampling design to identify the risk factors of malnutrition of children under five years, where only weight-for-height (known as underweight) was considered. In this chapter, the main objective is to utilize the multivariate joint model under GLMM in order to simultaneously identify the key determinants of malnutrition of children under five years based on three anthropometric indicators known as weight-for-height (underweight), height-for-age (stunting) and weight-for-age (wasting), to include random effects and to find out the possible correlation among these anthropometric indicators. In other words, a child might be in stunting status, or underweight status, or wasting status, or stunting and underweight, underweight and wasting or their combination. A separated generalized linear mixed model cannot determine the association between these three outcomes. The advantages of the joint model over the separate models include better control of type I error rates in multiple tests, possible gains in efficiency in the parameter estimates and the ability to answer intrinsically multivariate questions (Gueorguieva, 2001; Kandala et al., 2011b).

5.1. Model overview

Let us first consider a bivariate response variable and thereafter we will extend it to more than two response variables. The joint model formulation can be done in various approaches, such as Probit nomal formulation, Plackett-Dale formulation and generalized linear mixed model formulation among others. In this study we consider only the generalized linear mixed model formulation

Generalized linear mixed model formulation

In this case, the formulation uses both random effects and serial correlations. It is in general given by Molenberghs and Verbeke (2005) as

$$Y_i = \mu_i + \epsilon_i \tag{5.1}$$

where

$$\mu_i = \mu_i(\eta_i) = h\left(X_i\beta + Z_ib_i\right) \tag{5.2}$$

Assume $b_i \sim N(0, D)$ are the q-dimensional random effects. In this case, the components of the inverse link functions h are allowed to change with the nature of the various outcome variables in Y_i. Further, the variance of ϵ_i depends on the mean-variance links of the different outcome variables; in addition it contains a correlation matrix $R_i(\alpha)$ and the over-dispersion parameter ϕ. Once there is no random effects in expression (5.2), then it reduces to a marginal model referred to as marginal generalized linear models (MGLM). However, when there are no residual correlations in $R_i(\alpha)$, then it reduces to a purely random effects model or a conditional independence model that is also a generalized linear mixed model.
The variance-covariance matrix of Y_i is obtained from a general first order approximate expression (Molenberghs and Verbeke, 2005) given by

$$V_i = Var(Y_i) \simeq \Delta_i Z_i D Z_i' \Delta_i' + \Sigma_i \tag{5.3}$$

where

$$\Delta_i = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)|_{b_i=0} \tag{5.4}$$

and

$$\Sigma_i \simeq \Phi_i^{1/2} A_i^{1/2} R_i(\alpha) A_i^{1/2} \Phi_i^{1/2} \tag{5.5}$$

where A_i is a diagonal matrix containing the variance following from the generalized linear specification of $Y_{ik}, k = 1, 2$ for a given random effects $b_i = 0$; in other words the diagonal elements are given by $v\left(\mu_{ij}|b_i = 0\right)$. Similarly, Φ_i is also a diagonal matrix however with the overdispersion parameters along the diagonal. Σ_i captures the variance-covariance in residual error ϵ_i and first term of the right hand side of expression (5.3) corresponds to the random effects structure of $h\left(X_i\beta + Z_ib_i\right)$. $R_i(\alpha)$ is the correlation matrix. Furthermore, if the outcome component is normally distributed then the overdispersion parameter is σ_i^2 and the variance function is 1 (Molenberghs and Verbeke, 2005). In the case of binary outcome variable with logit link, we get

$$\mu_{ij}(b_i = 0)\left[1 - \mu_{ij}(b_i = 0)\right] \tag{5.6}$$

The evaluation under $b_i = 0$ derives from a Taylor series expansion of the mean component around $b_i = 0$. When the exponential family specification is considered for all components, with canonical link, $\Delta_i = A_i$, then variance covariance matrix of Y_i can be written as follows

$$Var(Y_i) \simeq \Delta_i Z_i D Z_i' \Delta_i' + \Phi_i^{1/2} \Delta_i^{1/2} R_i(\alpha) \Delta_i^{1/2} \Phi_i^{1/2} \tag{5.7}$$

under conditional independence R_i vanishes and

$$var(Y_i) = \Delta_i Z_i D Z_i' \Delta_i' + \Phi_i^{1/2} \Delta_i^{1/2} \Phi_i^{1/2} \tag{5.8}$$

A model with no random effects for the marginal generalized linear model (MGLM) has the form

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \mu_1 + \lambda b_i + \alpha X \\ \frac{exp[\mu_2 + b_i + \beta X_i]}{1 + exp[\mu_2 + b_i + \beta X_i]} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \tag{5.9}$$

where λ is the scale parameter included in the continuous of an otherwise random-intercept model, given the continuous and binary outcome are measured on different scales. Therefore, in this case

$$Z_i = \begin{pmatrix} \lambda \\ 1 \end{pmatrix}, \Delta_i = \begin{pmatrix} 1 & 0 \\ 0 & v_{i2} \end{pmatrix}, \Phi = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 1 \end{pmatrix}$$

with $v_{i2} = \mu_{i2}(b_i = 0)(1 - \mu_{12}(b_i = 0))$. In addition let ρ be the correlation between ϵ_{i1} and ϵ_{i2}. However, Z_i is not a design matrix as it contains unknown parameters. The variance-covariance matrix of Y_i from (5.1) becomes

$$\begin{aligned} V_i &= \begin{pmatrix} \lambda^2 & v_{i2}\lambda \\ v_{i2}\lambda & v_{i2}^2 \end{pmatrix} \tau^2 + \begin{pmatrix} \sigma^2 & \rho\sigma\sqrt{v_{i2}} \\ \rho\sigma\sqrt{v_{i2}} & v_{i2} \end{pmatrix} \\ &= \begin{pmatrix} \lambda^2\tau^2 + \sigma^2 & v_{i2}\lambda^2\tau^2 + \rho\sigma\sqrt{v_{i2}} \\ v_{i2}\lambda\tau^2 + \rho\sigma\sqrt{v_{i2}} & v_{i2}^2\tau^2 + v_{i2} \end{pmatrix} \end{aligned} \tag{5.10}$$

Thus, the derived approximate marginal correlation function is given by

$$\rho(\beta) = \frac{v_{i2}\lambda\tau^2 + \rho\sigma\sqrt{v_{i2}}}{\sqrt{\lambda^2\tau^2 + \sigma^2}\sqrt{v_{i2}^2\tau^2 + v_{i2}}} \tag{5.11}$$

The equation (5.11) depends on the fixed effects through v_{i2}. The model with no random effects is written as follows:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \mu_2 + \beta X_i \\ \frac{exp[\mu_1 + b_i + \beta X_i]}{1+exp[\mu_1 + \beta X_i]} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \tag{5.12}$$

and equation (5.10) reduces to ρ, by virtue of its fully marginal specification. Under conditional independence, ρ in equation (5.10) satisfies $\rho \equiv 0$ and equation (5.11) reduces

$$\rho(\beta) = \frac{v_{i2}\lambda\tau^2}{\sqrt{\lambda^2\tau^2 + \sigma^2}\sqrt{v_{i2}^2\tau^2 + v_{i2}}} \tag{5.13}$$

Equation (5.13) is simpler than equation (5.11) but equation 5.13 is a function of the fixed effects.

If both end points are binary, equation (5.13) can be reduced to

$$\rho(\beta) = \frac{v_{i2}v_{i2}\tau^2 + \rho\sigma\sqrt{v_{i1}v_{i2}}}{\sqrt{v_{i1}^2\tau^2 + v_{i1}}\sqrt{v_{i2}^2\tau^2 + v_{i2}}} \tag{5.14}$$

with again a constant correlation ρ when there are no random effects and with no residual correlation we get

$$\rho(\beta) = \frac{v_{i2}v_{i2}\tau^2}{\sqrt{v_{i1}^2\tau^2 + v_{i1}}\sqrt{v_{i2}^2\tau^2 + v_{i2}}} \tag{5.15}$$

The equation (5.15) can be performed with general random effects design matrices Z_i and for more than two components of arbitrary nature not necessarily continuous and binary.

Two binary responses

Similarly, when both sequences of outcomes are binary, a generalized linear mixed model (GLMM) can be assumed with correlated random effects (Faes et al., 2008) as follows

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \frac{exp[\alpha_1 + \beta_1 X_i + b_{i1}]}{1+exp[\alpha_0 + \beta_1 X_i + b_{i1}]} \\ \frac{exp[\alpha_2 + \beta_2 X_i + b_{i2}]}{1+exp[\alpha_0 + \beta_2 X_i + b_{i2}]} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1j} \\ \epsilon_{i2j} \end{pmatrix}$$

where the random effect $b_{i1}$and b_{i2} are normally distributed, ϵ_{i1j} and ϵ_{i2j} are independent. It is assumed that $Var(\epsilon_{i1j}) = v_{i1j} = \pi_{i1j}\,(b_{i1} = 0)\,[1 - \pi_{i1j}(b_{i1})]$ and $Var(\epsilon_{i2j}) = v_{i2j} = \pi_{i2j}\,(b_{2i} = 0)\,[1 - \pi_{i2j}(b_{i2})]$. The approximate variance-covariance matrix of the two binary measurements for subject i at time j is given by

$$V_{i1} = \begin{pmatrix} v_{i1j}^2\tau_1^2 + v_{i1j}\rho\tau_2 v_{i1j}v_{i2j} \\ v_{i1j}\rho\tau_2 v_{i1j}v_{i2j} + v_{i2j}^2\tau_2^2 + v_{i2j} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1j} \\ \epsilon_{i2j} \end{pmatrix}$$

and the correlation between the two outcomes in this case is given by

$$\rho_{Y_1Y_2} = \frac{\rho\tau_1\tau_2 v_{i1j}v_{i2j}}{\sqrt{v_{i1j}^2\tau_1^2 + v_{i1j}}\sqrt{v_{i2j}^2\tau_2^2 + v_{i2j}}} \tag{5.16}$$

Two continuous responses

If both responses variables are continuous, a linear mixed model can be used with correlated random effects. The correlation between the two response variables in this case is given by

$$\rho_{Y_1Y_2} = \frac{\rho\tau_1\tau_2}{\sqrt{\tau_1^2 + \sigma_1^2}\sqrt{\tau_2^2 + \sigma_2^2}} \tag{5.17}$$

It is possible to perform easily the above calculations in the case of general random effects design matrices Z_i and for more than two components of arbitrary nature and which are not necessarily continuous and binary.

In the case of general model, there is no need to specify full joint distribution, even when it is assumed that the first one is continuous and the second one to be Bernoulli distributed. We can still leave the specification of the joint moments to the second one, by way of marginal correlation.

A full joint specification would need full bivariate model specification, conditional upon the random effects, together with normality assumptions made about the random effects.

5.1.1. Maximum likelihood estimation.

The marginal likelihood in bivariate GLMM is obtained as in usual GLMM by integrating out the random effects (Gueorguieva, 2001)

$$\prod_{i=1}^{n}\int\int\left[\prod_{j=1}^{n_{i1}} f_1\left(y_{i1}|b_{1i};\beta_1,\phi_1\right)\prod_{j=1}^{n_{i2}} f_2\left(y_{i2}|b_{2i};\beta_2,\phi_2\right)\right] f\left(b_{i1},b_{i2};\Sigma\right) db_{i1}db_{i2} \qquad (5.18)$$

where f denotes the multivariate normal density of the random effects. The integral (5.18) is usually intractable and some numerical, stochastic or analytical approximation must be used. Methods for model fitting of the univariate GLMM include marginal maximization using Gaussian quadrature or Monte Carlo approximation Fahrmeir and Tutz (1994), penalized quasi-likelihood Breslow and Clayton (1993); Wolfinger et al. (1994), Monte Carlo EM algorithm (McCulloch, 1997; Booth and Hobert, 1999), Monte Carlo Newton-Raphson algorithm and simulated maximum likelihood (McCulloch, 1997). All these methods can be extended to multivariate GLMM. However, the maximum likelihood maximization is criticized when the number of outcomes increasing the computation becomes cumbersome; it is only feasible when the number of outcomes is sufficiently low (typically dimension 2 or 3 at most)(Molenberghs and Verbeke, 2005). In order to overcome this problem the next subsection presents model fitting procedure that is applicable irrespective to the dimensionality problem.

5.2. Extension to higher-dimensional data

Let m be the dimension or number of outcomes variables needed to be modeled jointly, defined as follows: $Y_{ik} = \left(Y_{ik1}, Y_{ik2}, ..., Y_{ikm_i}\right), k = 1, 2, ..., m$. The sequences Y_{ik} is a vector of m_{ki} measurements taken on subject i, for outcome k and Y_{ik} is not restricted to outcome of the same type; it can be either continuous or binary or mixed (Faes et al., 2008). Therefore, the m outcomes variables can then be simultaneously modeled by specifying a joint distribution for the random effects, in similar way as in the case of binary outcomes; however with an $m \times q$ dimensional random effects vector b_i.

All outcomes are not supposed to have the same type of model; a combination of linear, generalized linear, and non-linear mixed model is possible (Molenberghs and

Verbeke, 2005). Generally, in most applications, it will be assumed that conditionally on the random effects $b_{1i}, b_{2i}, ..., b_{mi}$, $y_{1i}, y_{2i}, ..., y_{mi}$ are independent. Finally, the model is completed by assuming that the vector b_i of all random effects for subjects i is multivariate normal with mean zero and covariance Σ given by

$$b_i = \begin{bmatrix} b_{1i} \\ b_{2i} \\ . \\ . \\ . \\ b_{mi} \end{bmatrix} \sim i.i.d.MVN\left(0, \Sigma\right) = MVN\left(\begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & . & . & . & \Sigma_{1m} \\ \Sigma_{12} & \Sigma_{22} & . & . & . & \Sigma_{2m} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \Sigma_{m1} & \Sigma_{m2} & . & . & . & \Sigma_{mm} \end{bmatrix}\right)$$

The matrices Σrs represent the covariances between b_{ri} and b_{si}, $r, s = 1, 2, ..., m$. Finally, Σ is the matrix with blocks Σ_{rs} as entries.

The estimation and inference are based on the marginal model of vector Y_i of all measurements of subject i. Therefore, assuming independence of outcomes conditional on the vector b_i of random effects, the likelihood contribution for subject i then becomes (Molenberghs and Verbeke, 2005; Fieuws and Verbeke, 2006; Faes et al., 2008)

$$L_i\left(\Theta|Y_{i1}, Y_{i2}, ..., Y_{im}\right) = \int_{\mathbb{R}^{mq}} \prod_{j=1}^{n_i} f_{ij}\left(y_{i1j}, y_{i2j}, ..., y_{imj}|b_i, \Theta\right) f(b_i|\Sigma) db_i \tag{5.19}$$

with $\Theta = (\beta, \alpha, \Sigma)$. However, computational problems often arise when m increases, owing to the $m \times q$-dimensional integral, especially when outcomes are of different type. In this case, rather than considering the full likelihood contribution for each subject i, one can avoid the computational complexity by using pseudo-likelihood approach, similar to the pairwise modelling approach proposed by Fieuws and Verbeke (2006). The full likelihood contribution for subject i is replaced by the pseudo-likelihood function

$$PL_i = \prod_{k=1}^{m-1} \prod_{l=k+1}^{m} L_{ikl}\left(\Theta|Y_{ik}, Y_{il}\right) = \prod_{k=1}^{m-1} \prod_{l=k+1}^{m} \int_{\mathbb{R}^{2q}} \prod_{j=1}^{n_i} f_{ij}\left(y_{ikj}, y_{ilj}|b_i^{kl}, \Theta\right) f(b_i^{kl}|\Sigma) db_i^{kl} \tag{5.20}$$

where each contribution L_{ikl} is equal to the bivariate likelihood function for outcomes k and l. Therefore, the $m \times q$-dimensional integration problem reduces to $2 \times q$-dimensional integrations. In practice, this is achieved by restructuring the data in all possible pairs of outcomes, and assuming, as working assumption, that conditional n random effects, all combinations of pair (k, l) and subject i are independent. Then, the inference for Θ follows from pseudo-likelihood theory, and is based on a sandwich-type robust variance estimator (Arnold and Strauss, 1991). The asymptotic multivariate normal distribution for $\hat{\Theta}$ is given by

$$\sqrt{N}\left(\hat{\Theta} - \Theta\right) \sim MVN\left(0, J(\Theta)^{-1}K(\Theta)J(\Theta)^{-1}\right) \tag{5.21}$$

where $J = J(\Theta)$ is a matrix with elements defined by

$$J_{pq} = -\sum_{k=1}^{m-1}\sum_{l=k+1}^{m} E\left(\frac{\partial^2 lnL_{ikl}(\Theta|Y_{ik}, Y_{il})}{\partial\theta_p\partial\theta_q}\right) \tag{5.22}$$

and $K = K(\Theta)$ is symmetric matrix with elements

$$K_{pq} = -\sum_{k=1}^{m-1}\sum_{l=k+1}^{m} E\left(\frac{\partial lnL_{ikl}(\Theta|Y_{ik}, Y_{il})}{\partial\theta_p}\frac{\partial lnL_{ikl}(\Theta|Y_{ik}, Y_{il})}{\partial\theta_q}\right) \tag{5.23}$$

The main advantage of pseudo-likelihood approach is the close connection with likelihood that enabled Geys et al. (1997) to construct pseudo-likelihood ratio test statistics. As it is known that Wald tests can yield erroneous results, especially when a variable has a large effect in the model (Geys et al., 1997), the pseudo-likelihood ratio test statistic is preferable in this situation. Suppose we are interested in testing the null hypothesis $H_0 : \gamma = \gamma_0$, where γ is an r-dimensional subvector of the p-dimensional vector of regression parameters β and write β as $\left(\gamma^T, \delta^T\right)^T$. Therefore, the pseudo-likelihood ratio test statistic, in this case is given by

$$G^{*2} = \frac{2}{\overline{\lambda}}\left[PL(\hat{\beta}_N) - PL\left(\gamma_0, \hat{\delta}(\gamma_0)\right)\right] \tag{5.24}$$

and is approximately χ^2_r distributed, where $\hat{\beta}_N$ is the pseudo-likelihood parameter estimate of β and $\hat{\delta}(\gamma_0)$ denotes the maximum pseudo-likelihood estimator in the subspace where $\gamma = \gamma_0$. In addition, $\overline{\lambda}$ is the mean of the eigenvalues of $(J^{\gamma\gamma})^{-1}\Sigma_{\gamma\gamma}$, where $J^{\gamma\gamma}$ is the $r \times r$ submatrix of the inverse of J and $\Sigma_{\gamma\gamma}$ is the submatrix of $\Sigma = J^{-1}KJ^{-1}$.

5.3. Application to the determinants of malnutrition of children under five years in Rwanda

Introduction

In literature, there are many studies done on determinants of malnutrition of children under five years of age, for instance Das and Rahman (2011); Kandala et al. (2011a) and Habyarimana et al. (2014) among others. All these studies considered underweight, stunting or wasting separately. However, a child may be well nourished or malnourished (stunted, or underweight, or wasted, or wasted and underweight, or underweight and stunted). A separate model cannot determine the association between these three outcomes. For this reason, the current research utilizes a joint model for a multivariate generalized linear mixed model to simultaneously identify the key determinants of stunting, wasting and underweight and to find out the possible correlation among them. The advantages of the joint model over separate models include better control of the Type I error rates in multiple tests, possible gains in efficiency in the parameter estimates and the ability to answer intrinsically multivariate questions (Gueorguieva, 2001; Kandala et al., 2011b).

Model formulation for three outcomes

Let us denote the response vector for the i^{th} subject as $Y_i = (Y_{i1}', Y_{i2}', Y_{i3}')'$, where $Y_{i1} = (y_{i11}, y_{i12}, ..., y_{i1n_{i1}})$, $Y_{i2} = (y_{i21}, y_{i22}, ..., y_{i2n_{i2}})$, $Y_{i3} = (y_{i31}, y_{i32}, ..., y_{i3n_{i3}})$ and are the repeated measurement of the first and the second variable. We assume that $y_{i1j}, j = 1, 2, ..., n_{i1}$, are conditional independent given b_{i1} with the density function $f_1(.)$ in the exponential family, $y_{i2j}, j = 1, 2, ..., n_{i2}$, are conditional independent given b_{i2} with the density function $f_2(.)$ in the exponential family. Similarly, $y_{i3j}, j = 1, 2, ..., n_{i3}$, are conditional independent given b_{i3} with the density function $f_3(.)$ in the exponential family also Y_{i1},Y_{i2} and Y_{i3} are conditional independent given $b_i = (b_{i1}', b_{i2}')', b_{i3}')'$ and the response on different subjects are independent. The conditional means of y_{i1j},y_{i2j} and y_{i3j} are denoted as μ_{i1j},μ_{i2j} and μ_{i3j} respectively. Let $\mu_{i1} = (\mu_{i11}, \mu_{i12}, ..., \mu_{i1n_{i1}})'$, $\mu_{i2} = (\mu_{i21}, \mu_{i22}, ..., \mu_{i2n_{i2}})'$ and $\mu_{i3} = (\mu_{i31}, \mu_{i32}, ..., \mu_{i3n_{i3}})'$. Thus, at first stage the mixed model specification is assumed to be

$$g_1(\mu_{i1}) = X_{i1}\beta_1 + Z_{i1}b_{i1} \tag{5.25}$$

$$g_2(\mu_{i2}) = X_{i2}\beta_2 + Z_{i2}b_{i2} \tag{5.26}$$

$$g_3(\mu_{i3}) = X_{i3}\beta_3 + Z_{i3}b_{i3} \tag{5.27}$$

where β_1, β_2 and β_3 are $(k_1 \times 1)$, $(k_2 \times 1)$ and $(k_3 \times 1)$ dimensional unknown parameter vectors, X_{i1}, X_{i2} and X_{i3} are $(n_{i1} \times k_1)$, $(n_{i2} \times k_2)$ and $(n_{i3} \times k_3)$ dimensional design matrices for the fixed effects, Z_{i1}, Z_{i2} and Z_{i3} are $(n_{i1} \times q_1)$, $(n_{i2} \times q_2)$ and $(n_{i3} \times q_3)$ dimensional design matrices for the random effects and g_1, g_2 and g_3 are applied componentwise to μ_{i1},μ_{i2} μ_{i3}. At second stage,

$$b_i = \begin{bmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix} \sim i.i.d.MVN(0, \sum) = MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sum_{11} & \sum_{12} & \sum_{13} \\ \sum'_{12} & \sum_{22} & \sum_{23} \\ \sum'_{13} & \sum'_{23} & \sum_{33} \end{bmatrix} \right) \tag{5.28}$$

where $\sum$, $\sum_{11}$, $\sum_{22}$ and $\sum_{33}$ are unknown positive definite matrix. If $\sum_{12} = \sum_{13} = \sum_{23} = 0$ then the above model is equivalent to three separate GLMMs for the three outcome variables. Advantages of joint model include the better control of type I error rates in multiple tests. This may lead to possible gains in efficiency in the parameter estimates and the ability to answer intrinsically multivariate questions (Gueorguieva, 2001; Molenberghs and Verbeke, 2005).

The marginal means and the marginal variance of Y_{i1}, Y_{i2} and Y_{i3} for the model defined by equation(5.25), (5.26) and (5.27) are the same as those of the GLMM considering one variable at time

$$E(y_{i1}) = E[\mu_{i1}(\beta_1, b_{i1})]$$

$$E(y_{i2}) = E[\mu_{i2}(\beta_2, b_{i2})]$$

$$E(y_{i3}) = E[\mu_{i3}(\beta_3, b_{i3})]$$

$$var(y_{i1}) = E[\phi_1 V(\mu_{i1})] + Var[\mu_{i1}]$$

$$var(y_{i2}) = E[\phi_2 V(\mu_{i2})] + Var[\mu_{i2}]$$

$$var(y_{i3}) = E[\phi_3 V(\mu_{i3})] + Var[\mu_{i3}]$$

where $V(\mu_{i1})$, $V(\mu_{i2})$ and $V(\mu_{i2})$ denote the the variance functions corresponding to the exponential family distributions for the three response variables, $Var[\mu_{i1}] = var[E(y_{i1}|b_{i1})]$, $Var[\mu_{i2}] = var[E(y_{i2}|b_{i2})]$ and $Var[\mu_{i3}] = var[E(y_{i3}|b_{i3})]$.

Data analysis

Several procedures are available for estimating the parameters involved in joint models. The parameter estimation in the joint models can be done using either numerical approximation such as Gaussian quadrature, adaptive Gaussian quadrature or Laplace approximation or approximation of the data by the pseudo-likelihood in which pseudo data are created based on linearization of the mean. More specifically, the pseudo-likelihood approach is used when estimating the parameters in marginal models and random effects with or without serial correlation, whilst quadrature or Laplace approximations can only estimate parameters in conditional independent random effects models. SAS procedure PROC GLIMMIX (SAS 9.3) can be used for estimating the parameter in case of a pseudo-likelihood approach while the NLMIXED procedure can be used for parameters estimation using Laplace approximation or Gaussian quadrature. In the current study PROC GLIMMIX (SAS 9.3) is utilized.

5.4. Results and interpretations

The current research considered many child malnutrition factors such as child characteristics (gender of the child, birth weight, birth order, child's age, incidence of fever during the two weeks prior to the survey, diarrhea), mother's characteristics such as: education level, mother's age at the birth, body mass index, incidence of anemia, mother's knowledge of nutrition, assistance at delivery, antenatal visits; environmental characteristics such as: region or province, source of drinking water, place of residence, toilet facilities; and household characteristics such as: size of household and household wealth index. In Table 5.3 any variable that is at least significant at one of the three anthropometric indicators is considered as a determinant of malnutrition and is hence reported. From Table 5.4 a strong positive correlation is observed between underweight and wasting as well as between underweight and stunting. This is not surprising because underweight is known to be the composite index between stunting and wasting; this is supported by the findings of the current study and is also consistent with the findings of other researchers (Onis, 2000; Nguefac-Tsague and Dapi, 2011; Nguefac-Tsague et al., 2013).

Stunting: This study reveals that child's age, birth order, mother's age at childbirth, mother's education, gender of the child, birth weight, province, mother's knowledge of nutrition and wealth index are the determinants of stunting (low height-for-age). These results are reported in Table 5.3. The age of the child significantly affects the height-for-age of the child. A child aged between 12 and 23 months is 3.428 (p-value $< .0001$) times more likely to be stunted than an infant (aged 0-11 months). Birth order significantly affects the height-for-age of the child. Sixth born children and those born thereafter are 1.652 (p-value=0.0002) times more likely to be stunted than first born children. Mother's age at childbirth significantly affects the height-for-age of the child Table 5.3. A child born to mother aged younger than 21 years old is 1.737 (p-value=0.0096) times more likely to be in stunting status than a child born to mother older than 21 years of age. A mother's level of education also significantly affects the height-for-age of the child. The z-score of height-for-age increases with increasing education levels of the mother Table 5.3. Therefore, stunting reduces as the mother's level of education increases. Further, a child born to a mother with a primary education or a secondary or higher education level is 0.0518 (p-value $< .0001$) or 0.0406 (p-value $< .0001$) times less likely to be underweight than a child born to a mother with no education, respectively. The gender of the child significantly affects his/her height-for-age Table 5.3.

The risk of having a low height-for-age z-score is 0.639 (p-value $< .0001$) times lower among female children than male children. Birth weight also significantly affects the weight-for-age of the child Table 5.3. A child born with low weight (< 2500g) is 1.786 (p-value =0.0115) times more likely to be underweight than a child born with a higher weight (≥ 2500g).

Province of birth significantly affects the height-for-age of the child Table 5.3. The risk of having a lower height-for-age z-score is 1.544 (p-value=0.0409) times higher among children born in Western province than children born in Eastern province. A child born in Southern province is 1.403 (p-value=0.023) more likely to be stunted as compared to a child born in Kigali city.

The mother's knowledge of nutrition is also seen to significantly affect the height-for-age of the child. A child born to a mother without knowledge of nutrition is 1.296 (p-value=0.0047) times more likely to be stunted than a child born to a mother with some knowledge of nutrition. The wealth index significantly affects the height-for-age of the child Table 5.3, as stunting increases with a decreasing wealth index. A child born into a poor family is 1.543 (p-value=0.0079) times more likely to be stunted than a child born into a rich family.

Wasting: The findings of this research show that a child's age, birth order, the birth weight of the child, wealth index, body mass index of the mother, recent incidence of fever, and source of drinking water all significantly affect the height-for-weight of the child Table 5.3. The age of the child is seen to significantly affect the height-for-weight of the child. A child aged between 12 and 23 months is 0.406 (p-value=0.0028) times less likely to be wasted than an infant (aged 0 to 11 months). Similarly, a child aged between 23 and 59 months is 1.826 (p-value=0.0442) times more likely to be wasted than an infant. The birth order also significantly affects the height-for-weight of the child Table 5.3. A sixth (or later) born child is 2.651 (p-value=0.0311) times more likely to be wasted than a first born child. Further, birth weight significantly affects the height-for-weight of the child Table 5.3. A child born with a higher weight ($\geq$ 2500g) is less likely to be wasted than a child born with a lower weight ($< 2500g$). The wealth index significantly affects the height-for-weight of the child Table 5.3. A child born into a poor family is 3.680 (p-value=0.0194) times more likely to be wasted than a child born into a rich family. Body mass index of the mother is also an indicator of wasting Table 5.3. A child born to an underweight mother (BMI $<$ 18.5) is 3.222 (p-value=0.0052) times more likely to be wasted than a child born to a normal or obese mother ($\geq$ 18.5). In other words, these results show that there is an association between weight of the mother and nutrition status of the child.

Incidence of fever is also seen to significantly affects the height-for-weight of the child Table 5.3. A child reported to have had a fever in the two weeks prior to the survey is 1.763 (p-value=0.0427) times more likely to be wasted than a child who did not have a fever during the last two weeks before the survey.

Source of drinking water is also associated with nutrition status (weight-for-height) (Table 5.3. A child born to a mother from a family where piped water is delivered into their dwelling or yard is 0.130 (p-value=0.0045) times less likely to be wasted than a child born into a family where water comes from other sources (not piped in dwelling/yard, public tap and protected spring or well).

A child born into a family where they use water from a public tap is 0.259 (p-value=0.0007) less likely to be wasted than a child from a family where they use water from other sources (not piped in dwelling/yard, not from protected spring or well). In other words, water that is not piped in dwelling/yard, water that is not from public taps or which is not from a protected spring or well may be associated with childhood diseases such as diarrhea, among others. Potable water is very important in order to fight wasting and other related consequences.

Underweight: The results from this study reveal that the child's age, birth order, education level of the mother, gender of the child, birth weight of the child, mother's knowledge of nutrition, multiple births, incidence of anemia and body mass index of the mother are the key determinants of malnutrition of the child Table 5.3. The child's age significantly affects the weight-for-age of the child Table 5.3. A child aged 23 months and more is 0.798 (p-value=0.0411) times less likely to be underweight than an infant. Birth order also significantly affects the weight-for-age of the child such that underweight increases with increasing the birth order Table 5.3. A second or third born child is at a 1.296(p-value=0.0473) times greater risk of being underweight first born. Similarly, a fourth or fifth born child is 1.346 (p-value=0.0341) times more likely to be underweight than first born. Further, a sixth or later born child is 2.948 (p-value$<$ .0001) times more likely to be underweight than first born.

The mother's level of education significantly affects the weight-for-age of the child. The degree to which a child is underweight decreases with an increase in the mother's level of education Table 5.3. Further, a child born to a mother with primary education or a secondary or higher education level is 0.097 (p-value $<$ .0001) or 0.058 (p-value $<$ .0001) less likely to have an underweight status than a child born to a mother with no education, respectively.

The gender of the child significantly affects the weight-for-age of the child Table 5.3. A female child is 0.617 (p-value $<$ .0001) less likely to be underweight than a male child. Birth weight significantly affects the weight-for-age of the child, as well Table 5.3. A child born with low birth weight (weight $<$ 2500g) is 3.16 (p-value $<$ .0001) times more likely to be underweight than a child born with a higher weight

($\geq 2500g$). The mother's knowledge of nutrition significantly affects the weight-for-age of the child Table 5.3. A child born to a mother without knowledge of nutrition is 1.416 (p-value=0.0015) times more likely to be underweight than a child born to a mother with some knowledge of nutrition. Multiple births significantly affect the weight-for-age of the child Table 5.3, where the degree of underweight increases with increasing the incidence of multiple birth. A child born as the first multiple (twin) is 3.842 (p-value=0.0002) times more likely to be underweight than a singleton child. Incidence of anemia also significantly affects the weight-for-age of the child. A child born to a non-anemic mother is 0.691 (p-value=0.0002) less likely to be underweight than a child born to an anemic mother. The body mass index of the mother is seen to significantly affect the weight-for-age of the child Table 5.3. A child born to an underweight mother (BMI < 18.5) is 3.096 (p-value $< .0001$) more likely to be underweight him/herself than a child born to a normal weight, overweight or obese mother (BMI≥ 18.5). Incidence of fever is also seen to significantly affect the weight-for-age of the child Table 5.3. A child who had a fever in the two weeks prior to the survey is 1.667 (p-value $< .0001$) times more likely to be underweight than a child who did not have a fever during the same time frame.

TABLE 5.1. Type 3 tests of fixed effects

Explanatory variable	Num. Df	Den. Df	F value	$Pr > F$
Child's age in months	6	9759	19.66	$< .0001$
Birth order	9	9759	7.74	$< .0001$
Mother's age at the birth	3	9759	2.4	0.0663
Mother's education level	6	9759	10.24	$< .0001$
Gender of child	3	9759	18.14	$< .0001$
Wealth index	6	9759	2.31	0.0311
Birth weights	3	9759	14.54	$< .0001$
Province/region	12	9759	2.53	0.0025
Knowledge on nutrition	3	9759	5.71	0.0007
Multiple birth	6	9759	3.13	0.0046
Incident of Anemia	9	9759	5.88	0.0005
Mother's marital status	9	9759	3.02	0.0013
Body Mass Index	3	9759	13.24	$< .0001$
Incidence of fever in last two weeks	3	9759	7.27	$< .0001$
Source of drinking water	9	9759	2.4	0.0043

Table 5.2. Fit statistics conditional distribution

-2log L(response—r.effect)	47534.31
Pearson Chi-square	7909.36
Pearson Chi-square/DF	0.94

Table 5.3. Parameter estimates for a joint marginal model for anthropometric measurements of malnutrition

	Wasting			Underweight			Stunting		
Indicator	Estimate	SE	P-Value	Estimate	SE	P-VALUE	Estimate	SE	P-value
Intercept	-1.232	0.686	0.0726	3.186	2.054	0.121	1.181	0.971	0.224
Child's age in moths									
0-11 months	reference								
12-23 months	-0.901	0.3017	0.0028	0.004	0.1583	0.9799	1.232	0.1529	$< .0001$
24^{+} months **Birth order**	0.602	0.299	0.0442	-0.225	0.1099	0.0411	-0.145	0.0960	0.1302
First	reference								
2-3	0.095	0.3533	0.7882	0.259	0.1307	0.0473	0.142	0.1214	0.2424
4-5	-0.380	0.362	0.2944	0.297	0.14	0.0341	0.130	0.1315	0.3236
6^{+}	0.975	0.4521	0.0311	1.081	0.1659	< 0.0001	0.5029	0.137	0.0002
Mother's age at the birth									
$21 >$	reference								
≤ 21 years	0.379	0.673	0.573	0.131	0.2641	0.6209	0.552	0.2129	0.0096
Mother's education level									
Secondary & higher	reference								
No education	-0.547	0.4822	0.2566	-2.334	0.3685	$< .0001$	-0.658	0.1857	0.0004
Primary	-0.3232	0.552	0.5582	-2.829	0.3816	$< .0001$	-0.902	0.2064	$< .0001$
Gender of the child									
Male Female	reference -0.3342	0.2452	0.1633	-0.482	0.0943	$< .0001$	-0.447	0.0833	$< .0001$
Birth weights									
$\geq 2500g$	reference								
$< 2500g$	1.4	0.4481	0.0018	1.151	0.2095	$< .0001$	0.580	0.2294	0.0.0115
Province/region									
Eastern Kigali	reference -0.259	0.4973	0.6028	0.2905	0.3128	0.3531	-0.159	0.212	0.3641
Southern	0.246	0.3746	0.5110	0.3516	0.222	0.1137	0.339	0.1491	0.023
Western	0.65	0.3756	0.836	0.0493	0.2137	0.8176	0.435	0.7142	0.0409
Northern	0.924	0.511	0.0701	0.414	0.2485	0.0954	-0.150	0.1618	0.3549
Knowledge on nutrition									
No Yes	reference -0.141	0.2543	0.58	0.348	0.1096	0.0015	0.259	0.0916	0.0047
Wealth index									
Rich	reference								
Middle	0.562	0.3548	0.1129	0.144	0.1335	0.2821	0.194	0.1105	0.0784
Poor	1.303	0.5574	0.0194	0.228	0.2056	0.2681	0.434	0.1636	0.0079
Multiple birth									
Singleton	reference								
First multiple	0.027	1.0951	0.9804	1.346	0.364	0.0002	0.409	0.4299	0.3413
Second multiple and more	0.043	1.3326	0.9743	0.445	0.4542	0.3272	0.138	0.5582	0.8046
Incident of anemia									
No anemic	reference								
Anemic	-0.494	0.2591	0.0568	-0.370	0.0992	0.0002	-0.093	0.0882	0.2922
Body mass index									
$BMI \geq 18.5$	reference								
$BMI < 18.5$	1.117	0.3993	0.0052	1.130	0.1914	$< .0001$	0.131	0.1961	0.5028
Incident of the fever									
Had fever last two weeks	reference								
No fever **Source of drinking water**	0.567	0.2795	0.0427	0.511	0.119	$< .0001$	0.010	0.1134	0.9273
Others/yard	reference								
Piped into dwelling/yard	-2.041	0.718	0.0045	-0.157	0.4249	0.7115	0.436	0.3186	0.1714
Public tap	-1.352	0.4005	0.0007	-0.288	0.1588	0.0699	-0.019	0.1288	0.8802
Protected spring/well	-0.462	0.3185	0.1472	0.081	0.1246	0.514	-0.010	0.1038	0.336

Table 5.4. Variance components

Label	Estimate	SE.	P-value
Var(stunting)	0.3161	0.07284	< .0001
Var(underweight)	0.8065	0.3639	0.0133
Var(wasting	1.3923	0.1785	< .0001
Correlation between stunting and underweight	0.9665	0.0947	< .0001
Correlation between wasting and underweight	0.9903	0.1778	< .0001

5.5. Summary

This study used joint multivariate generalized linear mixed model to identify simultaneously the key determinants of malnutrition of the child under age five in Rwanda on three anthropometric indices: underweight, wasting and stunting. These three response variables (underweight, stunting and wasting) could have been used separately but as the correlation between underweight and wasting and underweight and stunting is significant, it is better to use the joint model. If the correlation was not significant, we would have simply used GLMM. However, joint model has a number of advantages over separate fitting such as better control of the type I error rates in multiple tests, possible gains in efficiency in the parameter estimates and the ability to answer intrinsically multivariate questions (Kandala et al., 2011b; Verbeke and Molenberghs, 2003; Gueorguieva, 2001). This study measured simultaneously the determinants of underweight, stunting and wasting. Our findings revealed a positive correlation between underweight and wasting and underweight and stunting; this means that increasing height-for-age and height-for-weight increases the weight for height or decreasing height-for-age and height-for-weight also decreases the weight-for-age. In other words, reducing stunting has a positive consequence of reducing underweight.

The findings of this study revealed that the age of child, gender of child, birth weight, birth order, fever, mother's education level, mother's age at the birth, body mass index of the mother, anemia, knowledge on nutrition by mother, province, source of drinking water, multiple birth and wealth index of the household are the key determinants of malnutrition of children under five years in Rwanda. This research revealed that stunting and underweight are lower in female children compared to male children. This finding is consistent with other authors (Habyarimana et al., 2014; Kandala et al., 2011a). It also revealed that the nutrition status of the mother affects the nutrition status of the child, where a thin mother is more likely to deliver a wasted or underweight child; this finding is in line with Das and Rahman (2011). Mother's knowledge on nutrition is a very import factor of nutrition of the child. Some variables such as birth weights of the child and birth order significantly affect all three anthropometric indices. It was also found that malnutrition decreases with increasing the mother's level of education especially in the case of stunting and underweight; these results are in line with Habyarimana et al. (2014) and Kandala et al. (2011a). Improving the access (distance traveled to feature) to potable water may help to reduce wasting; sensitization to the population about nutrition may reduce stunting and underweight. Also improving education level of women may reduce underweight and stunting; to continue sensitizing women to get pregnant when they are mature enough (aged 21 years or more) may also contribute to reduce stunting; sensitizing how to take care of children may reduce not only stunting but also underweight and wasting. The birth order significantly affects all three anthropometric indices, where malnutrition increases with birth order; maybe improving the existing planning policy about limitations of birth might reduce the negative effect on nutrition. But the spatial variability was not considered. Therefore, in the next chapter we use spatial multivariate joint model to account for spatial variability that might exist between households.

CHAPTER 6

Accounting for spatial variability in modelling malnutrition

Introduction

In previous chapters the data was analyzed using classical binary and ordinal logistic regression models, binary and ordinal survey logistic (proportional model with sampling weight), generalized linear mixed model, and multivariate joint model (underweight, stunting and wasting) under GLMM; however none of these included the spatial random variability effects. Therefore, this chapter extends chapter 5 to include spatial variability and to produce the smooth maps of joint malnutrition prevalence of stunting, wasting and underweight.

6.1. Model overview

Spatial statistics is mainly divided into three methods such as point pattern analysis, methods of lattice data and geostatistics (Schaberger and Gotway, 2005; Cressie and Cassie, 1993).

We consider the basic terminology first.

Variogram: A variogram $2\gamma(h)$ represents the average variance between observations separated by the distance h, $\gamma(h)$ is the semivariogram. A variogram plays an important role in the description and interpretation of the structure of the spatial variability. It is given by (Journel, 1978) as

$$2\gamma(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \{Z(s_i) - Z(s_i + h)\}^2 \tag{6.1}$$

where $Z(s_i)$ is the measurement at location s_i with $N(h)$ the number of sampled points of distance(lag) length h.

The non-Gaussian spatial problems can be analyzed in the context of generalized linear mixed models, where the specification of the likelihood of the random variable is required. The spatial process can be incorporated as $y(s_i|\alpha)$, and this assumed to be conditionally independent for any location s_i with the conditional mean $\mu(s_i) =$

$E[Y(s_i|\alpha)]$, where the parameter α is used to define the distribution of s. Therefore, the spatial correlated random effect is incorporated into the linear predictor as

$$g(\mu(s_i)) = \eta(S_i) = X'(S_i) + w(s_i) \tag{6.2}$$

where X and W are the design matrices. The random effect at location (s_i), $\alpha \sim N\left(0, \sum_{\alpha}(\theta)\right)$ and $\epsilon \sim N\left(0, \sigma_{\epsilon}^2 I\right)$, where the spatial correlation is parameterized by θ in $\sum_{\alpha}(\theta)$ (Schaberger and Gotway, 2005).

There are three major functions used to describe the spatial correlation in Geostatistics. These functions are the correlogram, the covariance and semivariogram. A variogram represents the structural and random aspects of the data. A variogram has a number of properties to satisfy for instance, assuming that the mean is constant, and define

$$var[Z(s_1) - Z(s_2)] = 2\gamma(s_1 - s_2)$$

; the variance of s_1 and s_2 is through their difference. A process that satisfies this property is called intrinsically stationary. If the semivariogram depends only its vector argument h through its length $||h||$, then the process is called isotropic. A process that is both intrinsically and isotropic is known as homogeneous. Isotropic processes are more convenient to deal with because there are a number of commonly used parametric forms of semivariogram. semivariograms γ increase monotonically to reach a peak (sill) at range (r) with spatial variance called partial-sill σ_1^2 and non random variance starting at $(h > 0)$ referred to as nugget (c_1). Some of the examples are:

$$\text{Spherical} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 \left[\frac{3}{2} \frac{|h|}{r} \frac{1}{2} \left(\frac{|h|}{r} \right)^3 \right], & \text{if } 0 < |h| \leq r \\ c_1 + \sigma_1^2, & \text{if } |h| \geq r \end{cases} \tag{6.3}$$

This is valid in $(\mathbb{R}^d, d = 1, 2, 3)$. The spherical function reaches the sill at $|h| = r$. The model looks nearly linear at small lags. The spherical model is a commonly used variogram structure in practice, particularly for modelling spatial correlation that decreases linearly with the separation distance.

$$\text{Exponential} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 \left[1 - \exp(-\frac{|h|}{r}) \right], & \text{if } |h| > 0 \end{cases} \tag{6.4}$$

This function is valid for all dimension. However, it reaches the sill asymptotically when $|h| \to \infty$.

$$\text{Gaussian} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 \left[1 - \exp(-\frac{|h|^2}{r}) \right], & \text{if } |h| > 0 \end{cases} \tag{6.5}$$

This is valid for all dimension; the Gaussian model reaches the sill asymptotically. It is used when the data exhibit strong continuity at short lag distance, in other words when spatial correlation between two nearby points is very high. The Gaussian semivariogram is S-shaped, much like one-half of the Gaussian distribution.

$$\text{Exponential power form} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 \left[1 - \exp(-\left|\frac{h}{r}\right|^q) \right], & \text{if } |h| > 0 \end{cases} \tag{6.6}$$

where $0 \leq q \leq 2$. Note that the Gaussian and exponential forms are special cases of the exponential power form, in other words model (6.6) generalizes models (6.4) and (6.5).

$$\text{Cubic} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 \left[7\left(\frac{|h|}{r}\right)^2 - \frac{35}{4}\left(\frac{|h|}{r}\right)^3 + \frac{7}{2}\left(\frac{|h|}{r}\right)^5 - \frac{3}{4}\left(\frac{|h|}{r}\right)^7 \right], & \text{if } 0 < |h| \leq r \\ c_1 + \sigma_1^2, & \text{if } |h| \geq r \end{cases} \tag{6.7}$$

$$\text{Power law} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 h^p, & \text{if } |h| > 0 \end{cases} \tag{6.8}$$

This is valid for all dimensions. But the power model does not reach the sill. Any power between 0 and 2 may be used to constract a valid power variogram model. The power model is only appropriate if there is long-range correlation or if sample were not collected at a sufficiently large distance to reach the point where pairs of points are uncorrelated.

$$\text{Pentaspherical} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 \left[\frac{15}{8}\frac{|h|}{r} - \frac{5}{4}\left(\frac{|h|}{r}\right)^3 + \frac{3}{8}\left(\frac{|h|}{r}\right)^5 \right], & \text{if } 0 < |h| \leq r \\ c_1 + \sigma_1^2, & \text{if } |h| \geq r \end{cases} \tag{6.9}$$

The pentaspherical semivariogram behaves like cubic and spherical models in that $\gamma(h)$ increases with the distance until reaches the sill value $c_1 + \sigma_1^2$ at the distance h equals to the model range r.

$$\text{Sine hole effect} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 \left[1 - r\frac{sin(|h|/r)}{|h|} \right], & \text{if } |h| > 0 \end{cases} \tag{6.10}$$

The wave or hole-effect mode is generally used when there is some periodicity in the data resulting in a hole-effect. The range in the hole-effect is the shortest distance at which the semivariogram equals c_1. This will occur on the initial rise in the variogram function. Because of periodicity, this model contains both positive and negative correlation.

$$\text{Mathéron} : \gamma(h) = \begin{cases} 0, & \text{if } |h| = 0 \\ c_1 + \sigma_1^2 \left[1 - \frac{2}{\Gamma(\nu)}\left(\frac{|h|\sqrt{\nu}}{r}\right)^{\nu} K_\nu 2\left(\frac{|h|\sqrt{\nu}}{r}\right) \right], & \text{if } |h| > 0, \nu > 0 \end{cases} \tag{6.11}$$

K_ν is the modified bessel function of order ν. This model is a highly flexible model around nugget effect, and is the best when modelling complicated behaviour near the nugget effect (Handcock and Stein, 1993).

6.2. Valid covariance and semivariogram functions

Consider isotropic models for the covariance function and semivariogram of a spatial process. Let $C(h)$ be isotropic covariance function of the second order stationary field and $\gamma(h)$ the isotropic semivariogram of a second order or intrinsically stationary field. A valid covariance $C(h)$ is a positive-definite function, that is

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j C(s_i - s_j) \geq 0 \tag{6.12}$$

for any finite configurations of spatial locations $\{s_i : i = 1, 2, ..., m\}$ and all real numbers $\{a_i : i = 1, 2, ..., m\}$. Based on Bochner's theorem, the equation (6.12) means that the covariance function $C(h)$ can be represented in spectral form as follows

$$\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} exp\left(i\omega' h\right) dS(\omega) \tag{6.13}$$

where $S(d\omega) = s(\omega)d\omega$ the integral is over $\mathbb{R}^d$ and S is a positive bounded spectral measure. In isotropic case, the spectral representation of the covariance function in $\mathbb{R}^d$ becomes (Cressie and Cassie, 1993; Schaberger and Gotway, 2005)

$$C(h) = \int_{0}^{\infty} \Omega_d(h_\omega) dH(\omega) \tag{6.14}$$

with

$$\Omega_d(t) = \left(\frac{2}{t}\right)^{\nu} \Gamma(d/2) J_\nu(t)$$

and Ω_d is commonly known as basis function of the covariance model in $\mathbb{R}^d$, where $\nu = \frac{d}{2} - 1$, J_ν is the Bessel function of the first kind of order ν and H is a non-decreasing function on the interval $[0, \infty)$ with $\int_0^\infty dH(\omega) < \infty$ (Schaberger and Gotway, 2005). In addition, the model validity can also be defined based on the variogram theory as in Cressie and Cassie (1993) and Schaberger and Gotway (2005) by

$$2 \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \gamma(s_i - s_j) \leq 0 \tag{6.15}$$

for any finite configurations of spatial locations $\{s_i : i = 1, 2, ..., m\}$ and all real numbers $\{a_i : i = 1, 2, ..., m\}$ and satisfying $\Sigma_{i=1}^m a_i = 0$. A valid semivariogram as in the case of covariance has also a spectral representation given by

$$\gamma(h) = \frac{1}{2} \int_{0}^{\infty} \omega^{-2} \left(1 - \Omega_d(\omega h)\right) dH(\omega) \tag{6.16}$$

with $\int_0^\infty \left(1 + \omega^2\right)^{-1} dH(\omega) < \infty$. A necessary condition for $\gamma(h)$ to be a valid semi-variogram is $2\gamma(h)$ grows more slowly than $\|h\|^2$ which is often referred to as the intrinsic hypothesis.

6.3. Estimating semivariogram functions

The main idea is to find a valid variogram that, as a measure of the spatial dependence, is closest to the spatial dependence present in the data $Z = (Z(s_1), Z(s_2), ..., Z(s_n))'$

Fitting semivariogram and covariance models

After estimating the empirical semivariogram, the next step is to fit the theoretical model (for instance, spherical, Gaussian, exponential, etc.) to the empirical semivariogram. There are three main approaches for estimating the parameters of the semivariogram model: Visual, (weighted) least squares, and likelihood methods. The estimation of semivariograms is mainly based on the method of moments known as *Matheron's* estimator, the *Cressie-Hawkins* robust estimator, estimators based on order statistics and quantiles. However, the simplest method is Matheron's estimator also known as classical estimator; it was proposed by Matheron (1962). Let $Z(s_1), ..., Z(s_n)$ be a set of spatial data, one could plot the squared differences $(Z(s_i) - Z(s_j))^2$ against the lag distance h. This graph is referred to as the empirical semivariogram cloud. However, $(Z(s_i) - Z(s_j))^2$ estimates unbiasedly the variogram at lag $h = s_i - s_j$ if the mean of the random field is constant. A more useful estimator is obtained by summarizing the squared differences. The semivariogram estimator which averages the squared differences of point that are distance $s_i - s_j = h$ part is generally known as classical or Matheron estimator.

$$\hat{\gamma} = \frac{1}{2|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2 \tag{6.17}$$

where the set $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = \|h\|; i, j = 1, 2, ..., n\}$ consists of all location pairs (s_i, s_j) separated by the distance $\|h\|$ and $|N(h)|$ is the number of distinct pairs in N(h). In the case of sparse data, it is usually recommended to group the distances into bins according to chosen distance lags and lag tolerances. Therefore, the corresponding averaged $\frac{1}{2}(Z(s_i) - Z(s_j))^2$ in each bin is taken as the semivariogram estimate for that distance lag. The lag tolerance must be chosen in such a way that adequate spatial resolution and stability in the smoothed estimator are retained. Journel (1978) proposed choosing lag tolerance such that at least 30 locations-to-location pairs fall within each bin.

Cressie and Hawkins (1980) proposed an estimator that alleviates the negative impact of outlying observations by eliminating squared differences from the evaluations. This estimator is commonly known as the robust semivariogram estimator or the Cressie-Hawkins (CH) estimator formulated in Cressie and Cassie (1993) and Schaberger and Gotway (2005) as

$$\overline{\gamma}(h) = \frac{1}{2|N(h)|} \frac{\left(\sum_{N(h)} |Z(s_i) - Z(s_j)|^{\frac{1}{2}} \right)^4}{0.47 + \frac{0.494}{|N(h)|}} \tag{6.18}$$

This estimator was derived under the assumption that the differences $Z(s_i) - Z(s_j)$ are normally distributed for all station pairs (s_i, s_j) and the denominator in equation (6.18) is the bias correction (Genton, 1998). However, this estimator is not a resistant estimator, since is not stable under gross contamination of the data (Schaberger and Gotway, 2005). The CH and Matheron estimators have unbounded influence functions and a breakdown point of 0%. The influential function of an estimator measures the effect of infinitesimal contamination of the data on the statistical properties of the estimator (Hampel et al., 1986) and the breakdown point is the percentage of the data that can be replaced by arbitrary values without explosion of the estimator. The median absolute deviation (MAD), is an estimator of scale with 50% breakdown point and a smooth influence function. For a set of numbers $\{x_1, x_2, ..., x_n\}$, the MAD is given by

$$MAD = d \text{ median}_i \left(\|x_i - \text{median}_j(x_j)\| \right) \tag{6.19}$$

where $\text{median}_i(x_i)$ is the median of the x_i and d is chosen to produce approximate unbiasedness and consistency. Rousseeuw and Croux (1993) proposed a robust estimator of scale which also has a 50% breakdown point and smooth influence function. Their Q_n estimator is given by the k^{th} order statistic of the $n(n-1)/2$ inter-point distances. Let $h = n/2 + 1$ and $k = \begin{pmatrix} h \\ 2 \end{pmatrix}$. Then $Q_n = c\{\|x_i - x_j\| : i < j\}_{(k)}$. This method has positive small-sample bias that can be corrected (Croux and Rousseeuw, 1992). Genton (1998, 2001) considers the modification that leads from equations (6.17) to(6.18) not sufficient to impart robustness and develops a robust estimator of the semivariogram based on Q_n. If the spatial data $\{Z(s_1), ..., Z(s_n)\}$are observed, let $N(h)$ denote pairwise difference $T_i = Z(s_i) - Z(s_i + h), i = 1, 2, ..., n(n-1)/2$.

Next, calculate $Q\|N(h)\|$ for the T_i and return as the semivariogram estimator at lag h

$$\overline{\gamma}(h) = \frac{1}{2}Q^2_{|N(h)|} \tag{6.20}$$

Since Q_n has 50% breakdown points, $\overline{\gamma}(h)$ 50% also has breakdown points in terms of process of differences in T_i, but not necessarily in terms of the $Z(s_i)$. Genton (2001) established through simulation that roughly equation (6.20) will be resistant to 30% of outliers among the $Z(s_i)$.

Another method used for robustness of the empirical semivariogram estimator is to consider the quantiles of the distribution of $\{Z(s_i)-Z(s_j)\}^2$ or $|Z(s_i)-Z(s_j)|$ instead of considering the arithmetic averages as in equation (6.17) and (6.18) (Schaberger and Gotway, 2005). If $[Z(s_i), Z(s_i+h)]'$ are bivariate Gaussian with common mean, therefore

$$\frac{1}{2}\left(Z(s_i) - Z(s_i+h)\right)^2 \sim \gamma(h)\chi^2_1$$
$$\frac{1}{2}\|Z(s_i) - Z(s_i+h)\| \sim \sqrt{\frac{1}{2}\gamma(h)}|U|,\ U \sim G(0,1)$$

Let $q^{(p)}_{|N(h)|}$ denote the p^{th} quantile. Therefore,

$$\hat{\gamma}_p(h) = q^{(p)}_{|N(h)|}\{\frac{1}{2}\left(Z(s_i) - Z(s_i+h)\right)^2\} \tag{6.21}$$

estimates $\gamma(h) \times \chi^2_{p,1}$. If $p = 0.5$, then equation (6.21) reduces to median estimator as:

$$\hat{\gamma}(h) = \frac{1}{2}\text{median}_{|N(h)|}\{\frac{1}{2}\left(Z(s_i) - Z(s_i+h)\right)^2\}/0.455 \tag{6.22}$$

$$= \frac{1}{2}\left(\text{median}_{|N(h)|}\{\frac{1}{2}\left(Z(s_i) - Z(s_i+h)\right)^{\frac{1}{2}}\}\right)^4/0.455 \tag{6.23}$$

then $q^{(p)}_{|N(h)|}$ reduces to median based estimator. The empirical variogram provides a description of how the data are related with distance. The variogram function $\gamma(h)$ was originally defined by Matheron (1963) as a half of the average squared difference between points separated by a distance h. The semivariogram is $1/2\gamma(h)$.

The empirical semivariogram $\hat{\gamma}(h)$ is unbiased estimator of $\gamma(h)$, however, it only provides estimates at a finite set of lags or lag classes. In order to obtain estimates of $\gamma(h)$ at any arbitrary lag, the empirical semivariogram must be smoothed. A non parametric kernel smoother will not suffice because it is not guaranteed that the resulting fit is a conditionally negative-definite function. The common approach is to fit the parametric semivariogram models or to apply non parametric semivariogram. Although fitting a parametric semivariogram model to empirical semivariogram by the least squares method is by far the most common approach, it is not the only parametric technique. Maximum likelihood and restricted (residual) maximum likelihood (REML) estimation use observed data directly, usually assuming a Gaussian random field (Schaberger and Gotway, 2005). Other estimating function based methods such as generalized estimating equations (GEE) and composite likelihood also utilize pseudo-data. No single method can claim uniform superiority. To distinguish the empirical semivariogram $\gamma(h)$ from

the semivariogram model being fit, we introduce the notation $\gamma(h, \theta)$ for the latter. The vector θ contains all unknown parameters to be estimated from the data and its estimate $\hat{\gamma}(H)$

Least square estimation

Suppose that the semivariogram is estimated by $\gamma(h)$ at finite set of values of h, and wish to fit model specified by parametric function $\gamma(h, \theta)$ with respect to a finite parameter θ. Let us assume that the method of moment (MoM) estimator $\hat{\gamma}(h)$ has been used and let $\hat{\gamma}$ denote the vector of estimates of $\gamma(\theta)$, the vector model values at the same vector of h values. Generally, there are three common approaches of least squares estimator in literature known as *Ordinary least squares (OLS)*: in this approach θ can be minimized using $(\hat{\gamma} - \gamma(\theta))' (\hat{\gamma} - \gamma(\theta))$.

The second approach is *Generalized least squares or GLS*, in this approach θ can be minimized using $(\hat{\gamma} - \gamma(\theta))' V(\theta)^{-1} (\hat{\gamma} - \gamma(\theta))$ where $V(\theta)$ denotes the covariance matrix of $\hat{\gamma}$. This estimator depends on an unknown θ because the problem is nonlinear. The third approach is *Weighted least squares or WLS*. In this approach θ can be minimized using the following expression $(\hat{\gamma} - \gamma(\theta))' W(\theta) (\hat{\gamma} - \gamma(\theta))$, where $W(\theta)$ is a diagonal matrix whose diagonal entries are the variances of the entries of $\hat{\gamma}$. Therefore weighted least squares allows for the variance of $\hat{\gamma}$ but not the covariance, while GLS allows for both. Also the weights in matrix $W(\theta)$ may be proportional to $|N(h)|$ or inversely proportional to approximate variance of $\hat{\gamma}$ for more details see (Cressie, 1985). OLS is the most convenient estimator to use, it is immediately implementable by nonlinear least squares procedure, while WLS and GLS need specification of the matrices $W(\theta)$ and $V(\theta)$. However, in general, OLS, WLS and GLS are in increasing order of efficiency (Cressie and Cassie, 1993). Based on Gaussian process we get the following expressions (Cressie and Cassie, 1993):

$$E\left([Z(s_i+h)-Z(s_i)]^2\right)=2\gamma(h), \tag{6.24}$$

$$var\left([Z(s_i+h)-Z(s_i)]^2\right)=2\left[2\gamma(h)\right]^2, \tag{6.25}$$

$$corr\left([Z(s_1+h)-Z(s_1)]^2,[Z(s_2+h)-Z(s_2)]^2\right)= \tag{6.26}$$

$$\left[corr\left(\{Z(s_1+h)-Z(s_1)\}^2,\{Z(s_2+h)-Z(s_2)\}^2\right)\right]^2= \tag{6.27}$$
$$\frac{\left[\gamma\left(s_1-s_2+h_1\right)+\gamma\left(s_1-s_2-h_2\right)-\gamma\left(s_1-s_2+h_1-h_2\right)-\gamma\left(s_1-s_2\right)\right]^2}{2\gamma(h_1)2\gamma(h_2)}$$

The equation (6.27) may be used to find the matrices $W(\theta)$ and $V(\theta)$. Therefore, generalized least squares can be used in principle, however it is complicated to implement, for instance it is not guaranteed that the resulting minimization has a unique solution. Schaberger and Gotway (2005) proposed the following weighted least squares criterion for solving complicated

$$\sum_j |N(h_j)|\left(\frac{\hat{\gamma}(h_j)}{\hat{\gamma}(h_j,\theta)}-1\right)^2 \tag{6.28}$$

equation (6.28) can be derived as the WLS solution under the approximation

$$var\left(\hat{\gamma}(h)\right)\approx\frac{8\gamma^2(h)}{|N(h)|} \tag{6.29}$$

This follows from equation (6.25) if we assume that the individual $Z(s_i)-Z(s_j)$ terms are independent. This assumption is not exactly satisfied but may be a reasonable approximation if the pairs (s_i, s_j) lying in $N(h)$ are widely spread over the sampling space. The WLS in equation (6.28) is not difficult to implement than OLS and is more efficient.

Maximum likelihood

Maximum likelihood estimator for spatial model is only developed for the Gaussian case (Mardia and Marshall, 1984), and the Gaussian assumption for spatial model is given by $Z(s)\sim N\left(X(s)\beta,\Sigma(\theta)\right)$, where $\sum=\alpha V(\theta)$, α is a scale parameter and

$V(\theta)$ is a vector of standard covariance and θ is an unknown parameter. Maximum likelihood (ML) is a simultaneous estimation of mean and covariance parameters, where the ML estimates are the simultaneous solution to the problem of minimizing the negative of twice the Gaussian log likelihood

$$\varphi(\beta;\theta;Z(s)) = ln\left(|\Sigma(\theta)|\right) + nln(2\pi) + (Z(s)-X(s)\beta)'\,\Sigma(\theta)^{-1}\,(Z(s)-X(s)\beta) \tag{6.30}$$

To profile β, differentiate equation (6.30) with respect to β and solve. The result is the GLS estimator

$$\hat{\beta} = \left(X(s)'\Sigma(\theta)^{-1}X(s)\right)^{-1} X(s)'\Sigma(\theta)^{-1}Z(s) \tag{6.31}$$

Equation (6.31) in (6.30) yields an objective function for minimization profiled for β given by

$$\varphi_{\beta}(\theta;Z(s)) = ln\left(\left|\sigma^2\Sigma(\theta^*)\right|\right) + nln(2\pi) + \sigma^{-2}r'\Sigma(\theta^*)^{-1}r \tag{6.32}$$

where

$$r = Z(s) - \left(X(s)'\Sigma(\theta)^{-1}X(s)\right)^{-1} X(s)'\Sigma(\theta)^{-1}Z(s)$$

where r is the GLS residual. σ^2 can be profiled from the objective function (6.32), note that its MLE is

$$\hat{\sigma}^2_{ml} = \frac{1}{n}r'\Sigma(\theta^*)^{-1}r$$

substituting again yields the negative of twice the profiled log likelihood as

$$\varphi_{\beta,\sigma}(\theta^*;Z(s)) = ln\left(|\Sigma(\theta^*)|\right) + n\left(ln(2\pi)-1\right) \tag{6.33}$$

Therefore minimizing equation (6.33) is an optimization problem with only $(q-1)$ parameters. Upon convergence we obtain $\hat{\theta}_{ml}$ from $\hat{\sigma}^2_{ml}$ and $\hat{\sigma}^*_{ml}$,and $\hat{\beta}_{ml}$ by evaluating equation (6.34) at the maximum likelihood estimates $\hat{\theta}_{ml}$ of θ:

$$\hat{\beta} = \left(X(s)'\Sigma(\hat{\theta}_{ml})^{-1}X(s)\right)^{-1} X(s)^T\Sigma(\hat{\theta}_{ml})^{-1}Z(s) \tag{6.34}$$

One of the advantages of likelihood estimation is the ability to estimate the variance-covariance matrix of the parameter estimates based on the observed or expected information matrix.

Restricted maximum likelihood

Restricted maximum likelihood (REML) estimation is a method employed to estimate variance-covariance parameters from data that follow a Gaussian linear model. Restricted maximum likelihood estimates are frequently preferred over maximum likelihood estimates (MLE) since the latter exhibit greater negative bias for estimates of covariance parameters. In the case of the spatial model

$$Z(s) \sim N\left(X(s)\beta, \Sigma(\theta)\right)$$

the REML adjustment consists of performing maximum likelihood estimation not for $Z(s)$, but for $KZ(s)$,where the $((n-k)\times n)$ matrix K is chosen so that $E\left[KZ(s)\right] = 0$ and the rank of $K = n-k$. Because of these properties the matrix K is called a matrix of error contrast. An objective function about θ is given by

$$\varphi_R\left(\theta; KZ(s)\right) = ln\{|K\Sigma(\theta)K'|\} + (n-k)ln(2\pi) + Z(s)'K'\left(K\Sigma(\theta)K'\right)^{-1}KZ(s) \tag{6.35}$$

and

$$\hat{\beta}_{reml} = \left(X'\Sigma(\hat{\theta}_{reml})^{-1}X\right)X'\Sigma(\hat{\theta}_{reml})^{-1}Z(s) \tag{6.36}$$

If $E[KZ(s)] = 0$, then $KX(s) = 0$, in addition if $\Sigma(\theta)$ is positive definite, then equation (6.35) can be reduced (Searle et al., 1992) to

$$K'\left(K\Sigma(\theta)K'\right)^{-1}K = \Sigma(\theta)^{-1} \tag{6.37}$$

where $\Sigma(\theta) = \left(X(s)'\Sigma(\theta)^{-1}X(s)\right)^{-1}$. This is identity and $\Sigma X(s)'\Sigma(\theta)^{-1}Z(s) = \hat{\beta}$ yields

$$Z(s)'K'\left(K\Sigma(\theta)K'\right)KZ(s) = r'\Sigma(\theta)^{-1}r$$

where

$$r = Z(s) - \left(X(s)'\sum(\theta)^{-1}X(s)\right)^{-1}X(s)'\sum(\theta)^{-1}Z(s)$$

is the GLS residual. Harville (1974) based on the following identities

$$KK' = I - X(s)\left(X(s)'X(s)\right)^{-1}X(s)'$$

and

$$KK' = I$$

reduced the minus twice the log likelihood of $KZ(s)$ to

$$\begin{aligned}\varphi_R(\theta; KZ(s)) = & \quad ln(|\Sigma(\theta)) + ln\{|K\Sigma(\theta)K'|\} + (n-k)ln(2\pi) \\ & + \quad Z(s)'K'(K\Sigma(\theta)K')^{-1}KZ(s) \\ & - \quad ln(|X(s)'X(s)|) - r'\Sigma(\theta)^{-1}r + (n-k)ln(2\pi) \end{aligned} \tag{6.38}$$

He also pointed out that $(n-k) \times n$ matrices whose rows are linearly independent rows of $I - X(s)(X(s)'X(s))^{-1}X(s)'$ will lead to REML objective function that differ by a constant amount and this amount does not depend on θ or β. The obvious choice as a REML objective function for minimization is

$$\begin{aligned}\varphi_R(\theta; KZ(s)) = & \quad ln(|\Sigma(\theta)) + ln\{|K\Sigma(\theta)K'|\} + (n-k)ln(2\pi) \\ & + \quad Z(s)'K'(K\Sigma(\theta)K')^{-1}KZ(s) + r'\Sigma(\theta)^{-1}r + (n-k)ln(2\pi)\end{aligned} \tag{6.39}$$

In this form the minus twice the REML log likelihood differ by the terms $ln(|X(s)'\Sigma(\theta)^{-1}X(s)|)$ and $kln(2\pi)$. As with ML estimation, a scale parameter can be profiled from $\Sigma(\theta)$ and the REML estimator of this parameter is given by

$$\hat{\sigma}^2_{reml} = \frac{1}{n-k} r'\Sigma(\theta^*)^{-1}r$$

and upon substitution one obtains minus twice the profile REML log likelihood as follows

$$\varphi_R(\theta^*; KZ(s)) = ln(|\Sigma(\theta^*)) + ln\{|K\Sigma(\theta^*)K'X(s)|\} + (n-k)ln(\hat{\sigma}^2) + (n-k)(ln(2\pi) - 1) \tag{6.40}$$

Wolfinger et al. (1994) give expression for the gradient and Hessian of the REML log likelihood with and without profiling of σ^2.

Minimum norm quadratic estimation

Minimum norm quadratic (MINQ) estimation was developed by Rao (1979) for the spacial case where the variance matrix of the data is linear in its parameters and is given by

$$\Sigma(\theta) = \theta_1\Sigma_2 + \theta_1\Sigma_1 + ... + \theta_m\Sigma_m \tag{6.41}$$

This is used when finding an estimator of θ_j between those that can be written as $\hat{\theta}_j = W'F_jW$, with $W = A'Z$ (a vector of orthogonal contrast to X). The minimum norm estimator is obtained by minimizing $E\left(\hat{\theta}_j - \theta_j\right)$. Generally it is subjected to unbiasedness or invariance restrictions. The formula of MINQ estimator in spatial setting was given by Kitanidis (1985), where the data are sampled from a random process in $\mathbb{R}^d$, but he used the mean squared-error as norm. The minimum norm quadratic approach is particularly suitable for variance component model, however, in spatial setting $\sum(\theta)$ might be a nonlinear function of the small scale variation parameter θ

The advantage of this method over maximum likelihood estimator or restricted maximum likelihood estimator and WLS procedure is that for a fixed α it is a linear procedure and for this reason it does not need any iterated procedure.

The spatial autocorrelation measurements

Moran's I: Moran's I coefficient of autocorrelation is similar to Pearson's correlation coefficient, and quantifies the similarity of outcome variable among areas that are defined as spatially related (Moran, 1950; Pfeiffer et al., 2008) and is given by

$$I = \frac{n \sum_i \sum_j w_{ij} \left(Z_i - \bar{Z}\right)\left(Z_j - \bar{Z}\right)}{\left(\sum_i \sum_j w_{ij}\right) \sum_k \left(Z_k - \bar{Z}\right)^2} \tag{6.42}$$

where Z_i could be the residual $(O_i - E_i)$ and w_{ij} is a measure of closeness of the areas i and j. Moran's I is approximately normally distributed and has an expected value of $\frac{-1}{(N-1)}$, where N equals the number of area units within a study region. Moran's I generally lies between $+1$ and -1, Moran's I is not bound by these limits unlike Pearson's correlation coefficient. A Moran's I of zero indicates the null hypothesis of no clustering, whereas a positive Moran's I indicates positive spatial autocorrelation (this means clustering of areas of similar attribute values), while a negative coefficient indicates a negative spatial autocorrelation (this means that neighbouring areas tend to have dissimilar attribute values).

Geary's C

Geary's contiguity ratio commonly known as Geary's C is another weighted estimated of spatial autocorrelation (Geary, 1954; Pfeiffer et al., 2008), whereas Moran's I considers similarity between pairs of regions. The Geary's range from 0 indicating perfect positive spatial autocorrelation and 2 indicating perfect negative spatial autocorrelation for any pair of regions. Geary's C statistic is given by

$$C = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(y_i - y_{j=1}\right)^2}{2\left(\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2\right)\left(\sum_{j=1}^{n}\sum_{j=1}^{n} w_{ij}\right)} \tag{6.43}$$

where n is the number of polygons in the study area, w_{ij} is the $(i,j)^{th}$ value of the spatial proximity matrix, y_i the attribute under investigation, and $\bar{y}$ is the mean of the attribute under investigation.

Geary's approach considers similarities between the pairs of regions and C varies between 0 (highest value of positive autocorrelation) and 2 (strong negative autocorrelation. Moran's I is a more global measurement and sensitive to extreme values, but Geary's C is more sensitive to differences in small neighbourhoods. Generally, Moran's I and Geary's C result in similar conclusions. However, Moran's I is preferred in most cases.

6.4. Application to the risk factors of malnutrition of children under five years

Let us consider y_{ijk} to be child nutrition status (1 in malnourished case and 0 in nourished case) of the anthropometric indicators, with $k = 1$ for wasting, $k = 2$ for underweight and $k = 3$ for a stunting for a child j, in district i, $i = 1, 2, , 30$. Let us consider that the observed outcomes arise from a trivariate Bernoulli distribution, with p_{ijk} as the probability of anthropometric indicator k occurring in child j in district i, therefore the outcome is modeled using GLMM with spatial random effect as follows

$$g(\mu_k) = X_{ijk}\beta_k + Z_{jk}\alpha_k \tag{6.44}$$

Where $k = 1, 2, 3$, β_k are vectors of fixed regression parameters, X_{ijk} and Z_{jk} are the design matrices and α_k are random spatial variation.

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \sim i.i.d.MVN\left(0, \sum\right) = MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sum_{11} & \sum_{12} & \sum_{13} \\ \sum'_{12} & \sum_{22} & \sum_{23} \\ \sum'_{13} & \sum'_{23} & \sum_{33} \end{bmatrix}\right), \tag{6.45}$$

Where the above equation (6.45) is the covariance matrices of the spatial effects, Σ_{11},Σ_{22}, Σ_{33} are the variance components of wasting, underweight and stunting status respectively and Σ_{12}, Σ_{13} and Σ_{23} are correlation components between wasting and underweight, wasting and stunting and underweight and stunting respectively.

6.4.1. Data analysis.

The data was analyzed by fitting generalized linear mixed using SAS 9.3 PROC GLIMMIX procedure, several covariance structures were considered such as SP(EXP) (Exponential), SP(EXPA) (Anisotropic Exponential), SP(EXPGA) (2D Exponential), Geometric Anisotropic, SP(GAU) (Gaussian), SP(GAUGA) (2D Gaussian, Geometrically Anisotropic) SP(SPH) (Spherical), SP(SPHGA) (2D Spherical, Geometrically Anisotropic), SP(LIN) (Linear), SP(LINL) (Linear Log), SP(Matern) (Matrn) and SP(MATHSW) (Matrn (Handcoks-Stein-Wallis)) and ArcGIS was used to produce smooth maps of malnutrition prevalence corresponding to each outcome variable.

6.4.2. Interpretation of the results.

The results from Figure 6.1 represent the scatter plot for malnutrition prevalence for joint distribution of stunting, underweight and wasting. As can be seen from the figure, the plot suggested that the distribution is not an indicative of uniform distribution. The distribution is an indication of random spread of the response. Classical representation of Gaussian semivariogram is presented together with the robust semivariogram Figure 6.2. Based on this graphical representation, the Gaussian structure was found to perform better than any other spatial structure considered. Therefore, the variogram analysis was performed based on Gaussian structure given in equation (6.5). It is observed from the figure, that the origin of Y-axis does not correspond to that of x-axis; this indicates the possible presence of nugget effect. The estimate of the range was estimated based on SP(GAU) spatial structure and is given by 1.5864 Table 6.2. In the Gaussian model, the variance parameter estimated by 0.7574 in Table 6.2 is known as the partial sill. The null hypothesis states that the spatial distribution of feature values is the result of random spatial process. The results from Moran's I (Z=-129.81 and p-value$<$ .0001) and Geary's C(Z-value=-9.32 and p-value$<$ .0001, indicate that the spatial distribution of feature values is not the result of random spatial processes. The Z values are negative for Moran's I and Geary's C; this is an indication that spatial distribution of higher values and low values in the dataset is more spatially dispersed than would be expected if underlying processes were random. However, if the z-value for Moran'I and Geary's C were positive with significant p-values, these would mean that the spatial distribution of high values or low values in the dataset is more spatially clustered than would be expected if underlying spatial processes were random.

The spatial autocorrelation was measured by this study which considered different child malnutrition factors such as gender of the child, birth weight, birth order, child age, child had fever in two weeks before the survey, diarrhea, mother's education level, mother's age at the birth, body mass index of the mother, anemia, mother's knowledge on nutrition, assistance at delivery, antenatal visits, region or province, source of drinking water, place of residence, toilet facilities, wealth index, access to toilet, the size of household and household wealth index. But in Table 6.1, any variable which is at least significant at one of the three anthropometric indicators is considered as a determinant of malnutrition and is reported. For the test of model fit, the AIC and -2log likelihood (deviance) are the same and smaller for Gaussian, Exponential power and Spherical than AIC and -2log likelihood of any other considered model. However, based also on graphical representation, Gaussian was found to be the best spatial covariance structure for this study.

Stunting: This study revealed that birth order, mother's age, mother's education, child's age, gender of the child, birth weight, province, mother' s knowledge on nutrition and wealth index are the determinants of stunting of children under five years in Rwanda. From Table 6.1, we observe that the age of a child significantly affects height-for-age of the child. A child aged between twelve months and twenty three months is 3.8768(p-value$<$ .0001) times more likely to be stunted than infant. But a child aged twenty three months or more was not significant as compared to infant. Birth order significantly affects height-for-age of the child Table 6.1. It was found that a child born at sixth order or more is 1.7092 (p-value=0.0003) times more likely to be stunted than infant. Mother's age at the birth significantly affects height-for-age of the child. A child born to mother aged less than twenty one years old is 1.8738 (p-value=0.0066) times more likely to be stunted than a child born to mother aged twenty one years old or more. Mother's education level significantly affects height-for-age of the child. A child born to mother with primary education or mother with secondary or higher level is 0.4781 (p-value $<$ .0001) or 0.3798 (p-value=0.0002) respectively times less likely to be in stunting status than a child born to mother with no formal education. This means that the stunting status decreases with increasing the mother's level of education.

Gender of child also affects height-for-age of the child Table 6.1. We observe from the same table that a male child is 1.6537 (p-value $<$ .0001) times more likely to be stunted than a female child. Birth weight significantly affects height-for-age of the child. A child born with low weight (weight $<$ 2500g) is 1.7212 (p-value=0.0271) times more likely to have stunting status than a child born with weight greater or equal to 2500g (weight $\geq$ 2500g). Province also affects height-for-age of the child. A child born in Southern or Eastern province is 1.7109 (p-value=0.0061) or 1.9484 (p-value=0.0147) respectively times more likely to be stunted than a child born in Kigali city. Mother's knowledge on nutrition significantly affects height-for-age of the child. A child born to mother with some knowledge on nutrition is 0.7240 (p-value=0.0009) less likely to be stunted than a child born to mother without knowledge on nutrition. Wealth index also significantly affects height-for-age of the child. A child born in rich family is 0.6460 (p-value=0.0143) times less likely to be stunted than a child born in poor family. The prevalence of stunting is higher in Northern province and lower in Kigali city Figure 6.3; this is consistent with other findings such as NISR et al. (2012).

Wasting: This study revealed that source of drinking water, fever, wealth index, birth weight, birth order and age of the child are the determinants of wasting of children under five years of age in Rwanda. The age of the child significantly affects height-for-weight of the child Table 6.1. A child aged between twelve months and twenty three months or twenty three months and more is 0.3712 (p-value=0.0011) or 1.800 (p-value=0.0499) respectively times more likely to be wasted than infant. But a child aged twenty three months or more was not significant as compared to infant. Birth order also significantly affects height-for-weight of the child. A child born at sixth order is 2.6406 (p-value=0.0317) times more likely to have wasting status than a child born at the first order.

Birth weight of the child also significantly affects height-for-weight of the child Table 6.1. A child born with low weight is 3.1018 (p-value=0.0035) times more likely to be wasted than a child born with weight $\geq$ 2500g. The wealth index is also significantly affecting height-for-weight of the child. A child born in rich family is 0.2658 (p-value=0.0181) times less likely to have wasting status than a child born in poor family. Body mass index of the mother is significantly affecting height-for-weight of the child. A child born to thin mother (BMI$<$ 18.5) is 3.8923 (p-value=0.0007) times more likely to be wasted than a child born to normal mother or obese (BMI $\geq$ 18.5). Source of drinking water significantly affects height-for-weight of the child Table 6.1. A child born in a family who use water piped in their dwelling or from public tap is 4.0390 (p-value=0.0045) or 7.3749 (p-value=0.0058) respectively times more likely to have wasting status than a child born in family where they use water from not piped and protected spring. The prevalence of wasting is higher in Western province and lower in Kigali city Figure 6.4. However, the current research was expecting the highest prevalence of wasting in Kigali city. This difference might be the effect of other covariates included in the model.

Underweight: This study revealed that birth order, mother's education, gender of the child, birth weight of the child, province, mother's knowledge on nutrition, multiple birth, anemia, body mass index of the mother and fever are the determinants of underweight of children under age five in Rwanda. Birth order significantly affects weight-for-age of the child. A child born at fourth to fifth order or sixth order or more is 1.3445 (p-value=0.0285) and 2.8405 (p-value $<$ .0001) respectively times more likely to be in underweight status than a child born at the first order. Mother's level of education significantly affects weight-for-age of the child Table 6.1. A child born to mother with primary education or mother with secondary or higher level is 0.1139 (p-value $<$ .0001) and 0.0954 (p-value $<$ .0001) respectively times less likely to be in underweight status than a child born to mother with no formal education. This means that the underweight status decreases with increasing the mother's level of education.

Gender of the child also significantly affects weight-for-age of the child Table 6.1. A male child is 1.5809 (p-value $< .0001$) times more likely to be in underweight status than a female child. Birth weight of the child is also significantly affecting weight-for-age of the child. A child born with low weight is 3.1018 (p-value $< .0001$) times more likely to be underweight than

a child born with weight $\geq$ 2500g. Province significantly affects weight-for-age for child. A child born in Western province is 0.6518 (p-value=0.0389) times less likely to be underweight than a child born in Kigali. Mother's knowledge on nutrition significantly affects weight-for-age of the child Table 6.1. A child born to mother with some knowledge on nutrition is 0.7160 (p-value=0.0014) less likely to be underweight than a child born to mother without knowledge on nutrition. Multiple births significantly affects weight-for-age of the child Table 6.1. A child born at the first multiple is 3.6988 (p-value=0.002) times more likely to be in underweight status than a singleton child. Anemia significantly affects the weight-for-age of the child Table 6.1. A child born to anemic mother is 1.4519 (p-value$<$ .0001) times more likely to be in underweight status than a child born to no anemic mother. Body mass index of the mother significantly affects weight-for-age of the child. A child born to thin mother (BMI$<$ 18.5) is 3.2197 (p-value $< .0001$) times more likely to be in wasting status than a child born to normal mother or obese (BMI $\geq$ 18.5). Fever also significantly affects weight-for-age of the child. A child who had no fever two days before the survey is 0.6083 (p-value$<$ .0001) times less likely to be underweight than a child who had fever two days before the survey. The prevalence of underweight is higher in Northern Province and Lower in Kigali city Figure 6.5. These findings are consistent with (NISR et al., 2012).

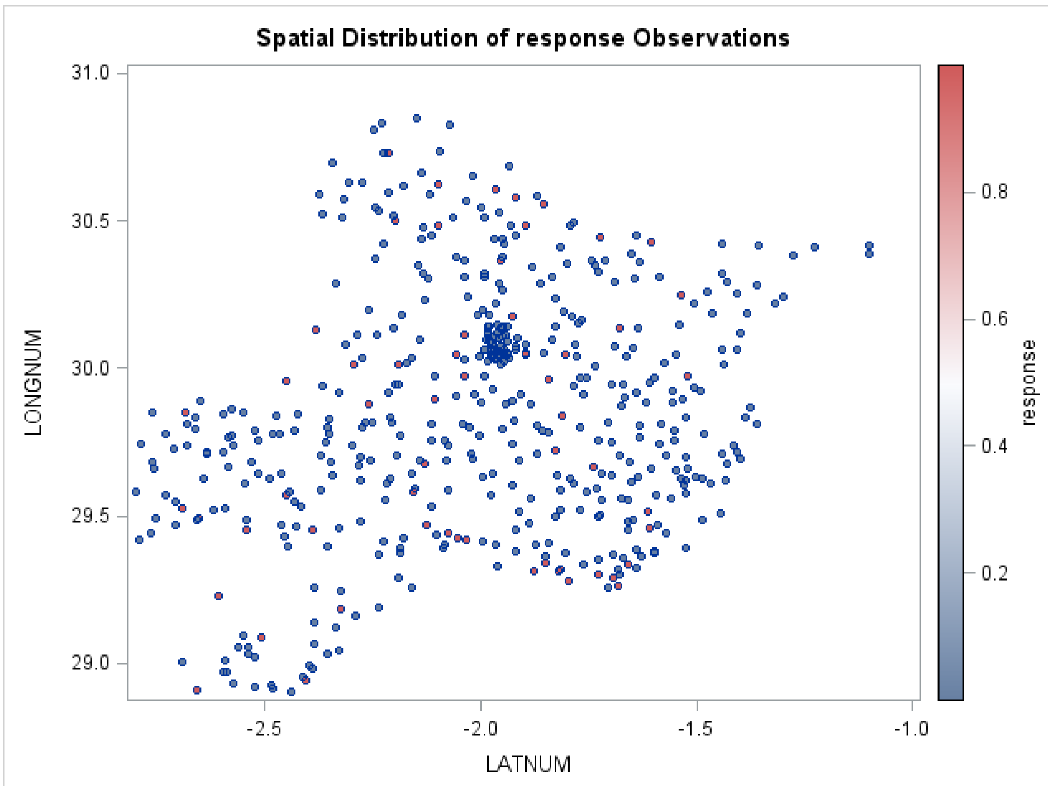
Table 6.1. Parameter estimates for a spatial joint marginal model for anthropometric measurements of malnutrition

	Wasting			Underweight			Stunting		
Indicator	Estimate	Std.Error	P-Value	Estimate	Std.Error	P-VALUE	Estimate	Std.Error	P-value
Intercept	-.817	0.790	0.3012	2.644	2.113	0.2108	1.250	1.063	0.2396
Child age in months									
0-11 months	reference								
12-23 months	-0.991	0.304	0.0011	-0.015	0.1539	0.9207	1.355	0.1625	$<.0001$
24^{+} months	0.588	0.300	0.0499	-0.200	0.107	0.0617	-0.148	0.1028	0.1507
Birth order									
1	reference								
2-3	0.040	0.3542	0.9109	0.239	0.1252	0.056	0.137	0.1308	0.2935
4-5	-0.388	0.3625	0.2847	0.296	0.1351	0.0285	0.148	0.1409	0.2922
6^{+}	0.971	0.4521	0.0317	0.1.044	0.1602	<0.0001	0.536	0.1472	0.0003
Mother's age									
$21 \geq$	reference								
<21	0.280	0.6759	0.6781	0.109	0.2576	0.6724	0.628	0.3213	0.0066
Mother's education									
No education & reference									
Primary	-0.482	0.4857	0.3212	-2.172	0.0.3556	$<.0001$	-0.738	0.1992	0.0002
Secondary & more	-0.3069	0.5555	0.5806	-2.350	0.3676	$<.0001$	-0.968	0.222	$<.0001$
Gender of the child									
Female	reference								
Male	0.361	0.2469	0.1436	-0.458	0.0918	$<.0001$	0.503	0892	$<.0001$
Birth weights									
$\geq 2500g$	reference								
$<2500g$	1.132	0.2013	0.0035	1.132	0.2013	$<.0001$	0.543	0.2458	0.0271
Province/region									
Kigali	reference								
South	-0.603	0.5163	0.2428	-0.043	0.2068	0.8356	0.537	0.1955	0.0061
West	-1.000	0.5126	0.0511	-0.428	0.2071	0.0389	0.118	0.1975	0.5502
North	-0.384	0.5246	0.4646	-0.332	0.2021	0.1008	-0.030	0.193	0.8777
Eastern	-1.178	0.6444	0.0667	-0.167	0.2885	0.5625	0.667	0.2736	0.0147
Knowledge on nutrition									
Yes	reference								
No	0.112	0.2564	-0.334	0.1045	0.0.0014	-0.323	-0.323	0.0996	0.0009
No	reference								
Yes	-0.141	0.2543	0.58	0.348	0.1096	0.0015	0.259	0.0916	0.0047
Wealth index									
Poor	reference								
Rich	-1.325	0.5604	0.0181	-0.307	1944	0.1145	-0.437	0.1785	0.0143
Middle	-0.805	0.4784	0.0924	-0.122	0.1688	0.4692	-0.260	0.1549	0.0.0931
Multiple birth									
Singleton	reference								
First multiple	0.029	1.0717	0.9781	1.308	0.3479	0.0002	0.376	0.4534	0.4063
Second multiple and more	-0.019	1.3151	0.9883	0.411	0.4473	0.3584	0.016	0.5875	0.9777
Incident of anemia									
No anemic	reference								
Anemic	0.490	0.2593	0.0586	0.3729	0.0958	$<.0001$	0.173	0.0948	0.2841
Body mass index									
$BMI \geq 18.5$	reference								
$BMI < 18.5$	1.359	0.4006	0.0007	1.1693	0.1813	$<.0001$	0.173	0.2124	0.4144
Incident of the fever									
Had fever last two weeks	reference								
No fever	-0.533	0.2814	0.0582	-0.497	0.1149	$<.0001$	-0.0005	0.1216	0.9697
Source of drinking water									
Others/yard	reference								
Piped into dwelling/yard	1.396	0.696	0.0045	0.096	0.4057	0.8122	-0.601	0.3427	0.079
Public tap	1.998	0.7249	0.0058	0.056	0.4071	0.891	-0.463	0.3431	0.1774
Protected spring/well	0.633	0.3236	0.1472	-0.219	0.3934	0.578	-0.046	0.1329	0.1571

TABLE 6.2. Random effect estimates

Effect	Estimate	SE.	Pr>z
Variance	0.7571	0.0924	< .0001
SP(GAU)	1.5864	0.4224	0.0165

FIGURE 6.1. Scatter plot for the malnutrition prevalence for joint distribution of stunting, underweight and wasting



6.5. Summary

In chapter 5, we used multivariate joint model of three anthropometric indices. However, this model does not allow us to include the spatial variability. This chapter

FIGURE 6.2. Classical and robust semivariogram for joint distribution of stunting, underweight and wasting

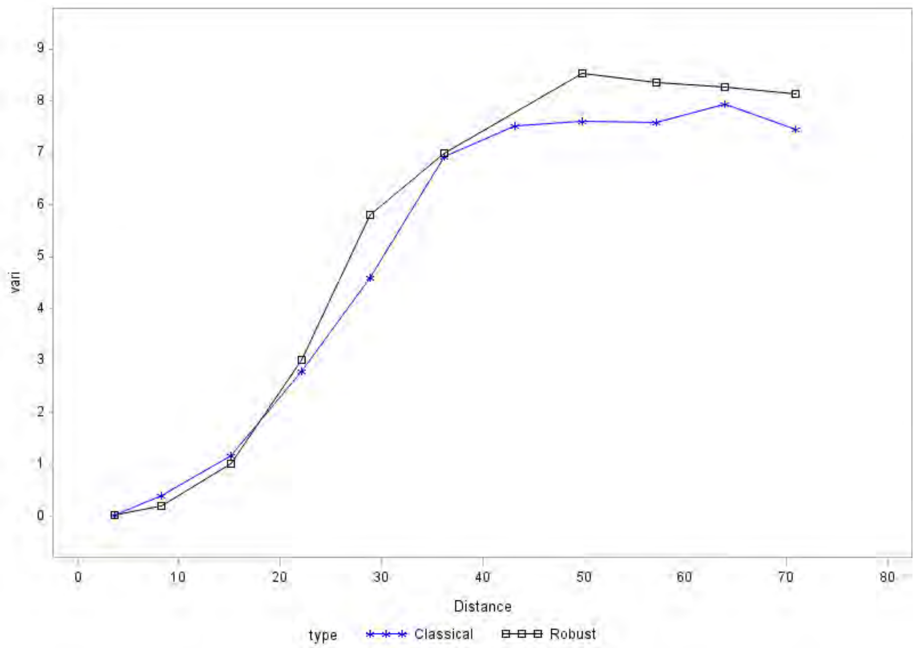


FIGURE 6.3. Predicted average spatial effects from the joint model for stunting

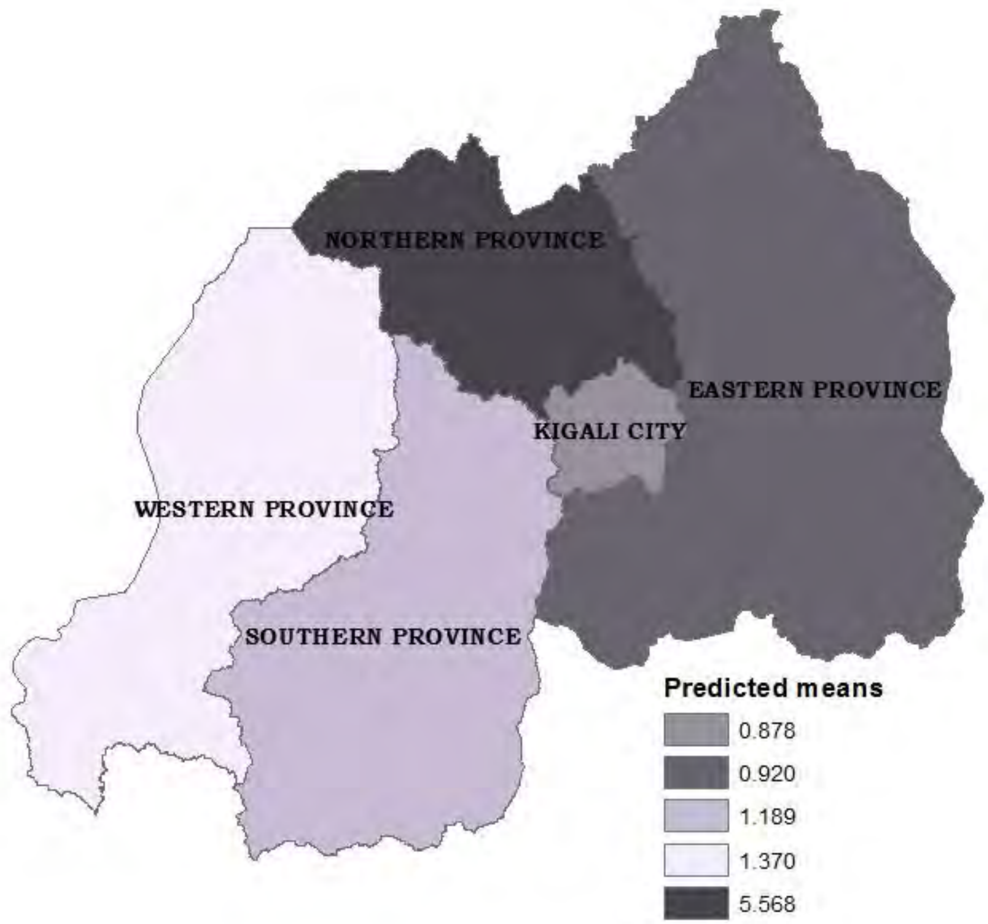


Figure 6.4. Predicted average spatial effects from the joint model for wasting

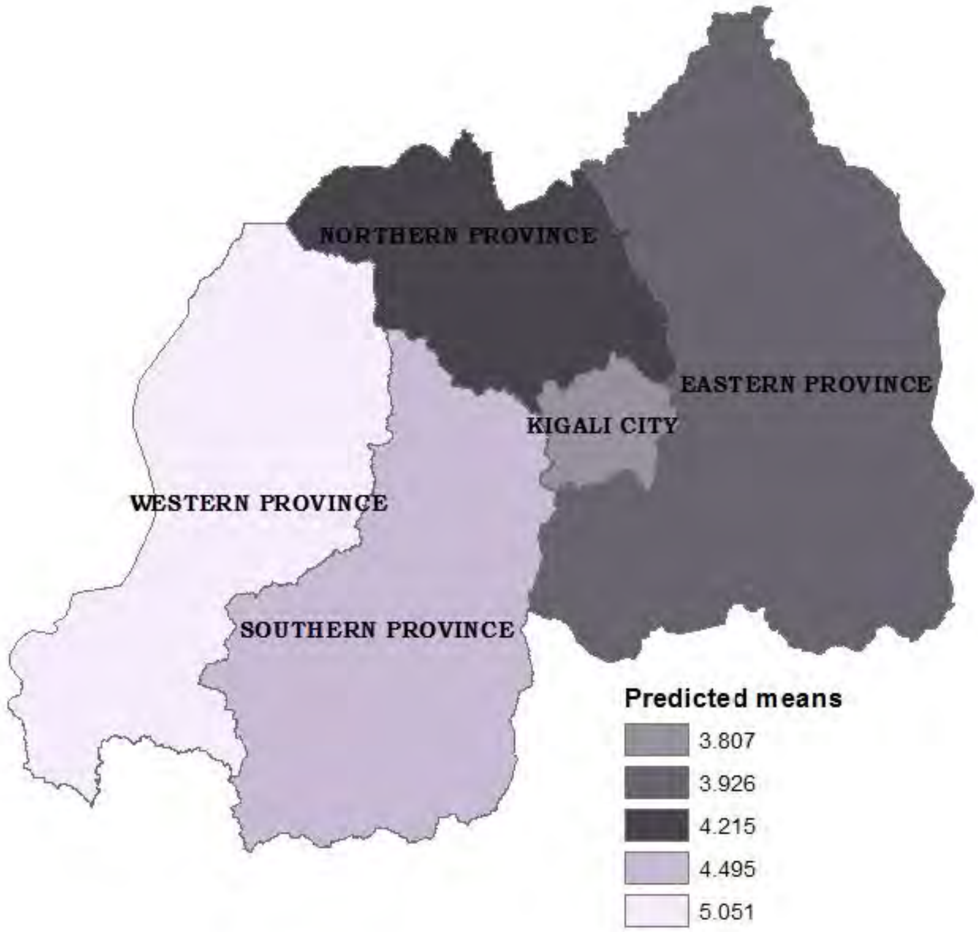
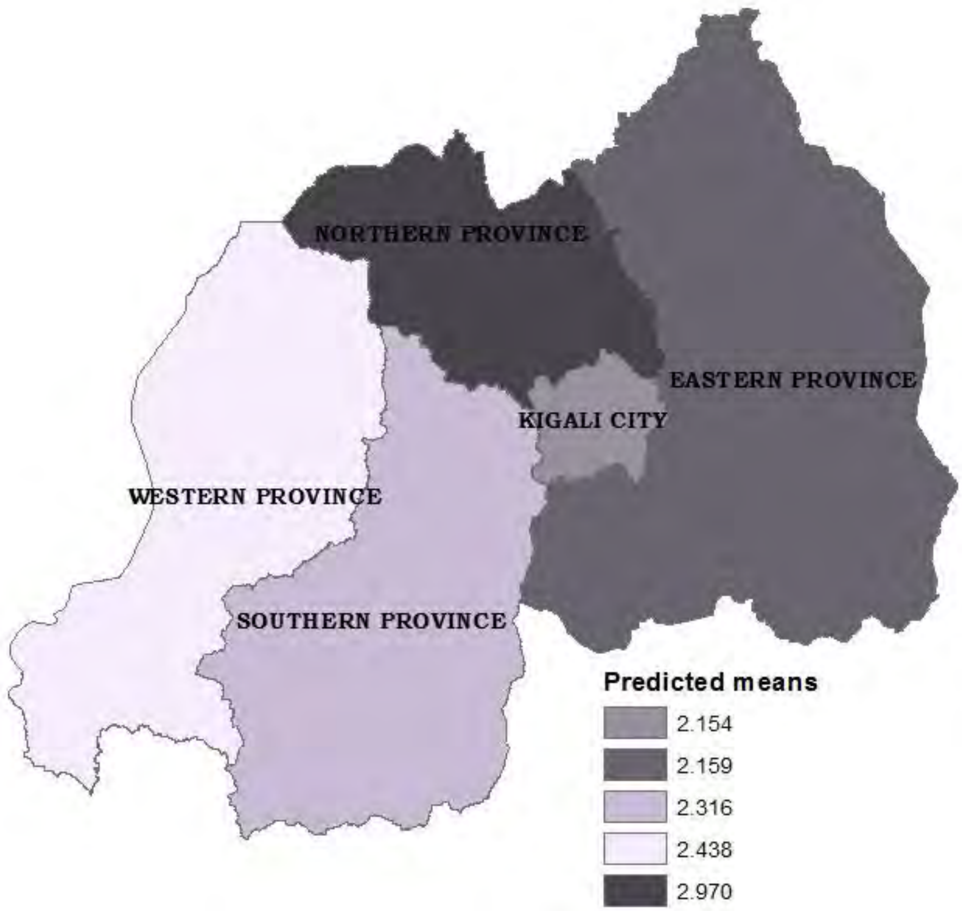


FIGURE 6.5. Predicted average spatial effects from the joint model for underweight



extended chapter 5 to include spatial variability and to produce the smooth maps of prevalence of malnutrition by predicting at unsampled location. Based on spatial generalized linear mixed model to wasting, stunting and underweight, we have identified the significant covariates and produced the prevalence map of each of the three responses. The findings of this study revealed that child age, birth order of the child, gender of the child, birth weight of the child, fever, multiple birth, mother's level of education, mother's age at the birth, anemia, body mass index of the mother, mother's knowledge on nutrition, wealth index of the family, source of drinking water and province are the key determinants of malnutrition of children under age five in Rwanda.

The findings of this study are consistent with previous studies (Das and Rahman, 2011; Kandala et al., 2011a; Habyarimana et al., 2014). This study found that prevalence of wasting is higher in Western province and lower in Kigali city, the prevalence of stunting is higher in Northern province and lower in Kigali city and the prevalence of underweight is higher in Northern Province and lower in Kigali city. These maps may be used for targeting programs in efforts to reduce children malnutrition. The findings of this study highlight, unexpected relationships which would be overlooked in analysis with separation of models or in cross-sectional analysis. The anthropometric indices and asset index are continuous distributions. In the next chapter we use quantile regression to account for the desired quantiles.

CHAPTER 7

Quantile regression models

7.1. Introduction

In the previous chapters, we have used binary logistic regression and survey logistic regression, ordinal logistic regression (proportional odds model with and without sampling design), generalized linear mixed model, multivariate joint model under GLMM and multivariate spatial joint model to include spatial variability. But all these methods estimate how the predictor variables are related to the mean value of the outcome variable. In this chapter we are interested to use the whole distribution of asset index in case of poverty of household and weight-for-height anthropometric index in the case of malnutrition of children under five years. Therefore in this chapter we consider quantile regression that allows for studying the impact of predictors on different desired quantiles of the response distribution, and thus provides a complete picture of the relationship between the response and predictor variables. Quantile regression is a flexible model in the sense that it does not involve link function that relates the variance and the mean of the response variable. The quantile regression method is robust to extreme points in the response space (outlier) but not to extreme points in the covariate space (leverage points); quantile regression is also a robust method in the sense that it makes no assumption about the distribution of error term in the model. These abilities of quantile regression, as introduced by Koenker and Basett (1978) to characterize the impact of variables on the whole distribution of the outcome of interest, motivated the use of quantile regression when assessing the risk factors associated to the poverty of households as well as the risk factors associated with the malnutrition of children under five years.

7.2. Model formulation and definition

Before defining the quantile regression, we highlight some of the notions of quantile function and give the definition of a sample quantile. Therefore, the word quantile is a synonym of percentile (Yu et al., 2003) and refers to the general case of dividing the population into 100 segments or sub-populations. A quartile separates the set into four sub areas or sub-populations containing an equal amount of observations within each sub-population, where the lower quarter is called the first quartile. The second quartile is well known as median. A quintile divides the reference population into five sub-population or groups, and a decile divides the population into ten sub-populations or groups; the median divides the population into two sub-populations. In quantile regression, equations are designed to estimate the relation of X with Y, conditional on quantiles (percentiles) of Y. In other words, this technique examines how the relation of X with Y changes depending on the score of Y. The quantile regression model is defined in Koenker and Basett (1978) as

$$y_i = x_i'\beta_\theta + u_{\theta_i} \tag{7.1}$$

with

$$Q_\theta(y_i|x_i) = x_i'\beta_\theta \tag{7.2}$$

and

$$Q_\theta(u_{\theta_i}|x_i) = F_u^{-1}(\theta|x_i) = 0 \tag{7.3}$$

where y_i is the i^{th} observation of the outcome variable, X_i is a vector of predictor (independent) variables, β_θ is a vector of unknown regression parameters and u_{θ_i} are independent identically distributed error terms with unspecified distribution; the quantities $Q_\theta(y_i|x_i)$ and $Q_\theta(u_{\theta_i}|x_i)$ mean the θ^{th} conditional quantile (percentile) of y_i and u_{θ_i} given x_i, respectively.

The θ^{th} sample quantile is given by $Q_Y(\theta) = \xi_\tau, 0 \leq \theta \leq 1$, of a random variable Y is the inverse of the cumulative distribution function written as $F_Y(y) = \theta$ defined as

$$Q_Y(\theta) = F_Y^{-1}(\theta) = inf\{y : F_Y(y) \geq \theta\} \tag{7.4}$$

The models in equation (7.1) and (7.2) are referred to as the linear location model where predictor variables affect only the location of the conditional distribution of the outcome variable. When the error terms are independent identically distributed (iid) the θ^{th} regression parameter

$$\beta_\theta = \beta + \left(F_u^{-1}(\theta), 0, 0, ..., 0\right)'.$$

In this case the conditional quantile planes are parallel and all parameters in β except the intercept, are similar for every value of θ. As a result the quantile regression slopes are constant for every quantile θ. In contrast, when error terms are not iid the quantile regression model is the linear location-scale model of heteroscedasticity, that can be defined as

$$y_i = x_i'\beta_\theta + (x_i'\gamma)u_{\theta_i} \tag{7.5}$$

with

$$Q_\theta\left(y_i|x_i\right) = x_i'\beta_\theta + x_i'\gamma F_u^{-1}(\theta) \tag{7.6}$$

with γ, an unknown scale parameter. The case of the linear location scale model of heteroscedasticity is essential for general class of quantile regression models (Koenker and Basett, 1982b). In this model the predictor variables affect the location as well as the scale of the response variable distribution and results change in the distribution since regression slopes vary across all parts of the distribution of the response variable. Therefore the θ^{th} regression parameter is given by $\beta_\theta = \beta + \phi F_Y^{-1}(\theta)$.

7.2.1. Parameter estimation.

The θ^{th} regression quantile estimator $\hat{\beta}_\theta$, also called regression quantile, is obtained by minimizing an asymmetric sum of weighted absolute deviation for the θ^{th} regression quantile $(0 \leq \theta \leq 1)$ defined by

$$\min_{\beta \in \mathbb{R}^p}\left[\sum_{i:y_i \geq x_i\beta_\theta} \theta|y_i - x_i'\beta_\theta| + \sum_{i:y_i < x_i'\beta_\theta}(1-\theta)|y_i - x_i'\beta_\theta|\right] = \min_{\beta \in \mathbb{R}^p}\sum_{i=1}^{n}\rho_\theta(y_i - x_i'\beta_\theta) \tag{7.7}$$

where $\rho_\theta(u) = \theta|u|I\left(u \geq 0\right) + \left(1-\theta\right)|u|I\left(u < 0\right)$, or simply $\rho_\theta(u) = \left(\theta - I\left(u < 0\right)\right)u$ is known as the check function, with the indicator function $I(.)$ that gives 1 to a positive residuals and 0 to a negative residuals (Koenker and Basett, 1978). The Least absolute deviation (LAD) estimator of β obtained by minimizing a symmetric sum of weighted absolute deviation is a special case of quantile for $\theta = 0.5$, which is the median and its estimate is also known as L_1-norm estimate. In the case of weighted quantile regression, it is straightforward by simply including the weight in equation (7.7) as

$$\min_{\beta_w \in \mathbb{R}^p} \left[\sum_{i:y_i \geq x_i \beta_{\theta w}} w_i \theta |y_i - x_i' \beta_{\theta w_i}| + \sum_{i:y_i < x_i' \beta_{\theta w}} w_i (1-\theta) |y_i - x_i' \beta_{\theta w}| \right] \tag{7.8}$$

where w_i, $i = 1, 2, ..., n$ are the weights.

The minimization of the weighted sum of absolute deviations in equations (7.7) and (7.8) can be formulated as a linear programming problem, which can be solved using a linear programming algorithm.

There are a number of algorithms in literature used to solve the linear programming problems for quantile regression. The simplex algorithm for median regression developed by Barrodale and Robert (1974) and extended to quantile regression by Koenker and D'Orey (1993), reduces the computing time required by the general simplex algorithm and it is suitable to the data sets less than 5000 observations and 50 variables. The interior point algorithm of Karmakar (1984), also known as the Frisch-Newton algorithm, was extended to quantile regression by Portnoy and Koenker (1997) and Koenker and Hallock (2000). This algorithm was developed as an alternative to solve large to huge linear programming problems. The finite smoothing algorithm was first developed by Clark and Osborne (1986) and later by Madsen and Nielsen (1993) to solve linear programming problems of L_1 regression and it was extended to quantile regression by Chen (2007). Each of these three algorithms has its own advantages; none of them can fully dominate the others. Based on the advantages of each of them Chen (2004) developed an adaptive algorithm combining these three algorithms. Interpretation of quantile regression parameter estimates is not different from that of the general linear model estimates as they are all rates of change when the effects of some variables in the model are adjusted for. The classical regression coefficient reflects the change in the mean of the distribution of the response variable Y, associated with a unit change in the predictor variable X that corresponds to the coefficient. However, the quantile regression coefficient reflects the change in a specified quantile of the response variable associated with a unit change in the predictor variable X that corresponds to the coefficient. The use of quantile regression allows for comparison of how some percentiles of the response variable may be more affected by the change in the size of the regression coefficients of different percentiles.

7.3. Properties of quantile regression

The quantile regression estimates have a number of equivariance properties, that are very important for meaningful interpretation of results from regression analysis, especially for transformed data. Koenker and Basett (1978) formulated four equivariance properties of quantile regression. Once we denote the quantile estimate for a given $\theta \in (0,1)$ and observations (y, X) by $\hat{\beta}(\theta; y, X)$, then for any $p \times p$ nonsingular matrix A, $\gamma \in \mathbb{R}^k$, and $a > 0$ holds

$$\hat{\beta}(\theta; ay, X) = a\hat{\beta}(\theta; y, X) \tag{7.9}$$

$$\hat{\beta}(\theta; -ay, X) = a\hat{\beta}(\theta; y, X) \tag{7.10}$$

$$\hat{\beta}(\theta y + X\gamma, X) = a\hat{\beta}(1-\theta; y, X) + \gamma \tag{7.11}$$

$$\hat{\beta}(\theta; y, AX) = A^{-1}\hat{\beta}(\theta; y, X) \tag{7.12}$$

where properties (7.9) and (7.10) imply a form of scale equivariance, (7.11) is normally called shift or regression equivariance, and property (7.12) is known as parametrization of design.

Invariance to monotonic transformations: Quantiles exhibit, besides usual equivariance properties, equivariance to monotone transformations. Let $f(.)$ be a nondecreasing function on $\mathbb{R}$, then for any random variable Y

$$Q_{f(Y)}(\theta) = f\{Q_Y(\theta)\} \tag{7.13}$$

This means that the quantiles of the transformed random variable in equation (7.13) are simply the transformed quantiles of the original variable Y. This is not the case of the conditional expectation $E\{f(Y)\} \neq f(EY)$ unless $f(.)$ is affine function. The property (7.13) follows immediately from the elementary fact that for any monotone function f then

$$P(Y \leq y) = P(f(Y) \leq f(y)) \tag{7.14}$$

holds; for more detail see (Koenker and Hallock, 2000; Koenker, 2005).

Robustness: The linear programming of quantile regression problem has many important implications from theoretical and practical points of view (standpoints). It is certain that the estimate of quantile regression will be obtained in a finite number of simplex iterations (Barrodale and Robert, 1974). Unlike the case of the mean type regression, the parameter vector estimate is robust to outliers. This means that if $y_i - x_i'\hat{\beta}_\theta > 0$, then y_i increases towards ∞, or $y_i - x_i'\hat{\beta}_\theta < 0$, then y_i decreases towards $-\infty$, without varying the solution of $\hat{\beta}_\theta$ (Buchinsky, 1998). In quantile regression, it is not the magnitude of the outcome variable that matters but on which side of the estimated hyperplane the observation is, which is not the case in the least squares estimates. However, quantile regression estimates lack robustness against observations that are extreme with respect to covariate variables known as higher leverage points.

7.4. Quantile regression goodness-of-fit

The goodness-of-fit of quantile regression as defined by Koenker and Machado (1999) derives from the familiar R^2 (coefficient of determination) of the classical ordinary least squares regression. It compares the quantile regression model fitted with intercept only and the quantile regression model fitted to a given number of predictor variables including the intercept. Let us consider the linear model for the conditional quantile function (Koenker and Machado, 1999) given by

$$Q_\theta\left(y_i|x_i\right) = x_i'\beta_\theta \tag{7.15}$$

the model (7.15) can be partitioned as follows

$$Q_\theta(y_i|x_i) = x_{i1}'\beta_{1\theta} + x_{i2}'\beta_{2\theta} \tag{7.16}$$

The partitioned model presented above results from partitioning the design matrix X into (X_1, X_2) and vector of parameter β_θ into $\beta_{1\theta}$ and $\beta_{2\theta}$. The components x_{i1} and x_2 of the model are the i^{th} rows of X_1 and X_2, which are the $m \times (p-k)$ and $m \times k$ design matrices, respectively. The components $\beta_{1\theta}$ and $\beta_{2\theta}$ are $(p-k) \times k$ and $k \times 1$ vectors of parameters respectively. The unrestricted θ^{th} quantile regression estimate $\hat{\beta}_\theta$ of the full model, minimizes the weighted sum of absolute deviations given by

$$\hat{V}_\theta = \min_{\hat{\beta}_\theta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\theta(y_i - x_i'\hat{\beta}_\theta) \tag{7.17}$$

Consider the restricted model, that can be defined as $Q_\theta(y_i|x_i) = x_{i1}'\beta_{1\theta}$. Thus the restricted estimator $\widetilde{\beta}_\theta = \left(\hat{\beta}_{1\theta}', 0'\right)'$, that is the θ^{th} quantile estimate under the k-dimensional linear restriction corresponding to null hypothesis

$$H_0 : \beta_{2\theta} = 0, \tag{7.18}$$

minimizes

$$\widetilde{V}_\theta = \min_{\hat{\beta}_\theta u \in \mathbb{R}^{p-k}} \sum_{i=1}^{n} \rho_\theta(y_i - x'_{1i}\hat{\beta}_{1\theta}) \tag{7.19}$$

The goodness-of-fit criterion may be defined (Koenker and Machado, 1999) as

$$R^1_\theta = 1 - \frac{\hat{V}_\theta}{\widetilde{V}_\theta} \tag{7.20}$$

like the classical R^2, $0 \leq R^1_\theta \leq 1$, since $\hat{V}_\theta u \leq \widetilde{V}_\theta$. Unlike R^2 that measures the relative success of the two models for the conditional mean function in terms of residual variances, R^1_θ measures the relative success of the unrestricted and restricted quantile regression models at a specific quantile in terms of an appropriately weighted sum of absolute residuals. Therefore, R^1_θ is a local measure of goodness-of-fit for a particular quantile rather than a global measure of goodness-of-fit over the entire conditional distribution as in classical R^2 from least squares regression. It is possible that under some circumstances a covariate might significantly affect one tail of the conditional distribution of the response variable and might have no effect in other tail. If R^1_θ is high at one tail of the distribution than at the other tail, this might be an indication of heteroscedasticity. If the full model in equation (7.16) is better at the θ^{th} quantile than the restricted model constrained by (7.18), then $\hat{V}_\theta$ should be significantly smaller than $\widetilde{V}_\theta$ as results $\hat{R}^1_\theta$ will be higher indicating a better model fit. Better in this case means that the predictor variables X_2 has a significant influence at the θ^{th} quantile (Koenker and Machado, 1999).

7.5. Inference for quantile regression

The conditional quantile functions of the response variable given predictor variables in the model are all supposed to be parallel to one another. In other words, the effects of covariate variables in the model shift the location of the conditional distribution of the outcome variable only, but do not alliterate its scale or shape and therefore the slope coefficients of different quantile regressions are equal. But in several applications of quantile regression, estimated slopes often differ considerable through quantiles and this makes the test of equality of slope parameters across quantiles to form a central component of inference in quantile regression (Koenker, 2005).

Even though there are no practical statistical inferences in the case of finite sample for quantile regression, like it is in least squares methods, the asymptotic theory offers practical statistical inferences for quantile regression. This is the foundation of several statistical approaches to inference such as asymptotic covariance matrix, the Wald test, rank tests and likelihood ratio tests as well as construction of some of the confidence intervals for regression quantiles.

7.5.1. Asymptotic distribution of quantile regression.

The asymptotic distribution of quantile regression estimator $\hat{\beta}_\theta$ results from that of sample quantiles. The asymptotic distribution of the sample quantile, $\hat{\xi}_\theta$, calculated from the n independent identical distributed (iid) observations of the outcome variable with the distribution F is given by

$$\sqrt{n}\left(\hat{\xi}_\theta - \xi_\theta\right) \rightarrow N\left(0, \omega^2\right) \tag{7.21}$$

with $\omega^2 = \theta(1-\theta)/\left(f^2\left(F^{-1}(\theta)\right)\right)$. There are two influences on the precision of the θ^{th} quantile of interest from the sample. The numerator $\theta(1-\theta)$ effect tends to make $\hat{\xi}_\theta$ more precise in the tail, however this would be dominated by the effect of the density term $1/f^2\left(F^{-1}(\theta)\right)$, that tends to make $\hat{\xi}_\theta$ less precise in the region of low density (Koenker, 2005); this term is the reciprocal of a density function referred to as the sparsity function by Turkey (1965) or quantile density function by Parzen (1979).

The sparsity function $s(\theta)$ reflects the density of observations near ξ_θ, such that the estimation of the quantile becomes difficult when the observations are very sparse at the close proximity of the quantile. Conversely, the quantile is precisely estimated when the sparsity of the data near ξ_θ is low, such that there are many observations near the quantile. In other words the sparsity of the data at the quantile of interest ξ_θ determines how precise is the estimated value of the quantile.

To generalize the asymptotic distribution of sample quantiles of regression quantiles, consider the quantile linear regression model $y_i = x_i'\beta_\theta + u_{\theta i}$ with independent identically distributed error terms $u_{\theta i}$. These terms have a common distribution function F associated with the density function f, and $f\left(F^{-1}(\theta_i)\right) > 0$, for $i = 1, 2, ..., n$. Then the asymptotic distribution of the quantile regression estimator $\hat{\beta}_\theta$ can be estimated as

$$\sqrt{n}\left(\hat{\beta}_\theta - \beta_\theta\right) \rightarrow N\left(0, \theta(1-\theta)K_\theta^{-1}GK_\theta^{-1}\right) = N\left(0, \Lambda_\theta\right) \tag{7.22}$$

where $G = \lim_{n\rightarrow\infty} n^{-1}\sum_i x_i x_i'$
and $K_\theta = \lim_{n\rightarrow\infty} n^{-1}\sum_i x_i x_i' f_i(\xi_{\theta i})$. The matrix G is a positive definite $p \times p$ matrix. If the error terms are assumed to be iid, then the density functions $f_i(\xi_{\theta i})$ are identical and the sandwich covariance matrix Λ_θ collapses to a simplified expression given by $\Lambda_\theta = \omega^2 = \frac{\theta(1-\theta)}{f^2(F^{-1}(\theta))} \lim_{n\rightarrow\infty} n\left(\sum_i x_i x_i'\right)^{-1}$
such that the asymptotic distribution of $\hat{\beta}_\theta$ is

$$\sqrt{n}\left(\hat{\beta}_\theta - \beta_\theta\right) \rightarrow N\left(0, \omega^2 G^{-1}\right) \tag{7.23}$$

The simplified expression of Λ_θ shows that under the iid error regression model, the asymptotic precision of quantile regression estimates depends on the sparsity function and the term $\theta(1-\theta)$. In the quantile regression model the sparsity function takes the role similar to that of the standard deviation of the error terms, σ, in the least squares estimation procedure of the iid error regression model.
But the assumption of iid error terms is very restrictive and sometimes it does not hold in practical application when the assumption holds the conditional quantiles are simple shifts of one another since all conditional quantiles planes are parallel. Therefore, the application of quantile regression does not provide any additional information to that provided by the least squares estimator since estimated regression coefficients for different quantiles $\hat{\beta}_{\theta j}$, have a common value, $\hat{\beta}_\theta$. However, in real life problems it is almost impossible to justify the assumption of iid error terms.
The asymptotic distribution of estimated regression coefficient in equation (7.22) can be extended to several regression coefficient vectors calculated at different quantiles see Koenker (2005) for more details.

7.5.2. Estimation of covariance matrix.

The precision of the θ^{th} quantile is measured by the covariance matrix. This covariance matrix can be estimated by several different methods. Some methods are direct and asymptotic that need the estimation of the sparsity function, whilst others are bootstrap mainly based on resampling. To estimate the precision of the θ^{th} quantile regression estimate directly, the nuisance quantity

$$s(\theta) = \left[f\left(F(\theta)\right)\right]^{-1} \tag{7.24}$$

must be estimated. There is a large literature on estimating equation (7.24), for instance Siddiqui (1960), Bofinger (1975) and Sheather and Maritz (1983). Differentiating the identity $F(F^{-1}(t)) = t$, it is found that the sparsity function is just the derivative of the quantile function and is given by

$$s(t) = \frac{d}{dt}F^{-1}(t) \tag{7.25}$$

Based on Siddiqui (1960) ideas, $s(t)$ is estimated using simple difference quotients of the empirical quantile functions (Koenker and Machado, 1999) as

$$\hat{s}_n(t) = \left[\hat{F}_n^{-1}(t + h_n) - \hat{F}_n^{-1}(t - h_n)\right]/2h_n \tag{7.26}$$

with $\hat{F}^{-1}$, an estimate of F^{-1}, and h_n is a bandwidth that tends to zero as $n \rightarrow \infty$. Hall and Sheather (1988) proposed a bandwidth rule based on Edgeworth expansions for studentized sample quantiles as

$$h_n = n^{-1/3} z_\alpha^{2/3} \left[1.5 s(t)/s''(t)\right]^{1/3} \tag{7.27}$$

where z_α satisfies $\Phi(z_\alpha) = 1 - \alpha/2$. In the absence of other information about the form of $s(.)$, we may use Gaussian model to select the bandwidth h_n, that produces

$$h_n = n^{-/3} z_\alpha^{2/3} \left[\frac{1.5\phi^2\left(\Phi^{-1}(t)\right)}{2\left(\Phi^{-1}(t)\right)^2 + 1}\right]^{1/3} \tag{7.28}$$

When the bandwidth is chosen, then $\hat{F}^{-1}$ can be estimated using the empirical quantile function residual from quantile regression fit or the empirical quantile function of Bassett and Koenker (1982) can be used to to estimate $\hat{F}^{-1}$.

The estimate of the asymptotic covariance matrix of $\hat{\beta}_\theta$ is simply obtained by substituting the estimate of the sparsity function in the simplified equation of Λ_θ. The Powell (1986) estimator for censored regression quantiles can be modified and used in the quantile regression to estimate both the sparsity function for an independent and identically distributed error and non independent and identically distributed error.

In the case of the iid error terms assumption, the sparsity function can be estimated by one sided estimator given by

$$\hat{f}\left(F^{-1}(\theta)\right) = (\hat{c}_n n)^{-1} \sum_{i}^{n} I\left(0 \leq \hat{u}_{\theta i} \leq \hat{c}_n\right) \tag{7.29}$$

where $\hat{u}_{\theta i} = y_i - x_i'\hat{\beta}_\theta$ and c_n is the kernel bandwidth. Then the cross validation methods such as log-likelihood and least squares may be used to obtain the optimal selection of c_n. Therefore, the resultant kernel estimator of the covariance matrix for β_θ may be given by

$$\hat{\Lambda}_\theta = \frac{\theta(1-\theta)}{\hat{f}^2(F^{-1}(\theta))}\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right) \tag{7.30}$$

The two sided kernel estimator in which the indicator function given in equation (7.29) is replaced by $I\left(-\hat{c}_n/2 \leq \hat{u}_{\theta i} \leq \hat{c}_n/2\right)$ may be used to estimate Λ_θ. When the error terms are heteroscedastic, K_θ can be estimated by $(\hat{c}_n n)^{-1}\sum_i^n I\left(0 \leq \hat{u}_{\theta i} \leq \hat{c}_n\right) x_i x_i'$. Instead of estimating the sparsity function, bootstrap method based on varying assumption about error terms and the form of the asymptotic covariance matrix may be used. He and Hu (2002) proposed the Markov chain marginal bootstrap (MCMB) method that differs from other bootstrap methods in two main aspects. The method solves one dimensional equations for parameters of any dimension, and produces a Markov chain instead of an independent sequence. The aim of the MCMB method is to simplify the computation problems associated with bootstrap in higher-dimensional problem.

7.5.3. Test of linear hypothesis.

After reviewing the estimation of parameters, it is very crucial to also review the statistical tests used in these methods.

Wald test: The Wald test is based on the regression coefficients estimated from unrestricted model (Koenker and Basett, 1982a). It tests the general linear hypothesis for $p \times 1$ vector of parameters, β_θ, in the case of single quantile regression coefficient, stated as $H_1 : K\beta_\theta = h$ against $H_0 : K\beta_\theta \neq h$, where K is a $k \times p$ matrix of the coefficient, h is a $k \times 1$ vector of constants that are commonly zeros (Koenker, 2005) and its test statistic is given by

$$T = n\left(K\hat{\beta}_\theta - h\right)'\left(K\hat{\Lambda}_\theta^{-1}K'\right)^{-1}\left(K\hat{\beta}_\theta - h\right) \tag{7.31}$$

The test (7.31) is asymptotically χ^2_q, with q the rank of the matrix K. This test of one quantile was generalized by Koenker and Basett (1982b) to account for several different quantiles and is defined as $H_0 : K\zeta = h$ and its test statistic is given by

$$T = n\left(K\hat{\zeta} - h\right)'\left(K\left(\Omega \otimes G^{-1}\right)^{-1}K'\right)^{-1}\left(K\hat{\zeta} - h\right) \tag{7.32}$$

which is asymptotically non-central χ^2 with rank q degree of freedom and noncentrality

$$\eta = \left(K\left(Q_\theta(u) \otimes \gamma_0\right)\right)\left(K\left(\Omega \otimes G^{-1}\right)K'\right)^{-1}\left(K\left(Q_\theta(u) \otimes \gamma_0\right)\right) \tag{7.33}$$

In the case of homoscedastic model, the slope parameters are identical at every quantile (Koenker and Basett, 1982b), and the test statistic T is asymptotically central χ^2 with $(n-1) \times (k-1)$ degrees of freedom, where k is the number of parameters in the model, and n is the number of quantiles for which the model is fitted (Koenker, 2005; Koenker and Basett, 1982b).

This formulation of the Wald test accommodates a wide variety of testing situations, from simple tests on one quantile regression coefficient to joint tests that involve different quantiles and several covariates. Therefore, based on this test, it is possible to test the equality of several slope coefficients across different quantiles. These tests provide a robust alternative to the classical least-squares based tests of heteroscedasticity as they are insensitive to the outliers in the response variable observations. Similar formulation can be used to accommodate nonlinear hypotheses (Koenker, 2005). Further Newey and Powell (1987) discussed the test for symmetry based on this approach.

Likelihood ratio test

The likelihood ratio (LR) test is based on the objective function values in the restricted and unrestricted models. The linear hypothesis to be tested in the case of the likelihood ratio test is the same as stated under Wald test above. Koenker and Machado (1999) adapted the Koenker and Basett (1982a) method and showed that under H_0 when the error terms are iid but drawn from the distribution function, the test statistic is given by

$$L_n(\theta) = \frac{2(\widetilde{V}_\theta - \hat{V}_\theta)}{\theta(1-\theta)s(\theta)} \tag{7.34}$$

where $\widetilde{V}_\theta$ and $\hat{V}_\theta$ are given by (7.17) and (7.19), and $s(\theta)$ is the sparsity function. $L_m(\theta)$ is asymptotically χ^2_q as in the Wald test statistics. Similarly, consider a location and scale form of the asymmetric Laplacean density

$$f\sigma(u) = \theta(1-\theta)\exp\left(-\rho_\theta(u)/\sigma\right) \tag{7.35}$$

that produces the LR statistics

$$-2\log\lambda^*_n(\theta) \equiv 2n\log\left(\widetilde{V}(\theta)/\widehat{V}(\theta)\right) \tag{7.36}$$

The asymptotic behaviour of this version of Likelihood ratio statistic follows from equation (7.34) results (Koenker and Machado, 1999)

$$\begin{aligned} -2\log\lambda^*_n(\theta) &= 2n\log\left(1+\left(\widetilde{V}(\theta)-\widehat{V}(\theta)\right)/\widehat{V}(\theta)\right) && (7.37)\\ &= 2n\left(\widetilde{V}(\theta)-\widehat{V}(\theta)\right)+0_p(1) && (7.38)\\ &= 2\left(\widetilde{V}(\theta)-\widehat{V}(\theta)\right)/\sigma(\theta)+0_p(1) \end{aligned}$$

where $\sigma(\theta) = E\rho_\theta(u) < \infty$ and $\hat{\sigma}(\theta u) = \frac{\widehat{V}(\theta)}{n} \to \sigma(\theta)$. Therefore, based on the null hypothesis H_0 in (7.35), the test statistic becomes

$$\Lambda_n(\theta) = \frac{2n\sigma(\theta)}{\theta(1-\theta)s(\theta)}\log\left(\widetilde{V}(\theta)/\widehat{V}\right) \tag{7.39}$$

is also asymptotically χ^2_q . Therefore, the likelihood ratio test can be used to test the global hypothesis that quantile regression slopes coefficients are identical across quantiles.

Koenker and Machado (1999) showed that the Wald test and the likelihood ratio test are asymptotically equivalent and that the distributions of the test statistics converges to χ^2_k. **Rank test of linear hypothesis**

Gutenbrunner et al. (1993) introduced tests of a general linear hypothesis for the linear regression model that are based on regression rank scores of Gutenbrunner and Jureckova (1992). The tests are robust to observations that are outlying with respect to the response variable, and are asymptotically distribution free; this means that no nuisance parameters that depend on the error term distribution need to be estimated for the computation of the test statistic.

The regression rank score process of the restricted form of linear location-scale model is given by

$$\hat{a}_n(\theta) = \text{argmax}\{y'a|X_1'a = (1-\theta)X_1'e, au[0,1]^n\} \tag{7.40}$$

with **e**, an n- vector of 1's and n by p matrix, $\mathbf{X}$ is partitioned into $(X_1 : X_2)$ as well as the vector of parameter,β_θ into $\beta_{1\theta}$ and $\beta_{2\theta}$. Therefore the linear hypothesis can be tested as $H_0 : \beta_{2\theta}$; $\beta_{1\theta}$ unspecified, against the local alternative $H_n : \beta_{2n\theta} = \beta_{0\theta}/\sqrt{n}$; with $\beta_{0\theta} \in R^q$, fixed. The regression rank scores are $n \times 1$ vector, $\hat{a}_n(\theta) = (\hat{a}_{n1}(\theta), ..., \hat{a}_{nn}(\theta))$. The test statistic for testing H_0 against H_n is given by

$$T_n = \frac{S_n' M_n^{-1} S_n}{A^2(\phi)} \tag{7.41}$$

where

$$S_n = n^{-1/2}\left(X_{n2} - \hat{X}_{n2}\right)' \hat{b}_n,$$

$$M_n = n^{-1}\left(X_2 - \hat{X}_2\right)'\left(X_2 - \hat{X}_2\right),$$

$$\hat{X}_2 = X_1\left(X_1'X_1\right)^{-1} X_1'X_2,$$

$$\hat{b}_n = \left(-\int \phi(t) d\hat{a}_{in}(t)\right)_{i=1}^{n}$$

$$A^2(\phi) = \int_0^1 \left(\phi(t) - \bar{\phi}\right)^2 dt$$

,

$$\bar{\phi} = \int_0^1 \phi(t) dt$$

and ϕ is a score generating function bounded variation. The test is based on the asymptotic distribution of T_n under the null hypothesis H_0. Under H_0, T_n is asymptotically distributed as central χ^2_q, whereas under the local alternatives hypothesis H_n, T_n is a noncentral χ^2_q and noncentral parameter η, defined under the Wald test. Koenker and Machado (1999) extended the work of Gutenbrunner et al. (1993) to the location scale linear model. In their approach, they replaced an ordinary least squares fit by a weighted least squares. Then the test statistic is defined as

$$T_n = \frac{S_n' M_n^{-1} S_n}{\theta(1-\theta)} \tag{7.42}$$

Under the null hypothesis the modified T_n has a central χ_q^2 distribution, however based on the local alternative hypothesis it has a noncentral χ_q^2 distribution with non-centrality parameter, $\eta(\phi, \xi)$. Then, this test statistic can be used to identify a global effect of the covariate variables on the outcome variable across quantiles, or local effect by choosing the score function ϕ to apply only on one quantile of interest θ (Koenker, 2005).

7.5.4. Confidence intervals of quantile regression.

There are different approaches in literature for constructing confidence intervals and bands for regression quantile parameter $\beta(\theta)$. These approaches are mainly classified into three methods: sparsity or direct estimation, rank score, and resampling (Kocherginky et al., 2005). The sparsity is the most direct and the fastest, but involves estimation of sparsity function, that is not robust for the data that are not iid. To circumvent this problem, a Huber sandwich estimate is computed using a local estimate of the sparsity function. Rank score methods avoid direct estimation of the error densities. It was first introduced by Gutenbrunner and Jureckova (1992) for an iid error model and Gutenbrunner et al. (1993) used it to construct a rank test for the null hypothesis and later on Koenker (1994) proposed an attractive method for constructing the confidence intervals based on inversion of a rank score test. This approach does not need the estimation of the sparsity function. Unlike the confidence intervals based on the estimation of the sparsity function, the confidence interval resulting from the inversion of rank tests are not symmetric. However, they are centered on the point estimate $\hat{\beta}_{2\theta}$ of the partitioned model consisting of one predictor variable X_2, $y = X_1\beta_{1\theta} + X_2\beta_{2\theta} + u_\theta$, in the sense that $T_n(\hat{\beta}_{2\theta}) = 0$. Koenker and Machado (1999) extended this method to location-scale regression model. However, the rank score method uses the simplex algorithm which is computationally expensive with large data sets.

The Bootstrap approach can be used to compute the most reliable confidence intervals for quantile regression estimates. Chen (2004) noted that resampling methods are not recommended for small data sets with sample size $n < 5000$, and the number of predictor variables, $p < 20$, as they can only achieve the stability for relatively larger data sets.

Parzen et al. (1994) proposed a general and simple resampling method based on pivotal estimating function for inferences about the true parameter β. This method can be adapted and used to construct confidence intervals for quantile regression estimates. The approach achieves robustness to some heteroscedastic quantile regression models by exploiting the asymptotical pivotal role of the quantile regression (Koenker, 1994).

He and Hu (2002) developed a new general resampling method, referred to as the Markov chain marginal bootstrap (MCMB). This method has an advantage over other bootstrap methods instead of solving a p-dimension system (or its equivalent) for each replication it solves only p one-dimensional equations, for moderate to large data sets. MCMB uses the same time needed for usual bootstrap method. Kocherginky et al. (2005) adapted MCMB to quantile regression which aims to provide faster computations, to construct confidence intervals for quantile regression and called it MCMB-A method.

7.6. Application on Demographic and Health Survey data to identify the determinants of poverty of household and malnutrition of children under five years in Rwanda

In this study, as application we have used the households data in case of poverty and women data in case of malnutrition. We first consider poverty and thereafter malnutrition.

7.6.1. Determinants of poverty of households.

In previous studies (Habyarimana et al., 2015a), we have used logistic regression and in chapter 4, we have used GLMM but in all these studies the response variable poverty was categorized into two levels namely poor and not poor. In the present study the main objective is to consider the whole outcome distribution based on quantile regression.

Model fitting: As the RDHS data was collected using multistage sampling, the researchers included sampling weights in the analysis to account for complex sampling design. PROC QUANTREG in SAS 9.3 was used to compute parameter estimates, statistical inferences as well as to plot quantile plots. As the data set is large enough 12540 > 5000, the researchers used a resampling method to compute the confidence intervals (Koenker and Machado, 1999) and the interior algorithm was used to compute the quantile regression estimates in SAS. The non-linearity between size of household head and the asset index was assessed by including the quadratic term for size in the analysis and their significance was then examined. The goodness-of-fit and the equality of slopes were tested as in Koenker and Machado (1999). Various researchers (Filmer and Pritchett, 1998; Booysen, 2002; Lokosang et al., 2014; Habyarimana et al., 2015a) created asset index, where households were classified into five quintiles as follows: first quintile (20%) as poorest, second quintile (20%) as poor, third quintile (20%)as middle, fourth quintile (20%) as rich and the fifth quintile (20%) as richest (highest). Based on this classification and the results from Tables 2.8 and 2.9, we used 10^{th} (lowest), 20^{th}, 40^{th}, 50^{th} and 80^{th} percentiles and Ordinary Least Square (OLS) was reported for comparison purposes.

Results and interpretations

The Wald test was used to test the hypothesis of pure location shift that all the slopes coefficients of the quantile regression model fitted to the household data are the same across the five quantiles. The joint test for equality slopes coefficients of household data for the following quantiles 0.10, 0.20, 0.40, 0.50 and 0.80 was significant (p-value< .0001); which means that the effects of explanatory variables on the household data are not the same across the five quantiles. This is the evidence that the quantile regression can show more information from different quantiles. Therefore, it is reasonable to use quantile regression. The goodness-of-fit of the quantile regression to the household data at each of the selected quantiles was assessed using pseudo R-square by Koenker and Machado (1999). The values of pseudo R-square at 10^{th}, 20^{th}, 40^{th}, 50^{th} and 80^{th} quantiles, together with the value of the measure of goodness-of-fit for the OLS R^2, are shown in the last row of Table 7.1; where the value of pseudo R-square increases with the quantile being increased by almost the same amount.

In the interpretation that follows any variable that is positively associated with household asset index decreases the poverty of the household, and conversely any variable that is negatively associated with the household asset index increases the poverty of the household. The level of education of the household head is highly significant at all five quantiles of the distribution. In addition, the coefficient increases with increasing the quantiles in all levels of education, where it is the highest at the upper quantile. The asset index is lower at the lower end (10^{th} percentile) and higher in the upper end (80^{th} percentile) in all levels of education. The household headed by an individual with primary, secondary or tertiary education level is found to increase the asset index, as compared to a household headed by a person with no formal education from 0.135 to 6.973, 0.185 to 7.779, 0.322 to 10.13, 0.407 to 11.21 and 0.695 to 15.54 for 0.10, 0.20, 0.40 and 0.50 and 0.80 quantiles respectively.

From Table 7.1, the researchers observe that a household headed by a female is negatively associated with the asset index, as compared to a household headed by a male. It is interesting to note that it decreases with increases from 10^{th} to 50^{th} percentiles. The size of the household is also negatively associated with asset index, but is only significant at the upper quantile (80^{th} percentile) and at the conditional mean from OLS. The place of residence of household is highly associated with household asset index (Table 7.1). From this table, it can be observed that an urban household is positively associated with household asset index in all five quantiles as compared to a rural household, where it increases from 0.424 (p-value$<$ .0001) of 10^{th} percentile to 3.361 (p-value$<$ .0001) of 80^{th} percentile.

From Table 7.1, it can be observed that the province is highly associated with the household asset index; a household from Kigali increases the asset index from lower tail to upper tail as compared to a household from Eastern province, whilst a household from Southern, Western or Northern province decreases the asset index, as compared with a household from Eastern province in all percentiles. It is interesting to note that in all provinces except Kigali, the asset index is higher at the lower quantile and lower at the upper quantile when compared to Eastern province. Whereas Southern province most negatively affects the household asset index. This means that Southern province is the poorest, compared to other provinces.

The quadratic term of household size is statistically significant in all quantiles as well as in OLS. The researchers examined the possible interaction effects and found only one significant interaction between gender of household head and the age of household head. From Figure 7.4 it can be observed that the asset index increases with increasing percentiles, but the effect is not significant at 80$^{\text{th}}$ percentile. Figures 7.1 to 7.4 present a summary of quantile regression results that show quantile regression estimates for the entire distribution and their confidence band.

TABLE 7.1. Quantile regression parameter estimates and OLS for poverty of household

Indicator	Q.10		Q.20		Q.40		Q.50		Q.80		OLS	
Indicator	β	P	β	P	β	P	β	P	β	P	β	P
Intercept	-2.967	< .0001	-2.682	< .0001	-1.977	< .0001	-1.679	< .0001	-0.445	0.023	-1.856	< .0001
Province												
Eastern	reference											
Kigali	1.185	< .0001	1.355	< .0001	3.971	< .0001	4.488	< .0001	6.052	< .0001	4.175	< .0001
South	-0.318	< .0001	-0.359	< .0001	-0.551	< .0001	-0.626	< .0001	-0.739	< .0001	-0.625	< .0001
West	-0.192	< .0001	-0.252	< .0001	-0.342	< .0001	-0.412	< .0001	-0.505	< .0001	-0.266	< .0001
North	-0.207	< .0001	-0.288	< .0001	-0.378	< .0001	-0.435	< .0001	-0.637	< .0001	-0.526	< .0001
Gender of the household head												
Female Male	reference -0.261	< .0001	-0.300	< .0001	-0.479	< .0001	-0.514	< .0001	-0.372	0.0491	-0.580	0.0007
Education of Household head												
No education Primary	reference 0.135	< .0001	0.185	< .0001	0.322	< .0001	0.407	< .0001	0.695	< .0001	0.648	< .0001
Secondary	0.889	< .0001	1.355	< .0001	1.987	< .0001	2.873	< .0001	5.032	< .0001	3.859	< .0001
Higher	6.973	< .0001	7.779	< .0001	10.13	< .0001	11.21	< .0001	15.54	< .0001	11.52	< .0001
Age of the household head	-0.001	0.124	-0.002	0.0883	-0.004	0.0036	-0.004	0.0030	-0.003	0.2055	-0.0001	0.9805
Size of household	0.0041	0.872	0.007	0.8163	-0.048	0.1889	-0.064	0.0751	-0.264	0.0002	-0.093	0.0247
Place of residence												
Rural Urban	reference 0.424	< .0001	0.583	< .0001	1.039	< .0001	1.107	< .0001	3.361	< .0001	2.137	< .0001
Size*Size	0.007	0.0078	0.008	0.0048	0.016	< .0001	0.018	< .0001	0.048	< .0001	0.030	< .0001
Age by gender	0.003	0.0120	0.003	0.474	0.006	0.0048	0.006	0.0030	0.005	0.1686	0.009	0.0116
R^1_T and R^2	0.107		0.147		0.220		0.267		0.446		0.540	

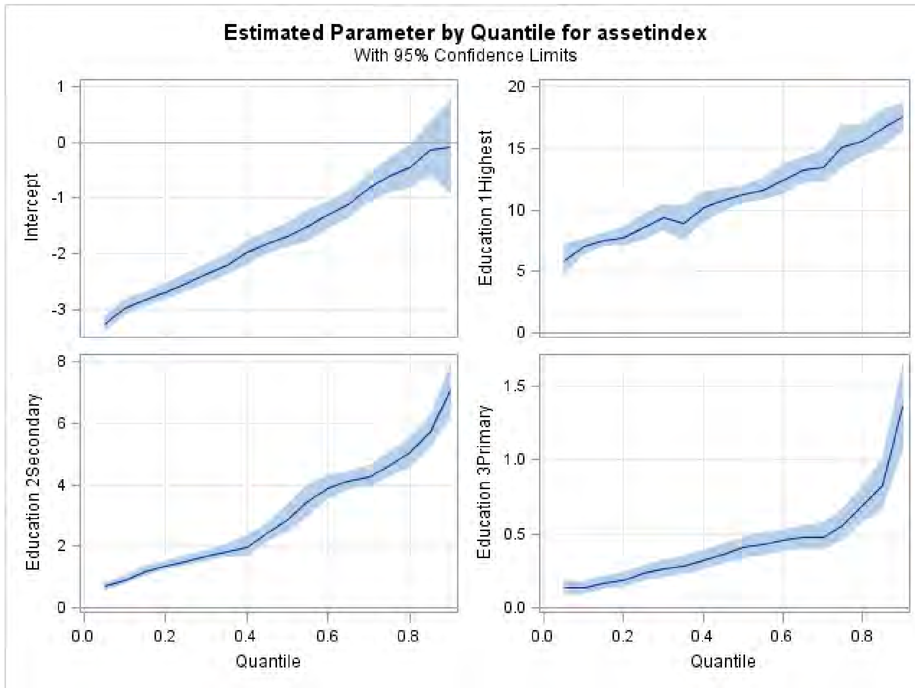


FIGURE 7.1. Summary of quantile regression estimates with 95% confidence bands by education level

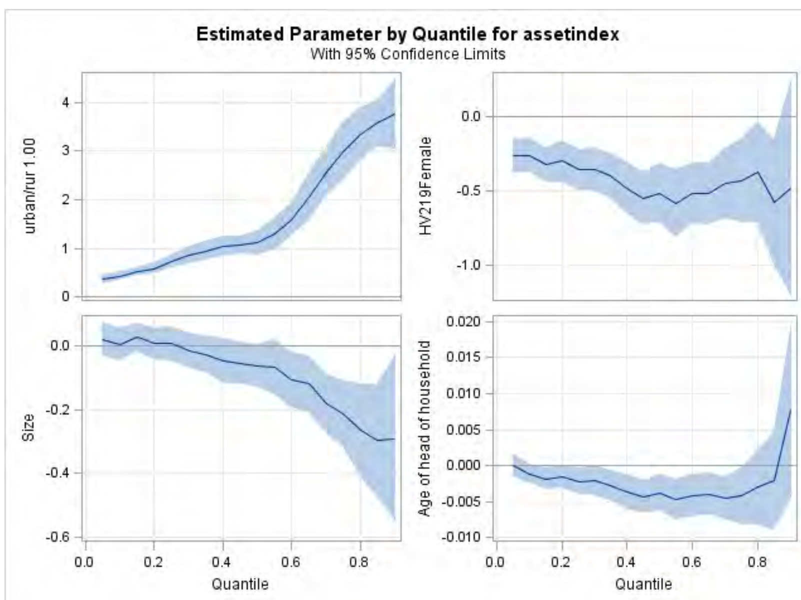


FIGURE 7.2. Summary of quantile regression estimates with 95% confidence bands by place of residence

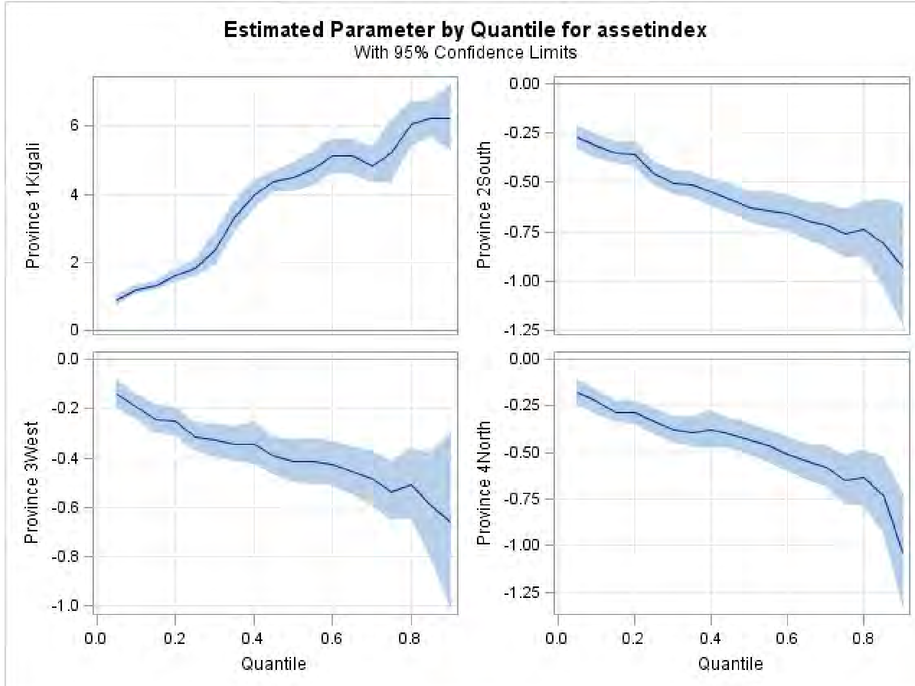


FIGURE 7.3. Summary of quantile regression estimates with 95% confidence bands by province

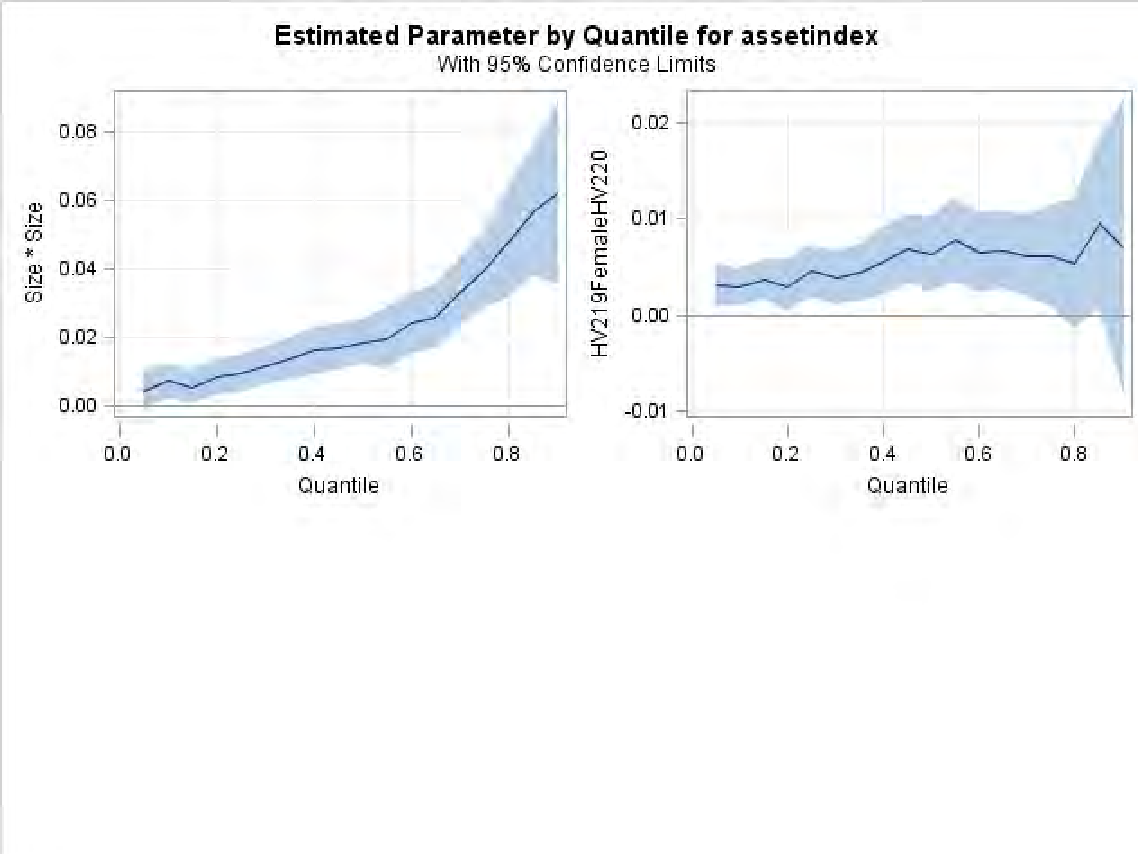


FIGURE 7.4. Summary of quantile regression estimates with 95% confidence bands by family size and gender

7.6.2. Application to Demographic and Health Survey data to identify the factors associated to malnutrition of children under five years.

Introduction

The anthropometric indicators are measured in Z-score for stunting, wasting and underweight and are defined as

$$Z_i = \frac{AI_i - MAI}{\sigma} \tag{7.43}$$

where AI_i refers to the individual anthropometric indicator, MA and σ refer to the median and the standard deviation of the reference population. Note that higher values of Z-scores indicate better nutrition and vice versa. Therefore a decrease of Z-score indicates an increase in malnutrition and vice versa. In a recent study by Habyarimana et al. (2014), we have considered the anthropometric measurements for underweight where the distribution of weight-for-age was categorized as severe underweight, moderate underweight and not underweight. In this chapter we consider the entire distribution of weight-for-age (underweight) and the quantiles for interest are 10, 25, 40, 50, 90, where 10^{th} quantile is the lower tail and 90^{th} quantile is the upper tail. In this analysis we only consider underweight; the analysis for stunting and wasting is done in a similar way.

Results and interpretation

In the interpretation that follows any variable that is positively associated with anthropometric index decreases the malnutrition of the child under five years, and conversely any variable that is negatively associated with the anthropometric index increases the malnutrition of the child under five years. The results are presented in Table 7.2.

The birth order significantly affects the weight-for-age Z-score of the child Table 7.2. However, it is not significant at the bottom of the distribution (10^{th} percentile)and at the top of distribution (90^{th} percentile). Further, the weight-for-age Z-score of the child decreases with increasing the child's birth order Figure 7.5.

Gender of child significantly affects the weight-for-age Z-score of the child. From the same Table, it is observed that the weight-for-age Z-score of the child decreases with increasing the quantiles; this is underestimated by least squares regression. It is observed that province slightly affects Z-score of weight-for-age of the child. However, it is only significant in 20^{th} percentile. Further, a child born in Southern province has a positive Z-score weight-for-age as compared to that of a child born in Eastern province.

Mother's knowledge on nutrition positively affects Z-score of weight-for-age of the child as compared to Z-score of weight-for-age of a child born to mother without some knowledge on nutrition.

Assistance of the mother at the delivery significantly affects the child Z-score. However, it is only significant in 40^{th},50^{th} and 90^{th} quantiles.

Mother's level of education significantly affects the Z-score of the child. The weight-for-age z-score of the child increases with increasing the level of education of the mother. However, the Z-score decreases with increasing the quantiles, this hidden in the case of OLS.

Marital status of the mother significantly affects weight-for-age Z-score of the child. However it is significant only in 40^{th} and 50^{th} quantiles. A child born to a married mother or to a mother living with a partner has a positive weight-for-age Z-score as compared to child born to divorced or separated mother. In addition, weight-for-age Z-score decreases with increasing percentiles from 40^{th} to 50^{th}.

Fever significantly affects the weight-for-age Z-score of the child. It is observed that the weight-for-age Z-score for a child who had fever two weeks before the survey was found to decrease as compared to that of a child who did not have fever in the same time frame. Further, the coefficients decrease with increasing the quantiles.

Anemia is significantly affecting the z-cores of the child. A child born to anemic mother has a negative Z-score as compared to a child born to non anemic mother; this means that a child born to anemic mother is more likely to be underweight than a child born to non anemic mother. However, this effect is only significant in 20^{th} and 50^{th} quantiles.

Birth weight significantly affects the Z-score of the child in all quantiles. The Z-score of a child born with a weight bigger or equal to 2500g are positive as compared to Z-score of a child born with lower birth weight. It is higher in 20^{th} quantile and lower in 10^{th} quantile.

The wealth index of the mother also affects the Z-score of the child. The Z-score of the child increases with increasing the wealth index of the mother. However, it is higher in 40^{th} quantile and lower in 50^{th} quantile and elsewhere is not significant.

Mother's age negatively affects the child's Z-score in lower tail (10^{th} quantile) and is not significant elsewhere Figure 7.7.

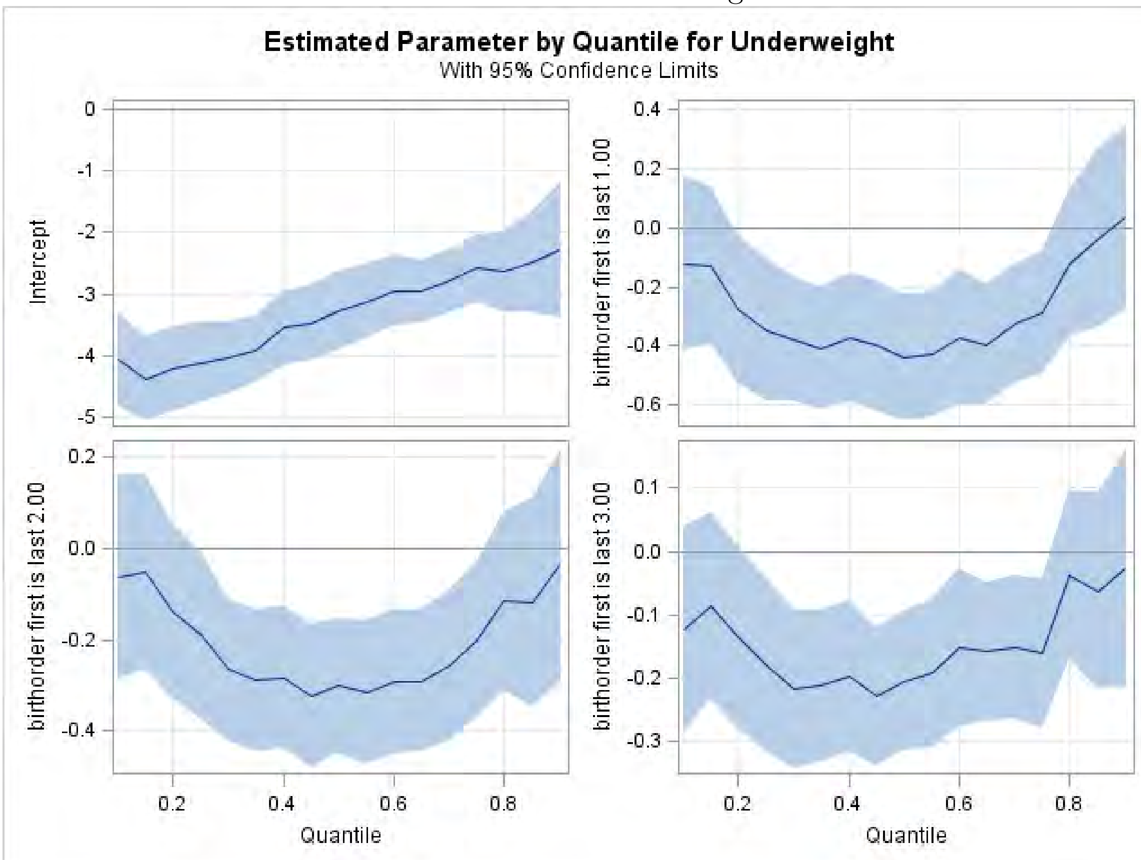
The age of the child significantly affects the Z-score of the child from 40^{th} quantile to 90^{th} quantile. The Z-scores of a child from age group 13 to 23 months and 23 months and more is positively affecting the child Z-score as compared to a child from 0 to 11 month age group. Further, the coefficients of a child belonging to age group 12-23 months are higher than the coefficients of a child aged 23 months and more. However, the coefficients are lower in 40^{th} quantile for a child aged 12-23 and 50^{th} quantile for a child aged 23 months and more Figure 7.8.

Body mass index of the mother significantly affects the Z-score of the child in all quantiles. The effect is smaller in lower tail and higher in upper tail (80^{th} quantile).

Table 7.2. Quantile regression parameter estimates and OLS for underweight

Indicator	Q.10		Q.20		Q.40		Q.50		Q.90		OLS	
Indicator	β	P	β	P	β	P	β	P	β	P	β	P
Intercept	-4.064	< .0001	-4.221	< .0001	-3.543	< .0001	-3.274	< .0001	-2.2279	0.3442	< .0001	
Birth order(6 and more=ref)												
4-5	-0.122	0.4092	-0.274	< .0001	-0.371	0.0014	-0.442	< .0001	-0.033	0.8362	-0.0622	0.03335
2-3	-0.064	0.5751	-0.138	0.158	-0.284	0.0003	-0.301	< .0001	-0.038	0.7594	-0.0400	0.0773
first	-0.124	0.1158	-0.135	0.0933	-0.196	0.0023	-0.205	< .0001	-0.027	0.7719	-0.0531	0.0019
Gender(Male=ref)												
Female	0.202	0.0004	0.174	0.0011	0.134	0.0031	0.096	0.0321	0.209	0.0021	0.0362	
Province												
Eastern	reference											
Kigali	0.0100	0.9520	-0.070	0.5916	-0.054	0.6327	-0.004	0.9691	0.122	0.3955	0.0076	0.7814
South	0.145	0.1236	0.172	0.0471	0.108	0.1467	0.114	0.0914	-0.073	0.5156	0.0421	0.0291
West	0.048	0.5988	-0.029	0.7077	-0.015	0.8258	-0.007	0.9077	-0.114	0.3053	0.021	0.2387
North	0.2131	0.0308	0.116	0.1796	0.043	0.5509	0.042	0.5381	-0.115	0.1870	0.0337	0.0969
Knowledge on nutrition(no=ref)												
Yes	0.152	0.0177	0.196	0.0005	0.0053	0.091	0.0313	0.0709	0.0.113	0.1307	0.0272	0.0404
Assistance(No=ref) Yes	-0.066	0.5950	-0.046	0.5831	0.132	0.0053	-0.148	0.0381	-0.229	0.0119	-0.0255	0.2280
Mother's education level(no education=ref)												
Primary	0.115	0.1670	0.060	0.4609	0.004	0.9463	0.063	0.5726	0.134	1093	0.0.1164	< .0001
Secondary& higher	0.570	< .0001	5138	< .0001	0.219	0.0169	0.242	0.0086	263	0.1628	0.0135	0.4057
Mother's marital status(divorced/separated=ref)												
Widowed	0.055	0.7889	0.176	0.2597	0.017	0.911	0.-0.0587	0.6987	0.303	0.2236	0.0355	0.3402
Married/living with partner	0.143	0.2003	0.435	0.0003	0.2850	0.0174	0.1929	0.1127	0.297	0.1633	0.0689	0.0101
Never in union	-0.274	0.2760	0.561	0.117	0.337	0.1520	0.3929	0.0898	0.577	0.3600	0.06200	< .0001
Had fever last two weeks(yes=ref) No	0.331	0.0003	0.162	0.0345	0.139	0.0257	0.158	0.0056	0.153	0.0.0687	0.0474	0.0042
Source of drinking water(Others=ref)												
Piped into dwelling/yard	0.060	0.8403	0.249	0.2336	0.164	0.3363	3175	0.0539	0.3417	0.071	-0.0365	0.3833
Public tap	-0.056	0.5111	-0.000	0.9959	-0.046	0.4483	-0.009	0.8982	-0.078	0.3737	-0.0381	0.0353
Protected spring/well	0.113	0.0989	0.109	0.0562	0.058	0.2635	0.073	0.1730	0.134	0.0957	-0.0045	0.7650
Anemia(noanemic=ref) Anemic	-0.130	0.0599	-0.122	0.0249	-0.059	0.2048	-0.096	0.0414	-0.096	0.1316	-0.0034	0.0126
Toilet facilities(Yes=ref)												
No	0.125	0.0783	0.0.0937	0.2127	0.015	0.8119	0-0.036	0.5352	-0.122	0.1198	0.0157	0.3281
Birth weight(≥ 2500g=ref)												
less< 2500g	0.398	0.0206	0.636	< .0001	0.615	< .0001	0.509	0.0005	0.509	0.0306	0.1581	< .0001
Wealth index(poor=ref)												
Middle	0.141	0.0710	0.111	0.0867	0.125	0.0319	0.111	0.0678	-0.049	0.6980	0.0279	0.2321
Rich	0.157	0.2178	0.1930	0.1188	0.285	0.0007	0.2350	0.0018	0.006	0.9471	0.0358	0.0254
Mother's age	-0.023	0.0049	-0.014	0.0802	0.005	0.3700	0.004	0.5728	-0.008	0.3718	-0.0001	0.9805
Child age(0-11months=ref												
12-23 months	0.006	0.9715	0.099	0.2965	0.316	< .0001	0.298	< .0001	0.677	< .0001	0.0249	0.1965
23+ months	0.104	0.1344	0.082	0.1778	0.174	0.0016	0.184	0.0003	0.393	< .0001	0.0141	0.3303
BMI	0.063	< .0001	0.067	< .0001	0.068	< .0001	0.067	< .0001	0.079	< .0001	0.0145	< .0001
R_T^1 and R^2	0.107		0.147		0.220		0.267		0.446		0.6207	

Figure 7.5. Summary of quantile regression estimates for the entire distribution and confidence band for underweight



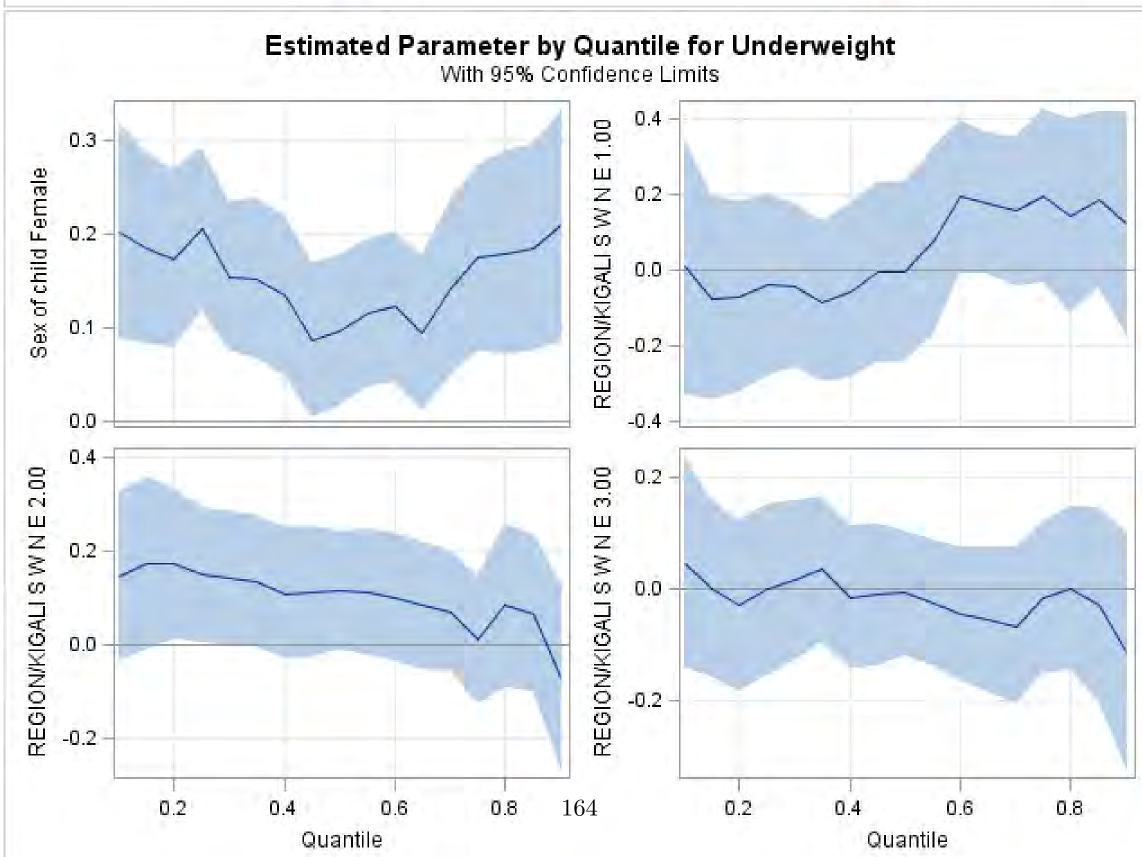


Figure 7.6. Summary of quantile regression estimates for the entire distribution and confidence band for underweight

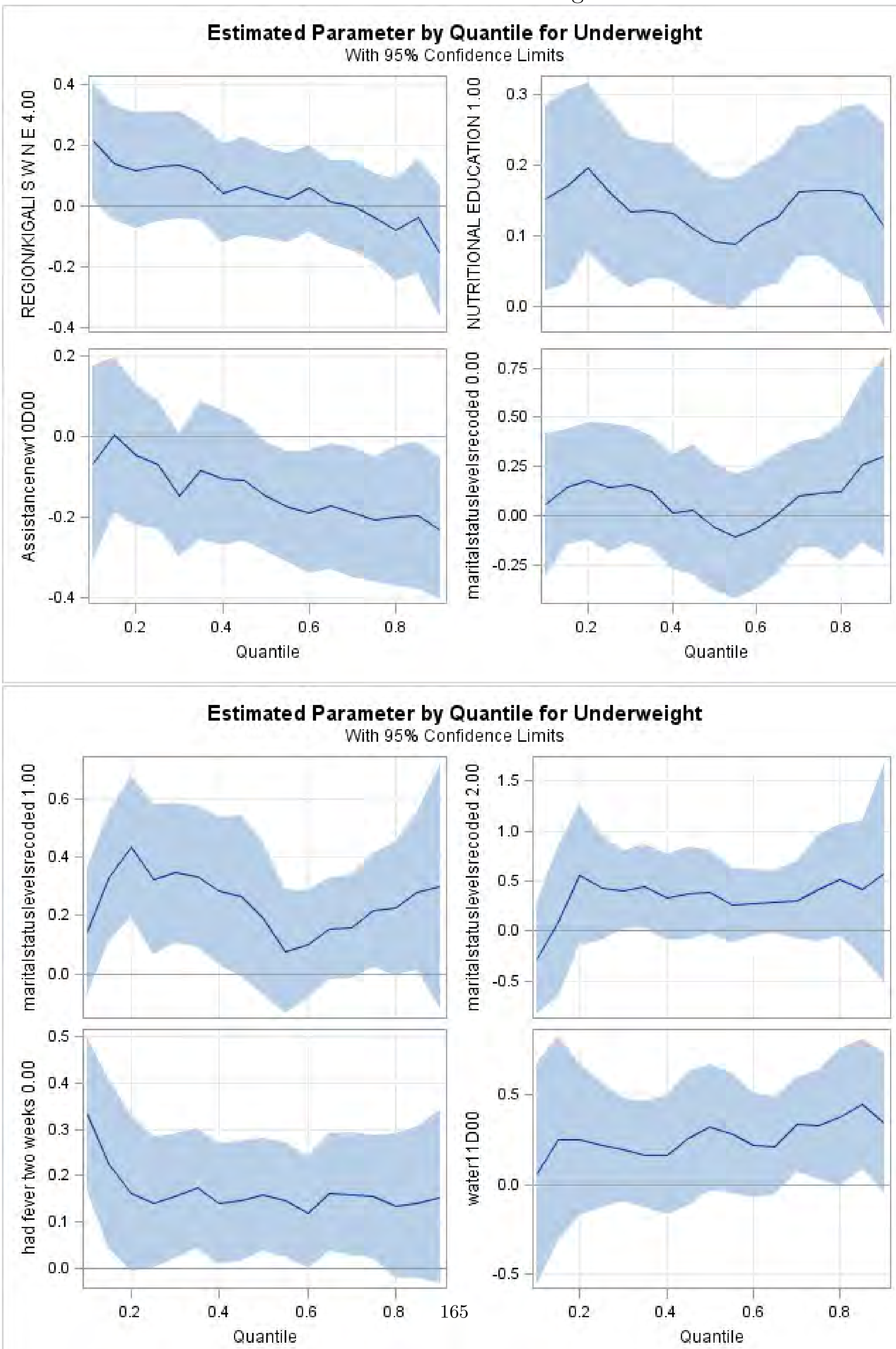


Figure 7.7. Summary of quantile regression estimates for the entire distribution and confidence band for underweight

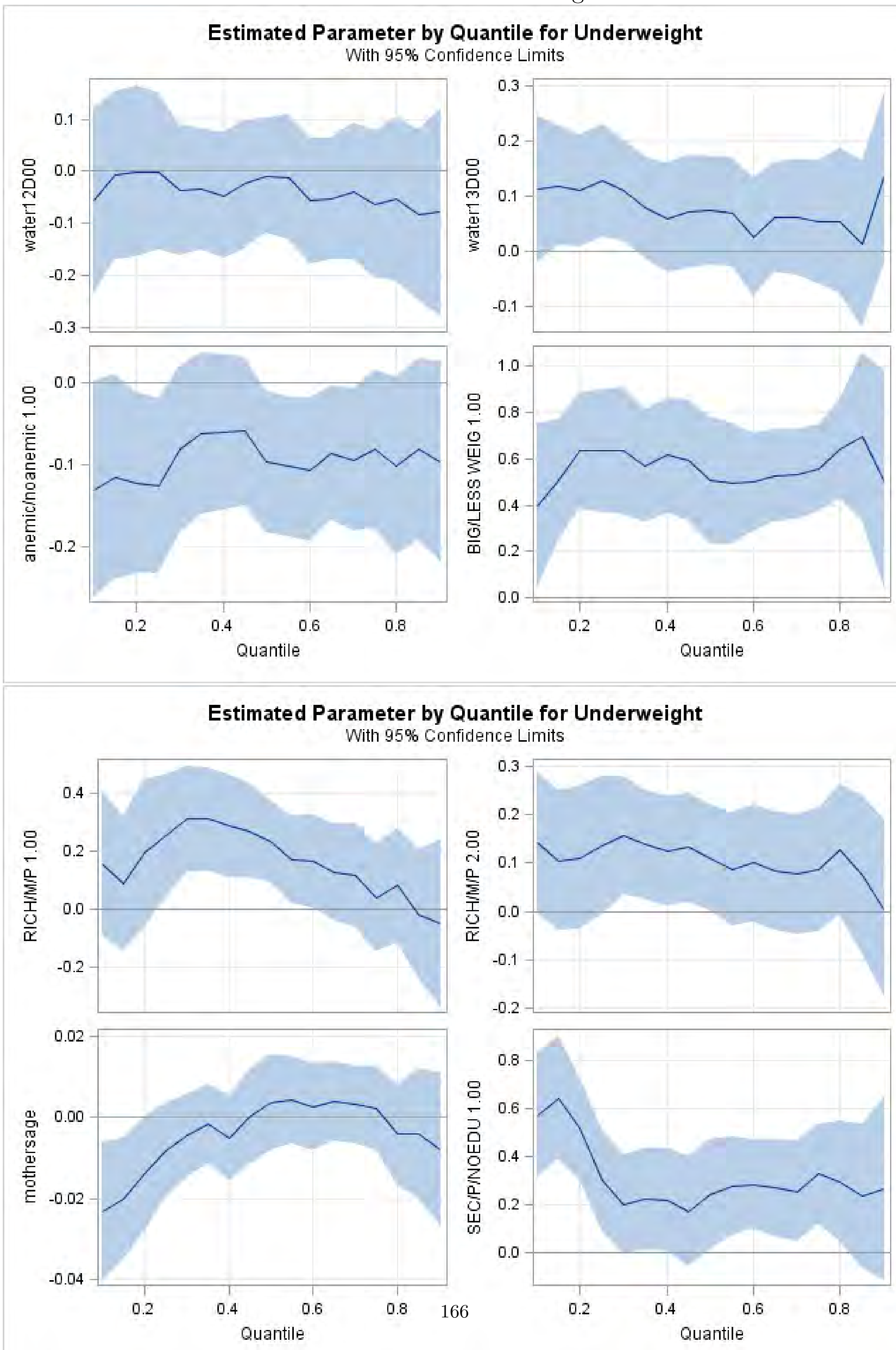
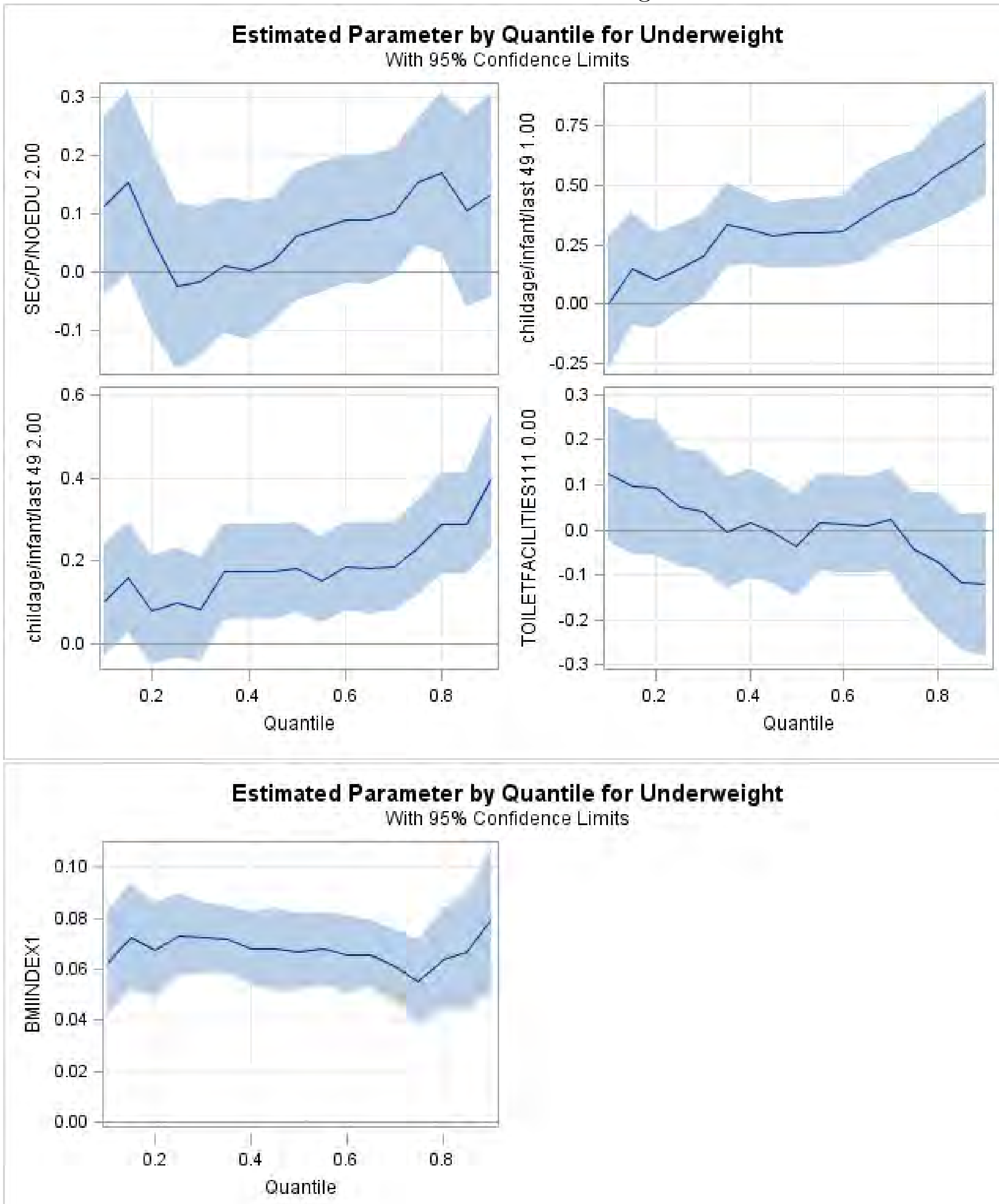


Figure 7.8. Summary of quantile regression estimates for the entire distribution and confidence band for underweight



7.7. Summary

The quantile regression model allows us to study the impact of predictors on different desired quantiles of the response distribution, and therefore to get a complete picture of the relationship between the response variable and the predictor variables. This is one of the drawbacks of OLS and logistic regression. Therefore, quantile regression procedures can reveal information about the dependence of the conditional distribution of the response variable on the predictor variable that are most of the cases hidden by OLS and logistic regression.

Based on the asset index and quantile regression, this chapter identified the determinants of poverty of households in Rwanda. The results confirmed the findings of the previous studies. In both studies, the key determinants of poverty are age of the household head, level of education of household head, gender of household head, place of residence (urban or rural), province of residence and the size of the household(number of the members of household). However, in this study, the findings from quantile regression method are more specific at each quantile of interest,

The level of education of the household head is highly significant at all five quantiles of the distribution. In addition, the coefficient increases with increasing the quantiles in all levels of education, where it is the highest at the higher level of education and in upper quantile. This means that education has a stronger effect on asset index in richer households.

A household headed by a female is negatively associated with the asset index, as compared to a household headed by a male. The size of the household is also negatively associated with the asset index. A household from Kigali was found to increase the asset index, as compared to a household from Eastern province, however, a household from Southern, Western or Northern provinces was found to decrease the asset index, compared to a household from Eastern province. This means that a household from Kigali is less likely to be poor as compared to a household from Eastern province. From Table 7.1, a household from Southern province is seen to most negatively affect the asset index; this shows that this province is the most poor as compared to other provinces. An urban household is positively associated with the asset index, whereas a rural household is negatively associated with asset index.

In malnutrition case, the results from this chapter supported the findings of Habyarimana et al. (2014). However in this chapter, as expected from the theory, it revealed some new information. It was found that some predictor variables were significantly affecting the weight-for-age Z-score of the child in some quantiles but these predictors were not significant in Habyarimana et al. (2014). These predictors are province of birth of the child, wealth index of his/her family and mother's age at the birth. In addition to these predictor variables, the study revealed that the key determinants of underweight among children under five years in Rwanda are birth order of the child, age group of the child, gender of the child, birth weight of the child, fever, mother's level of education, mother's marital status, assistance at the delivery, toilet facilities and source of drinking water. But almost all the results found at 50^{th} quantile are similar to the results from OLS.

CHAPTER 8

Generalized Additive Mixed Models

8.1. Introduction

In previous chapters, we modeled the households data as well as malnutrition data using various statistical models such as: generalized linear models through classical logistic regression and survey logistic regression(binary logistic regression and proportional odds models with complex survey designs) (Habyarimana et al., 2014), generalized linear mixed models, multivariate joint model, spatial multivariate joint model and quantile regression (Habyarimana et al., 2015b). All these models are parametric. The parametric models offer a strong tool for modelling the relationship between the outcome variable and predictor variables when their assumptions meet. However, these models may suffer from inflexibility in modelling complicated relationships between the outcome variable and the predictor variables in some applications and the parametric mean assumption may not always be desirable, as suitable functional forms of the predictor variables may not be known in advance and the response variables may depend on the covariates in a complicated manner (Lin and Zhang, 1999). The generalized additive mixed model (GAMM) relaxes the assumption of normality and linearity inherent in linear regression. The flexibility of nonparametric regression for continuous predictor variables, coupled with linear models for predictor variables, offers ways to reveal structure within the data that may miss linear assumptions. This flexibility of GAMM motivated the current research to use semiparametric logistic mixed model to assess the determinants of poverty of households as well as the risk factors associated to the malnutrition of children under five years. In literature there exists many nonparametric regression models and smoothing methods for independent data. The most commonly used are splines smoothers, kernel smoothers, locally-weighted running-line smoothers and running-mean smoothers. These methods are well detailed in Hastie and Tibshirani (1990); Hardle (1999) and Green and Silverman (1993).

8.2. Generalized additive mixed model

Generalized additive mixed model (GAMM) can be seen as an extension of GAM to incorporate random effect or an extension of generalized linear mixed models (GLMM) of Breslow and Clayton (1993) to allow the parametric fixed effects to be modeled nonparametrically using additive smooth functions in a similar spirit to Hastie and Tibshirani (1990). Suppose that observations of the j^{th} of k units consists of an outcome variable y_j and p covariates $x_j = (1, x_{j1}, ..., x_{jp})^T$ associated with fixed effects and $q \times 1$ of covariates z_j associated with random effects. Therefore, Lin and Zhang (1999) formulated GAMM as follows

$$g(\mu_j) = \beta_0 + f_1(x_{j1}) + ... + f_p(x_{jp}) + z_i b \tag{8.1}$$

where $g(.)$ is a monotonic differentiable link function,$\mu_j = E(y_j|b)$, $f_j(.)$ is a centred twice-differentiable smooth function, the random effect b is assumed to be distributed as $N\{0, K(\vartheta)\}$ and ϑ is a $c \times 1$ vector of variance components.

A fundamental feature of GAMM (8.1) over GAM is that the additive nonparametric functions are used to model covariate effects and random effects are used to model the correlation between observations (Lin and Zhang, 1999; Wang, 1998). If $f_j(.)$ is a linear function, then GAMM (8.1) reduces to generalized linear mixed model (GLMM) of Breslow and Clayton (1993).

For a given variance component ϑ, the log-quasi-likelihood function of $(\beta_0, f_j, \vartheta, j = 1, 2, ..., k)$ is given (Lin and Zhang, 1999) by

$$exp[\iota\{\beta_0, f_1(.), ..., f_k(.), \vartheta\}] \propto |K|^{\frac{-1}{2}} \int exp\left\{\frac{-1}{2\phi}\sum_{j=1}^{k} d_j(y_j; \mu_j) - \frac{1}{2}b'K^{-1}b\right\} db \tag{8.2}$$

where $y_j = (y_1, y_2, ..., y_k)$ and $d_j(y_j; \mu_j) \propto -2\int_{y_j}^{\mu_j} m_j(y_j - u)/v(u)du$ defines the conditional deviance function of $\{\beta_0, f_j(.), \vartheta\}$ given **b**. Statistical inference in GAMM includes inference on the nonparametric functions $f_j(.)$, that needs the estimation of smoothing parameter as well as inference on the variance components ϑ. The linear mixed models and the smoothing spline estimators have close connections (Green and Silverman, 1993; Lin and Zhang, 1999; Verbyla et al., 1999; Wang, 1998).

8.2.1. Natural cubic smoothing spline estimation.

Following the derivation of Greenland et al. (1994) and Lin and Zhang (1999), with a given λ and ϑ, the natural cubic smoothing spline estimators of the $f_j(.)$ maximize the penalized log-quasi-likelihood as follows

$$\iota\{\beta_0, f_1(.), ..., f_k(.), \vartheta\} - \frac{1}{2}\sum_{i=1}^{k}\lambda_i \int_{s_j}^{t_i} f"_i(x^2)dx \tag{8.3}$$

$$= \iota\{\beta_0, f_1(.), ..., f_k(.), \vartheta\} - \frac{1}{2}\sum_{i=1}^{k}\lambda_i f_i^T H_i f_i$$

where (s_i, t_i) defines the range of the i^{th} covariate and λ_i are smoothing parameters that regulate the tradeoff between the goodness-of-fit and smoothness of the estimated functions. In addition, $f_i(.)$ is an $r_i \times 1$ unknown vector of the values of $f_i(.)$, calculated at the r_i ordered distinct values of the $x_{ji}(i = 1, 2, ..., m)$ and H_i is the corresponding nonnegative definite smoothing matrix (Green and Silverman, 1993). GAMM, given in equation (8.1) can be formulated in matrix form as

$$g(\mu_i) = 1\beta_0 + M_1 f_1 + M_2 f_2 + ... + M_k f_k + Zb, \tag{8.4}$$

where $g(\mu_i) = \{g(\mu_1), g(\mu_2), ..., g(\mu_m)\}$, 1 is an $m \times 1$ vector of 1s, M_i is an $k \times r_i$ incident matrix defined in a way similar to that given in Green and Silverman (1994) such that the i^{th} component of $M_j f_j$ is $f_j(x_{ij})$ and $Z_i = (z_1, z_2, ..., z_m)^T$. The numerical integration is needed to estimate equation (8.2) except for Gaussian outcome. The natural cubic smoothing spline estimators of $f_i(.)$, evaluated by explicit maximization of equation (8.4), is sometimes challenging. To solve this problem, Lin and Zhang (1999) proposed the double penalized quasi-likelihood approach as an alternate approximation approach discussed in subsection 8.2.2.

8.2.2. Double penalized quasi-likelihood.

Since f_i is a centred parameter vector, it can be parameterized in terms of β_i and $a_i((r_i - 2)\times)1$ in a one-to-one transformation as

$$f_i = X_i\beta_i + \beta_i a_i, \tag{8.5}$$

where X_i is an $r_i \times$ 1 vector containing the r_h centred ordered distinct values of the x_{ij} $(i = 1, 2, ..., m)$, and $\beta_i = L_i(L_i^T L_i)^{-1}$ and L_i is an $r_i \times (r_i - 2)$ full rank matrix satisfying $H_i = L_i L_i^T$ and $L_i^T X_i = 0$ using the identity $f_i^T H_i f_i = a_i^T a_i$, the double penalized quasi-likelihood with respect to (β_0, f_i) and b is given by

$$-\frac{1}{2\varphi}\sum_{i=1}^{m} d_i(y;\mu_i) - \frac{1}{2}b^T K^{-1} b - \frac{1}{2}a^T \Gamma^{-1} a \tag{8.6}$$

where $a = \left(a_1^T, a_2^T, ..., a_k^T\right)^T$ and $\Gamma = diag\left(\tau_1 I, \tau_2 I, ..., \tau_k I\right)$ with $\tau_i = \frac{1}{\lambda_i}$. A small value of $\tau = (\tau_1, \tau_2, ..., \tau_k)^T$ corresponds to over-smoothing. Plugging equation (8.5) into (8.4), expression (8.4) suggests that given ϑ and τ, the DPQL estimators $\hat{f}_i$ can be obtained by fitting the following GLMM using (Breslow and Clayton, 1993) penalized quasi-likelihood approach:

$$g(\mu) = X\beta + Ba + zb, \tag{8.7}$$

where $X = (1, M_1 X_1, M_2 X_2, ..., M_k X_k)$, $B = (M_1 B_1, M_2 B_2, ..., M_k B_k)$, $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_k)^T$ is a $(k+1) \times 1$ vector of regression coefficients and **a** and **b** are independent random effects with distributions $a \sim N(0, \Gamma)$ and $b \sim N(0, K)$. Therefore DPQL estimator $\hat{f}_j$ is calculated as $\hat{f}_i = X_i \hat{\beta}_i + \beta_i \hat{a}_i$, that is a linear combination of the (Breslow and Clayton, 1993) penalized quasi-likelihood estimators of the fixed effect $\hat{\beta}_i$ and the random effects $\hat{a}_i$ in the working GLMM (8.7). The maximization of the expression (8.6) with respect to (β, a, b) can be proceeded by using the Fisher scoring algorithm to solve

$$\begin{pmatrix} X^T W X & X^T W B & X^T W Z \\ B^T W X & B^T W B + \Gamma^{-1} & B^T W Z \\ Z^T W X & Z^T W B & Z^T W Z + K^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ a \\ b \end{pmatrix} = \begin{pmatrix} X^T W Y \\ B^T W Y \\ Z^T W Y \end{pmatrix}, \tag{8.8}$$

where Y is the working vector defined as $Y = \beta_0 1 + \Sigma_{j=1}^{p} M_i f_i + Zb + \Delta(Y - \mu)$ and $\Delta = diag\left[g'(\mu_i)\right]$, $W = diag\left[\{\vartheta v(\mu_i) g'(\mu_i)^2\}^{-1}\right]$. An examination of the equation (8.8) shows that it corresponds to the normal equation of the best linear unbiased predictors (BLUPs) of β and (a, b) under linear mixed model

$$Y = X\beta_0 + Ba + Zb + \epsilon, \tag{8.9}$$

where a and b are independent random effects with $a \sim N(0, \Gamma)$, $b \sim N(0, K)$ and $\epsilon \sim N(0, W^{-1})$. This suggests that the DPQL estimators $\hat{f}_j$ and the random effects estimators $\hat{b}$ can be easily obtained using the BLUPs by iteratively fitting model (8.9) to the working vector Y (Lin and Zhang, 1999).

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

$$\begin{pmatrix} X^T R^{-1} X & X^T R^{-1} B \\ B^T R^{-1} X & B^T R^{-1} B + \Gamma^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ a \end{pmatrix} = \begin{pmatrix} X^T R^{-1} Y \\ B^T R^{-1} Y \end{pmatrix}, \tag{8.10}$$

where $R = W^{-1} + ZKZ^T$. Denoting by H the coefficient matrix on the left hand side of the equation (8.10) and $H_0 = (X, B)^T R^{-1} (X, B)$, the approximate covariance matrix of $\hat{\beta}$ and $\hat{a}$ is $cov(\hat{\beta}, \hat{a}) = H^{-1} H_0 H^{-1}$. It follows that the approximate covariance matrix of $\hat{f}_j$ is $(X_j, B_j)\, cov(\hat{\beta}, \hat{a})\, (X_j, B_j)^T$, where $cov(\hat{\beta}, \hat{a})$ can be easily found from the corresponding blocks of $H^{-1} H_0 H^{-1}$. It is assumed that the $\hat{f}_j(.)$ are smooth functions in calculating the covariances of the $\hat{f}_j$.

8.3. Estimating parameters and variance components

Previously, it was assumed that the smoothing parameters λ and the variance component ϑ are known when estimation was made on nonparametric function f_j. However, they usually need to be estimated from the data. Under the classical nonparametric regression model

$$y = f(X) + \epsilon, \tag{8.11}$$

where ϵ are independent random errors distributed as $N(0, \sigma^2)$, Whaba (1985) and Kohn et al. (1991) proposed to estimate the smoothing parameter λ by maximizing a marginal likelihood. The marginal likelihood of $\tau = \frac{1}{\lambda}$ is constructed by assuming that $f(X)$ has a prior specified in the form of equation (8.5) with $a \sim N(0, \tau I)$ and a flat prior for β and integrating out a and β as follows:

$$\exp\left\{\iota_M\left(y; \tau, \sigma^2\right)\right\} \propto \tau^{\frac{1}{2}} \int exp\left\{\iota(y; \beta, a, \sigma^2) - \frac{1}{2\tau} a^T a\right\} da d\beta, \tag{8.12}$$

where $\iota(y; \beta, a, \tau^2)$ is the log-likelihood of f under model (8.11). Robinson (1991) and Silverman (1985) pointed out that the marginal likelihood (8.12) of τ is indeed the restricted maximum likelihood (REML) under the linear mixed model

$$y = 1\beta_0 + X\beta_1 + Ba + \epsilon, \tag{8.13}$$

where $a \sim N(0, \tau I)$ and $\epsilon \sim N(0, \sigma^2 I)$ and B was defined earlier; τ is regarded as covariance component. Hence the marginal estimator of τ is a REML estimator.

Kohn et al. (1991) found that the maximum marginal likelihood estimator of τ can sometimes perform better than the generalized cross validation (GCV) estimator in estimating nonparametric function.

Zhang et al. (1998) extended these results to estimate the smoothing parameter λ and variance component ϑ jointly using REML in case of longitudinal data with normally distributed outcome and a nonparametric mean function and their model is formulated as follows

$$y = f(X) + Zb + \epsilon, \tag{8.14}$$

where $f(X)$ denotes the values of nonparametric function $f(.)$ evaluated at the design points of $X_{(m\times 1)}, b \sim N(0, K(\vartheta))$and $\epsilon \sim n(0, V(\vartheta))$. When $f(.)$ is estimated using a cubic smoothing spline (8.5), Zhang et al. (1998) rewrote the model (8.14) as a linear mixed model

$$y = 1\beta_0 + X\beta_1 + Ba + Zb + \epsilon, \tag{8.15}$$

where $a \sim N(0, \tau I)$ and distribution of $\mathbf{b}$ and ϵ are the same as those in model (8.14). They therefore proposed τ as an extra variance component in addition to ϑ in model (8.15) and to estimate ϑ and τ jointly by using REML. In this case, REML corresponds to the marginal likelihood of (τ, ϑ) constructed by assuming that f takes the form of (8.5) with $a \sim N(0, \tau I)$ and a flat prior for β and integration out a and β as follows:

$$\exp\{\iota_M(y; \tau, \vartheta)\} \propto K^{\frac{-1}{2}} \tau^{\frac{-1}{2}} \int \exp\left\{\iota(y; \beta, a, b) - \frac{1}{2} b^T K^{-1} b - \frac{1}{2\tau} a^T a\right\} dbdad\beta, \tag{8.16}$$

where $\iota(y; \beta, a, b) = \iota(y; f, b)$ is the conditional likelihood (normal) of f given the random effects $\mathbf{b}$ under the model (8.14). Note that the marginal log-likelihood $\iota_M(y; \tau, \theta)$ in (8.16) has a closed form. Whaba (1985) and Zhang et al. (1998) proposed to extend the marginal likelihood approach to GAMM (8.4) and to estimate τ and ϑ jointly by maximizing a marginal quasi-likelihood. Specifically, the GLMM representation of GAMM in (8.7) suggests that τ may be treated as extra variance components in addition to ϑ. Similarly to REML (8.16) the marginal quasi-likelihood of (τ, ϑ) can be constructed under the GAMM (8.4) by assuming that f_j takes the form (8.5) with $a_j \sim N(0, \tau_j I)(j = 1, 2, ..., p)$ and integrating a_j and β out as follows:

$$\exp\{\iota_M(y;\tau,\vartheta)\} \propto |\Lambda|^{\frac{-1}{2}} \int \exp\left\{\iota(y;\beta,a,\vartheta) - \frac{1}{2}a^T\Gamma^{-1}a\right\} da d\beta \tag{8.17}$$

$$\propto |K|^{\frac{-1}{2}}|\Gamma|^{\frac{-1}{2}} \int \left\{\sum_{i=1}^{n} -\frac{1}{2\phi}d_i(y_i;\mu_i) - \frac{1}{2}b^T K^{-1} b - \frac{1}{2}a^T\Gamma^{-1}a\right\}$$

where $\iota(y;\beta,a,\vartheta) = \iota(y;\beta_0,f_1,f_2,...,f_k,\vartheta)$ was defined in (8.2). Based on the Gaussian nonparametric mixed model (8.14) the marginal quasi-likelihood reduces to the Gaussian REML (8.16). An evaluation of the marginal quasi-likelihood (8.16) for non Gaussian outcomes is humped after intractable numerical integration. The Laplace's approximation method is an alternative method used to circumvent this problem. Specifically, taking the quadratic expansion exponent of the integrand of the expression (8.18) about its mode before integration and approximating the deviance statistic $d_i(y;\mu_i)$ by the Pearson χ^2-statistic (Breslow and Clayton, 1993), then the approximate marginal log-quasi-likelihood is given by

$$\iota_M(y;\tau,\vartheta) \approx -\frac{1}{2}log|V| - \frac{1}{2}log|X^TV^{-1}X| - \frac{1}{2}(Y - X\hat{\beta}^TV^{-1})(Y - X\hat{\beta}), \tag{8.18}$$

where $V = B\Gamma B^T + ZKZ^T + W^{-1}$. The equation (8.18) corresponds to the REML log-likelihood of the working vector y under the linear mixed model (8.9) with both **a** and **b** as random effects and τ and ϑ as variance components. Therefore τ and ϑ can be estimated by iteratively fitting model (8.9) using REML.

8.4. Application to the determinants of poverty of household in Rwanda

Introduction

In previous studies Habyarimana et al. (2015a), Habyarimana et al. (2015b) and in chapter 4, we have used GLMM. However, all these studies are based on parametric models. The main aim of this study is to model the effects of age of household head and the interaction of gender and age of household head nonparametrically while other covariates remain parametric using generalized additive mixed models.

8.4.1. Model fitting and interpretation of the results.

The various procedures for estimation discussed for fitting GAMM can be used when fitting the semiparametric logistic mixed model (8.19). The library mgcv from R package was used to fit the data. R package has many options for controlling the model smoothness, using splines such as cubic smoothing splines, locally-weighted running line smoothers, and kernel smoothers. For more details, see the following authors: Ruppert et al. (2003); Green and Silverman (1993); Hardle (1999) and Hastie and Tibshirani (1990). The shrinkage smoothers have several advantages, for instance, helping to circumvent the knot placement. In addition, the method is constructed to smooth any number of covariates. Moreover, the creation of shrinkage smoothers is made in a way that smooth terms are penalized away altogether (Wood, 2006). In this study, the main effect is considered, and also possible two-way interaction effects, where the AIC of each model is examined, the inference of smooth function and the p-value of the individual smooth term. Finally, the model with smaller AIC and higher value of degree of freedom and highly statistically significant was selected as follows

$$\begin{aligned} g(\mu_j) &= \beta_0 + \beta_1 Education_j + \beta_2 Gender_j + \beta_3 Place\ of\ residence_j \\ &+ \beta_4 Province_j + \beta_5 Size_j + \beta_6 Province_j * Place\ of\ residence_j \\ &+ f_1(Age_j) + f_2(Age_j) * Gender_j + b_{0j} \end{aligned} \tag{8.19}$$

where $g(\mu_i)$ is the logit link function, $\beta's$ are parametric regression coefficients, $f_j's$ are centered smooth functions and b_{0i} is the random effect distributed as $N(0, K(\vartheta))$. The common widely used methods for estimating additive models include cubic smoothing splines, locally-weighted running line smoothers, and kernel smoothers (Hardle, 1999; Hastie and Tibshirani, 1990; Ruppert et al., 2003).

The results from model (8.19) are presented in Tables 8.1, 8.2 and 8.3 and in Figure 8.1 and Figure 8.2.

From Table 8.1, it is observed that the level of education of the household head significantly affects the socio-economic status of the household, where the poverty of the household increases by decreasing the level of education of the household head. Furthermore, it is observed that a household with a household head with secondary education, primary education or no formal education is 4.1850 ($e^{1.4315}$), 14.2008 ($e^{2.6533}$) or 24.5154 ($e^{3.1993}$) respectively, times more likely to be poor as compared to a household headed by a household head with tertiary education.

A household from an urban area is 0.7703 ($e^{-0.2061}$) times less likely to be poor than a household from a rural area.

The size of the household significantly affects the socio-economic status of the household, also shown in Table 8.1. A family of four members or less is 0.6433 ($e^{-0.4411}$)times less likely to be poor than a family of five members or more Table 8.1.

Interaction effects

In this study, not only are the main parametric effects considered, but the two-way interaction effects are also considered. Of interest are the interaction effects between province or region and place of residence (urban or rural). Figure 8.1 shows that in all provinces a rural household is more likely to be poor as compared to an urban one. In the same figure, it is observed that there is a big gap, in terms of poverty, between a rural and urban household from Southern province and Western province. However, this gap is smaller in Kigali and Eastern province.

Approximate smooth function

In Figure 8.2, the estimated smoothing components for household socio-economic status are observed. The Y-axis represents the contribution of smooth function to the fitted values for household socio-economic status. In each figure, the smooth curve denotes the estimated trend of GAMM; **s** is a smooth term and the number in parentheses represents the estimated degree of freedom (edf). The effects of age and gender (female) on household socio-economic status is presented in Figure 8.2 B; the trend shows that the poverty of a household headed by a younger female increases with the age of the household head to approximately 35, and from there, the poverty decreases up to the age of approximately 60 years. The test statistics is 2.110 with 3.7492 degrees of freedom with a high significance (p-value=0.000184) against the assumption that the interaction of age and female gender is linearly associated to the socio-economic status of the household Table 8.3. In Figure 8.2 panel D the poverty of a household headed by young male decreases with increasing age up to approximately 30 years old. However, the poverty decreases with the increasing age of the head from approximately 35 to 60 years old. In addition, from 60 years of age, the poverty of a household increases with the increasing age of the household head regardless of the gender of the household head. The statistic test is 1.484 with 4.0044 degrees of freedom (p-value=0.004930) against the assumption that the interaction of age and male gender is linearly associated to the socio-economic status of the household.

Table 8.1. The parameter estimates of the poverty of households for the fixed part of GAMM

Variables	Estimate	S.E	t-Value	P-value
Intercept	-2.9738	0.5666	-5.249	1.56e-07***
Education(Tertiary=ref				
Secondary	1.4315	0.5675	2.523	0.011663*
Primary	2.6533	0.5608	4.732	2.25e-06***
No education	3.1993	0.5618	5.694	1.27e-08***
Province (Eastern=ref)				
Kigali	-1.1111	0.3021	-3.678	0.000236***
South	0.9197	0.1094	8.409	<2e-16***
West	0.5754	0.1113	5.168	2.40e-07***
North	0.6429	0.1214	5.297	1.19e-07***
Gender (female=ref)				
Male	-004408	0.0462	-9.550	<2e-16***
Place of residence(rural=ref)				
Urban	-0.2061	0.3814	-0.540	0.588879
Size of household(> 4=ref)				
1-4 member(s)	-0.4398	0.0448	-9.810	<2e-16**

FIGURE 8.1. Log odds associated with asset index and province with place of residence (urban or rural)

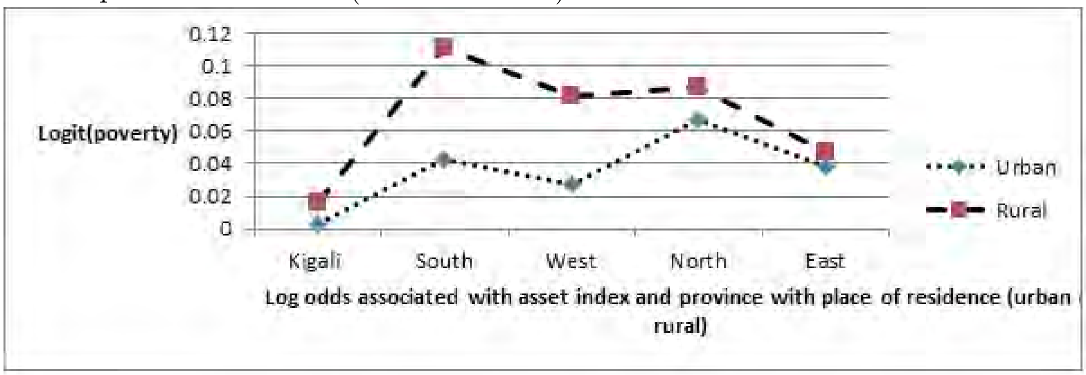


FIGURE 8.2. Smooth function of household socio-economic status with age by gender and confidence interval

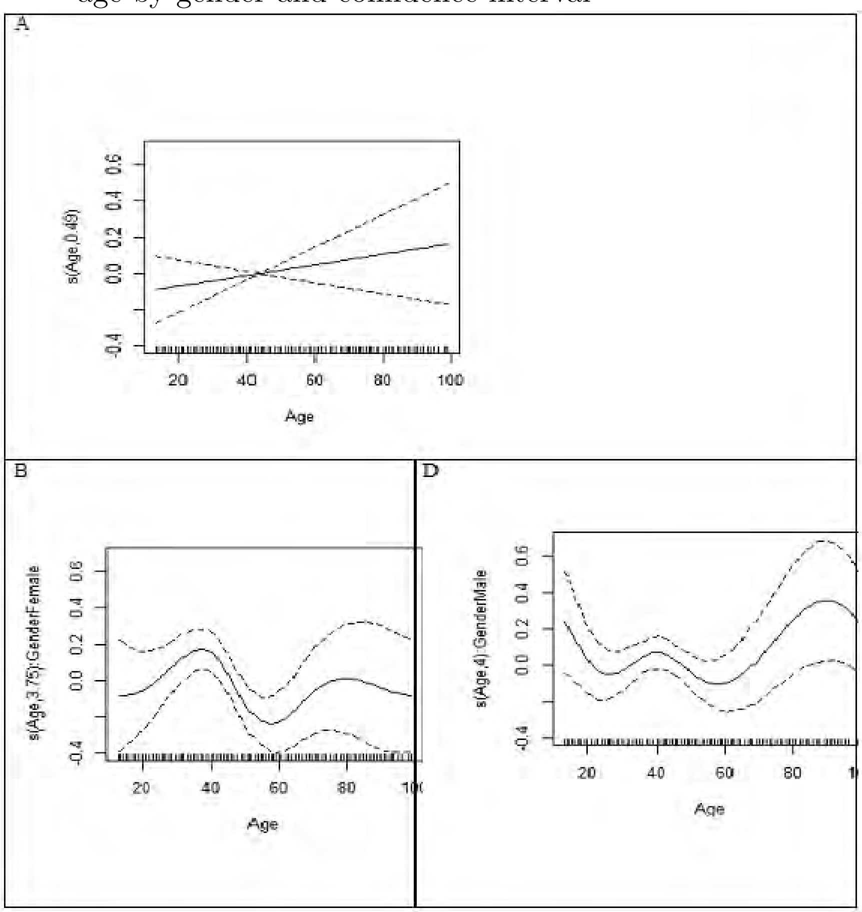


TABLE 8.2. The parameter estimates of GAMM in two way interaction effect for poverty of households

Variables	Estimate	S.E	t-Value	P-value
Province and place of residence				
Eastern and rural=ref				
Kigali and urban	-1.4308	0.5232	-2.735	0.006253**
South and urban	-0.8129	0.4436	-1.831	0.067144.
West and urban	-0.9656	0.5530	-1.746	0.080820.
North and urban	-0.0693	0.57064	-0.121	0.903381

TABLE 8.3. Approximate significance of the smooth term

Smooth terms	Edf	F-value	P-value
S(Age)	0.4882	0.0318	0.062208.
S(Age):Female	3.7492	2.110	0.000184***
S(Age):Male	4.0044	1.484	0.004930**

8.5. Application to the determinants of risk factors of malnutrition of children under five years: case of Rwanda

Introduction

In previous studies (Habyarimana et al., 2015a), we used proportional odds models with complex sampling design, and in chapter 5, we used multivariate joint model to simultaneously identify the risk factors of height-for-age, weight-for-height and weight-for-age; in chapter 6 spatial multivariate joint model is used, and in chapter 7 quantile regression is used. However, all these models are parametric models and sometimes they may suffer from inflexibility in modelling complicated relationships between outcomes variables. In this study the main objective is to model the effect of body mass index of mother and child's age nonparametrically while keeping other covariates parametric using generalized additive mixed model.

The analysis of the data and the model testing was done in mgcv from R package.

8.5.1. Results and interpretations.

The main effect is considered, and also possible two-way interaction effects, where the AIC of each model is examined and also the inference of smooth function and the p-value of the individual smooth term. Finally, the model with smaller AIC and higher value of degree of freedom and highly statistically significant was selected as follows:

$$\begin{aligned} g(\mu_j) &= \beta_0 + \beta_1 Education_j + \beta_2 Gender\ of\ the\ child_j \\ &+ \beta_3 Marital\ status\ of\ the\ mother_j + \beta_4 Multiple\ birth_j + \beta_5 Anemia_j \\ &+ \beta_6 Birth\ order_j + \beta_7 Kwoldge\ on\ nutrition_j + \beta_8 Fever_j \\ &+ f_1(Age\ of\ child_j) + f_2 BMI_j\ of\ the\ mother_j + b_{0j} \end{aligned} \tag{8.20}$$

where $g(.)$ is the logit link function, $\beta's$ are parametric regression coefficients, $f_j's$ are centered smooth functions and b_{0i} is the random effect distributed as $N(0, K(\vartheta))$.
The results from model (8.20) are presented in Table 8.4, Table 8.5 and Figure 8.3.
It is observed that the mother's education level significantly affects weight-for-age (underweight) of the child Table 8.4. Underweight reduces with increasing the level of education of the mother. The degree to which a child is underweight decreases with an increase in the mother's level of education. Further, a child born to a mother with primary education or a secondary or higher education level is 0.12916 ($e^{-2.0467}$) (p-value= 1.48e-06***) or 0.10105 ($e^{-2.2921}$)(p-value= $1.68e-07***$) less likely to have an underweight status than a child born to mother with no education, respectively.
A child born to a widow is 1.94391($e^{0.6647}$) with (p-value=0.001802**) times more likely to be underweight than a child born to a mother who has never been in union.
The gender of a child significantly affects the weight-for-age of the child Table. A male child is 1.51286 ($e^{0.4140}$) (p-value= 0.000205***) times more likely to be underweight than a female child.
Incident of fever significantly affects weight-for-age of the child. A child who did not have fever in the two weeks prior to the survey is 0.62556($e^{-0.4691}$) (p-value=0.000755***) times less likely to be underweight than a child who had a fever

during the same time frame. The birth weight significantly affects weight-for-age of the child. A child born with low birth weight (< 2500g) is 3.00928($e^{1.1017}$) with (p-value=5.30e-06***) times more likely to be underweight than a child born with a higher weight(≥ 2500g).

The mother's knowledge on nutrition significantly affects weight-for-age of the child. A child born to a mother without knowledge of nutrition is 1.35256($e^{0.3020}$) with (p-value=0.015787*) times more likely to be underweight than a child born to a mother with some knowledge of nutrition.

Incident of anemia significantly affects weight-for-age of the child. A child born to an anemic mother is 1.43763($e^{0.3630}$)(p-value=0.001680**) times more likely to be underweight than a child born to a non-anemic mother.

Multiple birth significantly affects the weight-for-age of the child. The degree of underweight increases with increasing the incident of multiple birth. A child born as the first multiple (twin) is 04014($e^{-0.9128}$) times less likely to be underweight than a child born at second or more multiple with (p-value=0.021325*). Whilst, a child born as singleton is 0.26232($e^{-1.3382}$) with (p-value=0.0011625**) times less likely to be underweight than a child born at the second or multiples.

The birth order significantly affects the weight-for-age of the child. A fourth or fifth born child is 0.47043 ($e^{-0.7541}$) (p-value=4.32e-05***) times less risk of being underweight than a sixth or later born child. Similarly, a second or third born child is 0.46213 ($e^{-0.7719}$) times less likely to be underweight than a sixth or later born child. Further, a first born child is 0.34480 ($e^{-1.0648}$ (p-value=1.09e-08***) times less likely to be underweight than a sixth or later born child.

Approximate smooth function: In this study we have also fitted continuous covariates (age of the child and body mass index of the mother nonparametricaly. From Figure 8.3, the estimated smoothing components of weight-for-age status are observed. The Y-axis represents the contribution of smooth function to the fitted values for weight-for-age status. In each figure, the smooth curve denotes the estimated trend of GAMM; **s** is a smooth term and the number in parentheses represents the estimated degree of freedom (edf). The statistic test is 2.673 (p-value=0.000112) with 4.300 degrees of freedom against the assumption that the age of the child is linearly associated to underweight status Table 8.5. The test statistics for BMI is 8.018 (p-value=5.91e-16***) that is highly significant with 3.016 degrees of freedom against the assumption that the body mass index of the mother is linearly associated to underweight status.

From Figure 8.3 we observe that underweight increases with increasing BMI of the mother up to approximately 20 and thereafter it decreases. This is in line with common knowledge on the effect of BMI of the mother where a child born to a normal or obese mother ($18.5 \leq BMI \leq 25.5$) is better than a child born to underweight (thin) mother ($BMI < 18.5$). Child's age significantly affects the weight-for-age of the child Figure 8.3.

It is observed from the same figure that malnutrition increases with increasing age of child from 0 to approximately 12 months (one year) and then decreases with increasing age up to 26 months. From 26-36 months it is increasing with increasing age and then underweight sharply decreases with increasing age up to 48 months and thereafter it increases up to 59 months.

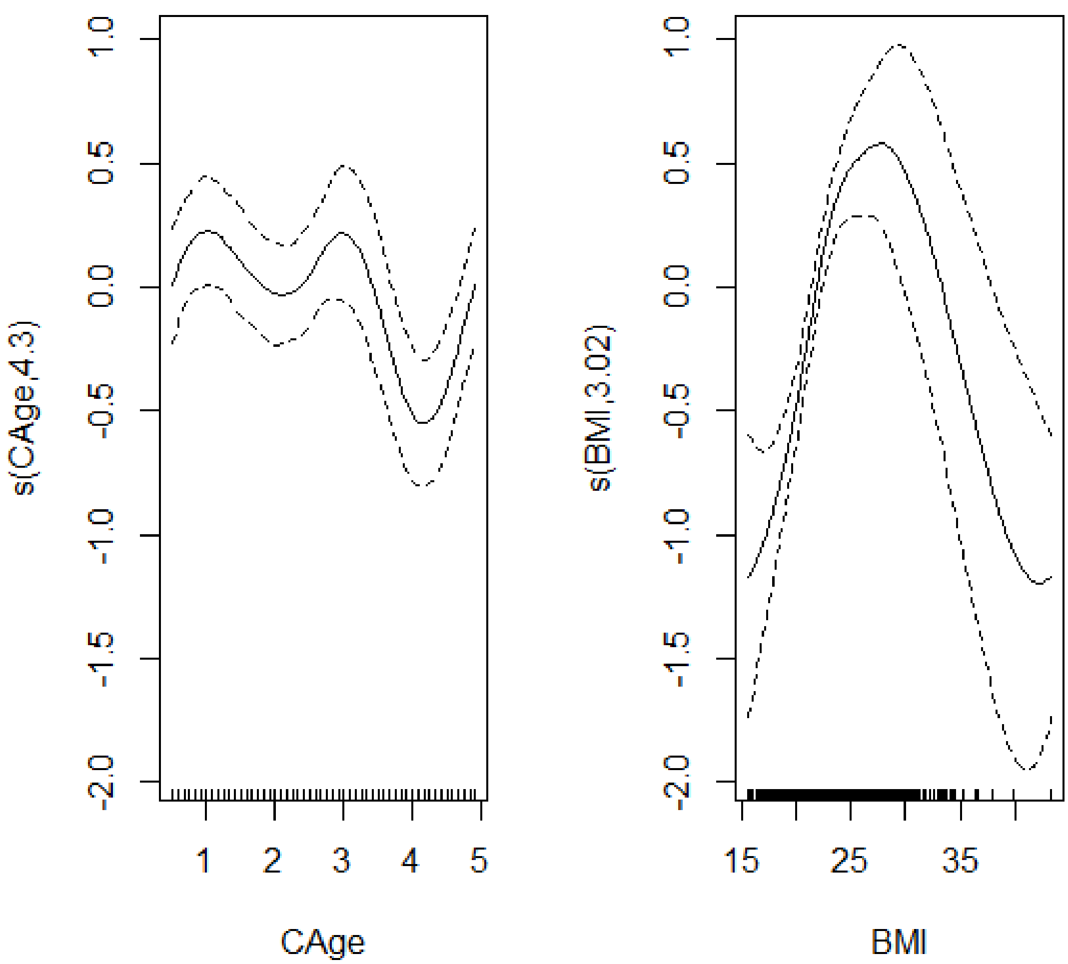
Table 8.4. The parameter estimates of the fixed part of GAMM for malnutrition (underweight) of children under five years

Variables	Estimate	S.E	t-Value	P-value
Intercept	2.7987	0.5330	5.251	1.62e-07***
Mother's education(no education=ref)				
Secondary &higher	-2.2921	0.4370	-5.245	1.68e-07***
Primary	-2.0467	0.4242	-4.825	1.48e-06***
Marital status (never in union=ref)				
Married/living with partner	0.4079	0.3939	1.036	0.300503
Widowed	0.6647	0.2128	3.124	0.001802*
Divorced/separated	0.2134	0.3120	0.684	0.494100
Gender of child (female=ref)				
Male	0.4140	0.1114	3.718	0.000205***
Had fever(yes=ref)				
No	-0.4691	0.1391	-3.373	0.000755***
Birth weights($\geq$ 2500g=ref)				
$<$ 2500g	1.1017	0.2415	4.561	5.30e-06***
Knowledge on nutrition(Yes=ref)				
No	0.3020	0.1250	2.415	0.015787*
Anemia (Anemic mother=ref)				
No anemic	0.3630	0.1154	3.145	0.001680**
Multiple birth(second multiple and more=ref)				
First multiple	-0.9128	0.3963	-2.303	0.021325*
Singleton	-1.3382	0.4242	-3.154	0.001625**
Birth order(6 and more=ref)				
4-5	-0.7541	0.1841	-4.096	4.32e-05***
2-3	-0.7719	0.1643	-4.700	2.73e-05***
1	-1.0648	0.1857	-5.733	1.09e-08***

Table 8.5. Approximate significance of the smooth term

Smooth terms	Edf	F-value	P-value
S(Age of child)	4.300	2.673	0.000112***
S(BMI of the mother)	3.016	8.018	5.91e-16e-16***

Figure 8.3. Smooth function of underweight with age of the child and BMI of the mother



8.6. Summary

In this chapter, we used GAMM to identify the risk factors associated to the poverty of households as well as malnutrition of children under five years.

The results from generalized additive mixed models validate the results from previous other models fitted to the households data as well as malnutrition data. Furthermore, the results from GAMM give more insight (understanding) concerning especially the distribution of continuous covariates.

In the case of household data, the results from parametric part supported that poverty is higher among rural households than urban households. The results from this study also confirmed that poverty decreases with increasing the level of education of household head. In addition, the findings from this study also supported that poverty of household increases with increasing the number of household members. The results from nonparametric part of the model support that the poverty is higher among the households headed by female. However, the use of semiparametric logistic mixed model revealed that it is only true when both male and female are young (approximately up to 35 years old) and this finding is hidden when parametric model is used. Otherwise, the household headed by a female is slightly better off than a household headed by a male.

In malnutrition case, the results confirmed the findings from previous studies especially the parametric part model. The results from this study confirmed that underweight decreases with increasing the mother's level of education. The gender of the child significantly affects the weight-for-age of the child in such a way that the prevalence of underweight is higher among male children than female children. Further, this study confirmed that birth weight significantly affects weight-for-age of the child, where the prevalence of underweight is higher among children born with low birth weight. It has also supported the previous findings where the prevalence of underweight increases with increasing birth order.

The results from nonparametric part model also validated the findings on child's age. However, the results of BMI revealed that underweight increases with increasing the BMI of the mother up to approximately 20 and thereafter it decreases.

CHAPTER 9

Joint modelling of poverty of households and malnutrition of children under five years

In previous chapters we have measured poverty of households and malnutrition separately. We have identified the factors associated with malnutrition using separate anthropometric indices (Habyarimana et al., 2014) or using a multivariate joint model of three anthropometric indices (Habyarimana et al., 2015d,e). In addition we have also studied the poverty of households separately to malnutrition of children under five years (Habyarimana et al., 2015a,b,c). In this chapter we are interested in creating a composite index from the classical three anthropometric indices as an alternative for measuring malnutrition and thereafter use it to study jointly poverty and malnutrition of children under five years. According to our knowledge there is no current study in literature using DHS data for studying the correlation between malnutrition of children under five years and poverty of households.

9.1. Composite index of malnutrition

In this section, based on the principal component analysis technique, we create a composite index from the three commonly used anthropometric indices known as height-for-age (stunting), weight-for-age (underweight) and weight-for-height (wasting). The theory of principal component technique discussed in chapter 2 is also used to compute the composite index of malnutrition. We have used SPSS 22 to compute the index and the results are in Table 9.1 and Table 9.2. We observe from Table 9.1 that the first component alone explains 99.386 % of the total variation of all anthropometric indices and Table 9.2 presents Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy which is good and Bartlett's test of sphericity which is significant. The scree plot presented in Figure 9.1 is used to show the number of components needed and the proportion of the variance explained by each principal component. It is observed from this Figure that the first component suffices to explain the total variation of the original data. Therefore, the first component in this study is used as the composite index of malnutrition.

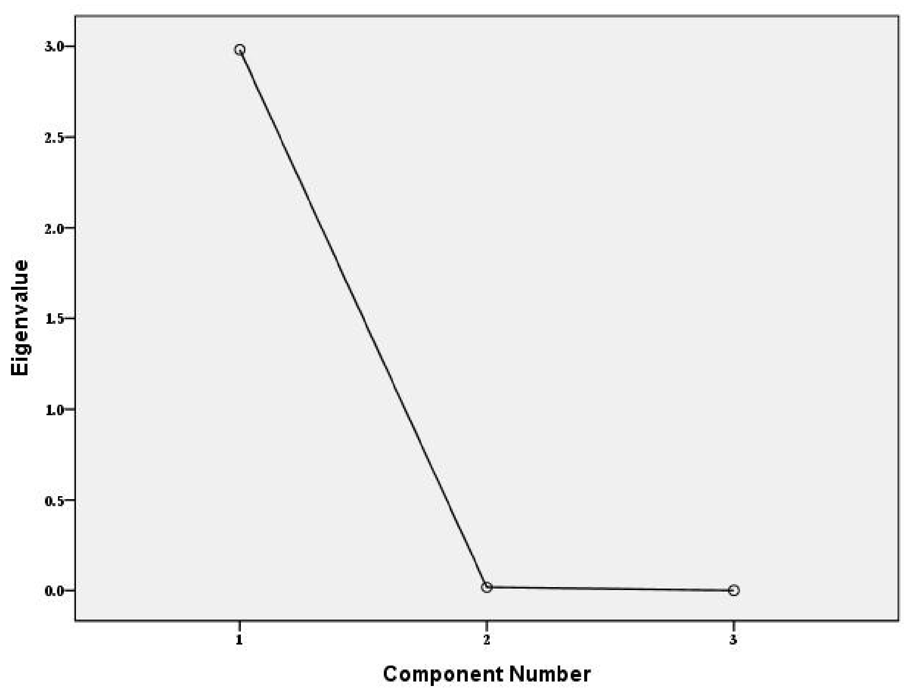
TABLE 9.1. Total variance explained

Component	Total	% Variance	Cumulative %
1	2.982	99.386	99.386
2	0.0189	.598	99.984
3	0.000	0.16	100.00

TABLE 9.2. KMO and Bartlett's test

KMO measure of sampling adequacy	.534
Bartlett's test of approximate χ^2	43454.952
df	3
Significance	.000

FIGURE 9.1. Scree plot test for composite anthropometric index



9.2. Model formulation

The theory of joint model of two binary outcomes discussed in Chapter 5 is also considered in this chapter. Let us consider the response vector of i^{th} subject to be $y_i = (y_{i1}', y_{i2}')'$, where y_{i1} is the child nutrition status (1= malnourished child and 0= nourished child) and y_{i2} is the socio-economic status of the household (1=poor household and 0=otherwise). Therefore joint multivariate binary generalized linear mixed model can be formulated as follows:

$$g_1(\mu_{i1}) = X_{i1}\beta_1 + Z_{i1}b_{i1} \tag{9.1}$$

$$g_2(\mu_{i2}) = X_{i2}\beta_2 + Z_{i2}b_{i2} \tag{9.2}$$

where β_1 and β_2 are vectors of unknown fixed effects, b_{i1} and b_{i2} are the vectors of random effects, X_{i1} , X_{i2}, Z_{i1} and Z_{i2} are the designs matrices for fixed effects and random effects respectively

$$b_i = \begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} \sim i.i.d.MVN(0, \sum) = MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sum_{11} & \sum_{12} \\ \sum_{12}' & \sum_{22} \end{bmatrix}\right), \tag{9.3}$$

where equation (9.3) is the covariance matrices of the random effects, $\sum_{11}$, $\sum_{22}$ are the variance components of malnutrition and poverty respectively. $\sum_{12} = \sum_{21}$ is the correlation component between malnutrition of children under five years and poverty of household. If $\sum_{12} = \sum_{21} = 0$, then the above model is equivalent to the separate generalized linear mixed model for two outcome variables. This means that the two outcomes are independent (Gueorguieva, 2001; Molenberghs and Verbeke, 2005).

Data analysis

We have used SAS 9.3 PROC GLIMMIX procedure to fit two binary outcomes (composite of malnutrition and asset index of household). The SAS GLIMMIX 9.3 allows to jointly model two outcomes with the same distributions or different distributions or the same link functions or different link functions. In this study same distributions are considered and the same link functions for both outcome variables. We have considered various covariance structures but Unstructured (UN) was found to be suitable to our analysis; based on the convergence criteria some of the covariance structures led to non-convergence.

9.3. Results and interpretations

The findings obtained from joint modelling of poverty and malnutrition revealed a significant positive correlation between poverty and malnutrition Table 9.6. This means that poverty and malnutrition change in same direction; when the poverty of a household increases malnutrition of children under five years in that household also increases, or in contrast, when the poverty of household reduces, in general the malnutrition also reduces.

The results are presented in Table 9.4 and Table 9.5. The findings of the study confirms the findings from chapter 5 and chapter 6. From the same Tables, it is observed that mother's level of education significantly affects the nutrition status of child as well as the socio-economic status of the household. The malnutrition of children under five years and poverty of household reduces with increasing the mother's level of education.

The age of the child significantly affects the child's nutrition status. A child aged between 12 and 23 months is 0.5689 (p-value=0.0049) times less likely to be malnourished than infant (0-11 months). The birth order of the child positively affects malnutrition. A first born child is 0.4742 (p-value $< .0001$) times less likely to be malnourished than a sixth born child or those born thereafter.

The gender of child is found to significantly affect the nutrition status of the child. A male child is 1.6242(p-value $< .0001$) times more likely to be malnourished than a female child.

Birth weight significantly affects the children's nutrition status. A child born with a higher weight is 0.3128 (p-value $< .0001$) times less likely to be malnourished than a child born with a lower weight(weight $< 2500g$).

The mother's knowledge on nutrition also significantly affects the child's nutrition status. A child born to a mother with some knowledge of nutrition is 0.6880 (p-value=0.0036) times less likely to be malnourished as compared to a child born to a mother without knowledge on nutrition.

Multiple births significantly affect the children nutrition status. A child born singleton is 0.3712 (p-value=0.0317) times less likely to be malnourished as compared to a child born second multiple or more.

The incident of anemia of the mother significantly affects the nutrition status of the children under five years . A child born to anemic mother is 1.3661 (p-value=0.0088) times more likely to be malnourished than a child born to a non-anemic mother.

The body mass index of the mother is found to significantly affect the nutrition status of the child. A child born to normal or obese mother (BMI$\geq$ 18.5) is 0.3723 (p-value $<$ .0001) times less likely to be malnourished compared to a child born to underweight mother (BMI$<$ 18.5). This result shows that there is an association between weight of the mother and nutrition status of the child.

The incident of fever is also seen to significantly affect the nutrition status of the child. A child who did not have a fever during the two weeks before the survey is 0.6623 (p-value=0.0043) times less likely to be malnourished than a child who was reported to have had a fever in the two weeks prior to the survey.

The age of household head is found to positively affect the malnutrition of children under five years.

The place significantly affects the poverty of household. A urban household is 0.7718 (p-value $<$ .0001) times less likely to be poor than a rural household.

The province significantly affects the household socio-economic status. A household from Western, Northern and Eastern provinces is 15.7053, 7.8853, and 3.5715 respectively poorer (p-value$<$ .0001) as compared to Kigali city.

Gueorguieva (2001) proposed an approach for validating the correlation between two outcomes. However, asset index (known as wealth index) in the case of malnutrition was considered as predictor variable in previous studies (Habyarimana et al., 2014, 2015d,e), but the results from all the models fitted show that poverty of household and malnutrition of children under five years are positively correlated. In addition it was observed that reducing the poverty of household also reduces the malnutrition of children under five years (Habyarimana et al., 2015e).

Table 9.3. Type 3 tests of fixed effects

Effect	Num. Df	Den. Df	F value	$Pr > F$
Child's age in moths	4	4656	2.82	0.0237
Birth order	6	4656	3.37	0.0026
Mother's education level	4	4656	9.27	$< .0001$
Gender of child	2	4656	8.78	0.0002
Birth weights	2	4656	11.57	$< .0001$
Province	8	4656	13.42	$< .0001$
Knowledge on nutrition	2	4656	5.51	0.0041
Multiple birth	4	4656	1.94	0.1005
Incident of Anemia	2	4656	5.14	0.0059
Place of residence	2	4656	12.82	$< .0001$
Body Mass Index	2	4656	10.88	$< .0001$
Incidence of fever	2	4656	4.41	0.0122
Source of drinking water	6	4656	32.99	$< .0001$
Toilet facilities	6	4656	3.73	0.001
Age of household head	2	4656	5.16	0.0058

Table 9.4. Parameter estimates for a joint model of malnutrition and poverty

	Malnutrition				Poverty			
Indicator	Estimate	Std.Error	P-Value	OR	Estimate	Std.Error	P-VALUE	OR
Intercept	2.901	1.420	0.043	17.779	-0.564	0.200	0.0049	0.5689
Child age in months								
0-11 months	reference							
12-23 months	-0.564	0.200	0.0049	0.5689	0.344	0.215	0.1097	1.4106
24^{+} months	-0.167	0.129	0.1928	0.8462	0.002	0.148	0.9818	1.0020
Birth order								
$6\&more$	reference							
4-5	-0.108	0.160	0.5007	0.8976	-0.085	0.199	0.6702	0.9185
2-3	-0.213	0.171	0.2122	0.8081	-0.188	0.204	0.3588	0.8286
1	-0.746	0.187	$< .0001$	0.4743	-0.139	0.212	0.5105	0.8702
Mother's education								
Secondary & more &reference								
Primary	1.678	0.399	$< .0001$	5.3548	1.479	0.357	$< .0001$	4.3885
No education	1.788	0.413	$< .0001$	5.9775	1.612	0.383	$< .0001$	5.0128
Gender of the child								
Female	reference							
Male	0.485	0.116	$< .0001$	1.6242	-0.029	0.128	0.8219	0.9714
birth weights								
$< 2500g$	reference							
$\geq 2500g$	-1.162	0.255	$< .0001$	4.0552	0.516	0.348	0.1379	3.1613
Province								
Kigali	reference							
South	0.036	0.313	0.9073	1.0366	-0.264	0.667	0.6929	0.7680
West	0.060	0.170	0.7198	1.0618	2.754	0.301	$< .0001$	15.784
North	-0.225	0.180	0.2132	0.7985	2.065	0.300	$< .0001$	7.8853
Eastern	-0.343	0.209	0.0997	0.7096	1.273	0.328	0.0001	3.5715
Knowledge on nutrition								
No	reference							
Yes	-0.374	0.129	0.0036	0.6880	-0.246	0.148	0.0969	0.7819
Multiple birth								
2^{nd}&more	reference							
First multiple	-0.350	0.609	0.5657	0.7047	-0.655	0.839	0.4355	0.5194
Singleton	-0.991	0.461	0.0317	0.3712	-0.815	0.691	0.2386	0.4426

Table 9.5. Continuation of parameter estimates for a joint model of malnutrition and poverty

Incident of anemia								
No anemic	reference							
Anemic	0.312	0.119	0.0088	0.312	0.260	0.138	0.0596	1.2969
Place of residence								
Rural	reference							
Urban	-0.259	0.278	0.352	0.7718	-2.530	0.503	< .0001	0.0796
Age of household head	0.016	0.005	0.0027	1.0161	0.007	0.006	0.2271	1.0070
Body mass index								
$BMI < 18.5$	reference							
$BMI \geq 18.5$	-0.988	0.221	< .0001	0.3723	0.374	0.303	0.217	1.4535
Incident of the fever								
Had fever last two weeks	reference							
No fever	-0.412	0.144	0.0043	0.6623	0.132	0.171	0.4401	1.1411
Source of drinking water								
Others/yard	reference							
Piped into dwelling/yard	-0.319	0.419	0.446	0.7276	0.531	0.624	0.3951	1.7006
Public tap	1.531	0.807	0.058	4.6228	-5.282	2.682	0.049	0.0051
Protected spring/well	0.865	0.584	0.1384	2.3750	-0.051	0.85	0.9518	0.9503
Toilet facilities								
Other toilets	reference							
Latrine	-0.318	0.419	0.4482	0.7276	0.59	0.620	0.3416	1.8040
Ventilated	1.541	0.807	0.0561	4.6692	0.-5.268	0.2.612	0.0438	0.0051
Flushed	0.870	0.584	0.1364	2.3869	0.093	0.849	0.9131	1.0975

Table 9.6. Covariance parameter estimates

Covariance parameter	Estimate	SE.	P-value
Var(Malnutrition)	0.223	0.124	0.0362
Var(Poverty)	2.403	0.395	< .0001
Correlation between malnutrition and poverty of household	0.417	0.191	0.0293

9.4. Summary

In this chapter we have created a composite index from height-for-age, weight-for-age and weight-for-height. This index is good when one is interested to assess the factors associated to malnutrition and to identify the correlation between malnutrition of children under five years and the poverty of the households.

We have used multivariate joint model to assess the possible correlation between the asset index and malnutrition as well as to assess the factors associated to malnutrition of children under five years. The findings of this study revealed a positive correlation between malnutrition of children under five years and poverty of household. This means that malnutrition and poverty increase or decrease in the same direction. This suggests that any policy change made to poverty also affects malnutrition. The findings of this chapter confirmed other findings obtained in the previous chapters. The factors associated to malnutrition are child's age, birth order of the child, mother's education level, gender of the child, birth weight of the child, mother's knowledge on nutrition, incident of anemia, body mass index of the mother, incident of fever, multiple birth, age of household head. It was also found that mother's education level affects poverty of household as well as malnutrition in the same direction.

CHAPTER 10

Conclusion and discussion

The measurements of poverty of household and malnutrition of children under five years are commonly measured based on income of household in developed countries whereas in developing countries they are measured by expenditure or consumption. However, collecting data on income and expenditure can be time and money consuming. In addition, in low-income countries, measurement of consumption and expenditure is fraught with difficulties such as the problem of recall and reluctance to divulge information. Additionally, prices are likely to differ substantially across times and areas, necessitating complex adjustment of the expenditure figures to reflect these price differences. Therefore, the main objective of this study was to develop an alternative method for measuring poverty of household and malnutrition and thereby examine the various statistical methods which are suitable to identify the risk factors associated to the poverty of households as well as the risk factors associated to the malnutrition of children under five years. To achieve these objectives we have used principal components analysis technique to create the poverty index of each household included in the survey and thereafter based on the household ranking into five quintiles from the poorest to the richest; we classified the households into socio-economic status as poor or not. We have tested the reliability of asset index by first testing the internal coherence and then testing robustness. We fitted various statistical models to poverty data and malnutrition data. Binary logistic regression and binary survey logistic regression were first applied to the household data to identify the key determinants of poverty of households and their results were compared. The findings from the comparison of the results showed that the sampling weights and sampling stratification have significant effects on parameter estimates and standards errors. Therefore, in order to get valid statistical inference, it is better to use survey logistic instead of classical logistic regression when the data were collected under multi-stage stratified sampling design. However, survey logistic regression does not account for the variability between the villages. Therefore, generalized linear mixed model was used to include the random effects.

While generalized linear models and generalized linear mixed models estimated how the predictor variables are related to the mean value of the dependent variable, quantile regression allows for studying the impact of predictors on different desired quantiles on the asset index distribution, and thus provides a complete picture of the relationship between the asset index and predictor variables. Therefore, quantile regression method was used in order to reveal some information that may be hidden when binary logistic regression, binary survey logistic and GLMM were used. In order to relax the assumption of normality and linearity inherent in linear regression models, we have used generalized additive mixed model (semiparametric), where the categorical covariates were modeled parametrically and continuous covariates non-parametrically. GAMM can reveal some information that may be hidden when only parametric models are used. The findings from all these models revealed that, in general, the level of education of household head, gender of household head, age of household head, size of the household, place of residence and the province are the key determinants of poverty of households in Rwanda. The asset index based model has a number of advantages over the money metric (income, expenditure or consumption) based model. The asset based index avoids many measurement problems associated with the classical method based on income and expenditure such as recall bias and seasonality and this is one of its main advantages over the classical methods based on income and expenditure. This method also may be very important for countries which not only lack the requisite household survey data to design policies and evaluate program effectiveness, but which also do not have the financial or human resources to generate such information. It is also very useful when considering inequality between households. However, it also has some limitations such as the Demographic and Health Survey data set is more reflective of longer-run household wealth or living standards. Therefore, in the case of Rwanda, if the need is knowledge of the current resources available to households an asset index may not be the most appropriate measure. The asset index cannot also provide information on absolute levels of poverty within the community. We were interested to compare the results from asset index to other results from other indices used to measure poverty such as Gini index (it measures the degree of inequality in the distribution of family income in developed countries or consumption in developing countries in a country) and poverty gap index (is a measure of the intensity of poverty) but these indices require information on income or consumption expenditure which are not available in DHS data. Malnutrition is measured based on anthropometric indices known as stunting (height-for-age), wasting (weight-for-height) and underweight (weight-for-age) variable. We have categorized the nutrition status of the child as severely malnourished, moderately malnourished and nourished and we have also considered the whole distribution of the index.

However, the Demographic and Health Survey data do not provide information on household income (household economic level). To circumvent this problem we created a household asset index based on ownership of consumer items and the characteristic of dwelling. Thereafter, a proportional odds model without and with complex sampling design was fitted to the data and the results were compared. It was also found that when multistage sampling was used to collect data, in order to get valid statistical inference, it is better to use a model that accounts for complexity of sampling design.

The malnutrition of children under five years is usually measured based separately on the three anthropometric indices, namely weight-for-age (underweight), height-for-age (stunting) and weight-for-height (wasting). We have used joint multivariate generalized linear mixed model to simultaneously identify the risk factors of malnutrition and also to possibly investigate the correlation between them. This model has a number of advantages over the separate models, such as: better control of type I error rates, possible gain in efficiency in parameter estimates and the ability to answer intrinsically multivariate question. The findings of this study revealed a positive correlation between stunting and underweight, wasting and underweight. The results from joint model showed that all significant covariates for underweight were still almost the same except mother's marital status which is not significant. In order to account for spatial variability between primary sampling units, we have used extended multivariate joint model to spatial multivariate joint model. But the results from this model confirmed the results from multivariate joint model. A quantile regression model was also used in order to study the impact of predictors on different quantiles of weight-for-age distribution, and therefore to provide a complete picture of the relationship between weight-for-age and predictor variables. In contrast, ordinary least squares, generalized linear models and generalized linear mixed models estimate how predictor variables are related to the mean value of the dependent variable. In order to relax the assumption of normality and linearity inherent in linear regression models, we have used generalized additive mixed model (semiparametric), where the categorical covariates were modeled parametrically and continuous covariates nonparametrically. GAMM can reveal some information which is hidden when only parametric models are used.

This study revealed that the key determinants of malnutrition of children under five years in Rwanda are: child's age in months, gender of child, birth weight, birth order, incident of fever, mother's education level, mother's age at the birth, body mass index of the mother,incident of anemia, knowledge on nutrition by mother, province, source of drinking water, multiple birth and wealth index of the household. In addition, in multivariate spatial joint model, we produced smooth maps showing the prevalence of stunting, wasting and underweight.

Further, we used principal component analysis technique to create a composite index of malnutrition from three anthropometric indices. Thereafter, multivariate joint model was used to ascertain the relationship between poverty and malnutrition and the risk factors of malnutrition and poverty simultaneously. The findings showed a positive correlation between them and this correlation means that the poverty of household and malnutrition of children from these households go in the same direction. This index is very good when identifying the risk factors of malnutrition; however it is also limited to identify the specific type of malnutrition.

The findings of this study recommended the following:

To continue supporting the existing policy of grouped settlements where people are advised to build their houses in townships known as *Imidugudu*. Since poverty levels are different by province it is important to understand poverty from a provincial perspective.

Improving access to potable water may help to reduce wasting. Also improving sensitization to the population about nutrition may reduce stunting and underweight. Improving the education level of women may reduce stunting and underweight. Sensitizing on how to take care of children may reduce not only stunting but also underweight and wasting. The sensitizing may be in the form of education that includes workshops, pamphlets, mobile clinics disseminating appropriate information on malnutrition and visits by malnutrition experts and health workers, or alternatively it can be done through *Umuganda*(community service of each last Saturday of every month). It could be better if DHS can collect GPS data at household level instead of primary sampling unit level. One must be aware of the fact that the Demographic and Health Survey data is cross-sectional and may not be able to address causality, hence longitudinal studies which will solve the problem of causality are recommended for future research. In malnutrition case, our future study is the structured additive model that includes the semiparametric (quantile regression) and the spatial variability to identify the risk factors of malnutrition.

Bibliography

Abreu, M. N. S. (2009). Ordinal logistic regression in epidemiological studies. *Revista de Saude Publica, 43*(1), 183–194.

Abreu, M. S., A. L. Siqueira, C. S. Cardosos, and W. T. Caiaffa (2008). Ordinal logistic regression models:application in quality of life studies. *Cadernos de Saude Publica, 24*(4), 581–591.

Achia, T. N. O. and N. Khadioli (2010). A logistic model to identify key determinants of poverty using demographic and health survey data. *European Journal of Social Sciences, 13*(1), 38–45.

Agresti, A. (2002). *Categorical data analysis.* New Jersey:Wiley & Sons.

Agresti, A. (2007). *An introduction to categorical data analysis,* 2^{nd} *ed.* Gainesville:John Wiley & Sons.

Agresti, A. (2010). *Analysis of ordinal categorical data,* 2^{nd} *ed.* New Jersey: John Wiley & Sons.

Allison, P. D. (1999). *Logistic regression using SAS system: Theory and application.* SAS Institute, Cary, NC.

Ananth, C. V. and D. G. Kleinbaum (1997). Regression models for ordinal response: A review of methods and applications. *International Journal of Epidemiology, 26*(2), 1323–1333.

Anthony, B. A. (2002). Performing logistic regression on survey data with new surveylogistic procedure. *Proceeding of the twenty-seventh annual SAS user group international conference*, 258–327.

Antonio, K. and J. Beirlant (2007). Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and economics, 40*(1), 58–76.

Arnold, B. C. and D. Strauss (1991). Pseudo-likelihood estimation: Some example. *Sankhya B, 53*(1), 233–243.

Barrodale, I. and F. Robert (1974). Solution of an overdetermined system of equations in ℓ_1. *Communications of the ACM, 17*(6), 319–320.

Bassett, G. J. and R. Koenker (1982). An empirical quantile function for linear models with iid errors. *Journal of american statistical association, 77*(378), 407–415.

Bender, R. and U. Grouven (1998). Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of clinal epidemiology, 51*(10), 809–816.

Binder, D. (1983). On the variances of asymptotically normal estimators from complex surveys. *International statistical review 51*(3), 279–292.

Bofinger, E. (1975). Estimation of a density function using order statistics. *Australian journal of statistics, 17*(1), 1–7.

Booth, J. G. and J. P. Hobert (1999). Maximizing generalized linear mixed model likelihhods with an automated monte carlo em algorithm. *Journal of Royal statistical society, Series B, 61*(1), 265–285.

Booysen, F. l. R. (2002). Using demographic and health surveys to measure poverty. an application to south africa. *Journal for Studies in Economics and Econometrics 26*(3), 53–69.

Brant, R. (1990). Assessing proportionality in proportional odds model for ordinal logistic regression. *Biometrics, 46*(4), 1171–1178.

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

Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of human ressources, 33*(1), 88–126.

Canty, A. J. and A. C. Davison (1999). Resampling-based variance estimation for labour force surveys. *The Statistician, 48*(3), 379–391.

Cattel, R. B. and J. Jaspers (1967). A general plasmode for factor analytic exercises and research. *Multivariate behavioural research monographs, 67-3*, 1–212.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research, 1*(2), 245–276.

Chen, C. (2004). An adaptative algorithm for quantile regression:. *in Theory and applications of recent robust methods, Birkauser Basel'*, 39–48.

Chen, C. (2007). A finite smoothing algorithm for quantile regression. *Journal for computational and graphical and statistics, 16*(1), 136–164.

Clark, D. I. and M. R. Osborne (1986). Finite algorithm for huber's m-estimator. *SIAM journal of scientific and statistical computing., 7*(1), 72–85.

Collet, D. (2003). *Modelling binary data,* 2^{nd} *ed.* Chapman & Hall/CRC.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *J. Internat. Assoc. Math. Geol., 17*(5), 563–586.

Cressie, N. and D. M. Hawkins (1980). Robust estimation of the variogram: I. *Mathematical geology, 12*(2), 115–125.

Cressie, N. A. and N. A. Cassie (1993). *Statistics for spatial data.* New York: John Wiley & Sons.

Croux, C. and P. J. Rousseeuw (1992). Time-efficient algorithms for two highly robust estimator of scale. *Computation statistics, 1*, 411–428.

Das, S. and R. M. Rahman (2011). Application of ordinal logistic regression analysis in determining risk factors of children malnutrition in bangladesh. *Nutrition Journal, 10*(1), 124.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

Deming, W. E. (1956). On simplifications of sampling design through replication with equal probabilities and without stages. *Journal of American statistical association, 51*(273), 24–53.

Dempster, A. P., R. B. Rubin, and R. K. Tsutakawa (1981). Estimation incovariance components models. *Journal of the American Statistical Association, 76*(374), 341–353.

Diggle, P. J., P. J. Heagerty, K. Y. Liang, and S. L. Zeger (2002). *Analysis of longitudinal data,* 2^{nd} *ed.* Oxford science publications. Oxford: Clarendon press.

Dobson, A. J. (2001). *Introduction to generalized linear models,* 2^{nd} *ed.* London: Chapman &Hall/CRC Press.

Durbin, J. (1958). A note on the application of quenouille's method of bias reduction to the estimation of ratios. *Biometrika, 46*(3-4), 477–480.

Efron, B. (1979). Bootstrap methods: Another look at jackknife. *Annals of statistics, 7*(1), 1–26.

Efron, B. (1982). *The Jackknife, the bootstrap and other resampling plans.* Philadelphia: SIAM.

Efron, B. and R. J. Tibshirani (1993). *An introduction to the bootstrap (Monographs on statistics and applied probability No. 57).* Boca Raton, FL: Chapman &Hall.

Faes, C., M. Aerts, G. Molenberghs, H. Geys, G. Teuns, and L. Bijnens (2008). A high-dimensional joint model for longitudinal outcomes of different nature. *Statistical in medicine, 27*(22), 4408–4427.

Fahrmeir, L. and G. Tutz (1994). *Multivariate statistical modelling based on generalized linear models.* New York: Springer-Verlag.

Fahrmeir, L. and G. Tutz (2001). *Multivariate statistical modelling on generalized linear models,* 2^{nd} *ed.* New York: Springer series in statistics, Springer-Verlag.

Falkingham, J. and C. Namazie (2002). Measuring health and poverty: A review of the appraoches to identify the poor. *Department for international development health systems resource centre (DFID HSRC).*

Fay, R. E. (1989). Theory and application of replicate weighting for variance calculations. *Proceedings of the survey research methods section: American statistical association,*, 212–217.

Fienberg, S. E. (1980). The analysis of cross-classified categorical data. *Cambridge: MIT Press*.

Fieuws, S. and G. Verbeke (2006). Pairwise fitting of mixed models for the joint modelling of multivariate longitudinal profile. *Biometrics, 62*(2), 424–431.

Filmer, D. and L. Pritchett (1998). The effect of household wealth on education attainment. evidence from 35 countries. *Population and development review, 25*(1), 85–120.

Filmer, D. and L. Pritchett (2001). Estimating wealth effects without expenditure data,or teras: An application to educational enrollments in state of india. *Demography, 38*(1), 115–132.

Freese, J. and J. S. Long (2006). Regression models for categorical dependent variables using stata. *Stata Press*.

Gameroff, M. (2005). Using the proportional odds model for health-related outcome: Why,when and how with various sas procedure. *Proceedings of the Thirtieth Annual SAS Users Group International Conference*, 205–230.

Geary, R., R. E. (1954). The contiguity ratio and statistical mapping. *The incorporated statistician, 5*(3), 115–145.

Genton, M. G. (1998). Higly robust variogram estimation. *Mathematical geology, 30*(2), 213–221.

Genton, M. G. (2001). Robustness problems in the analysis of spatial data. in spatial statistics. *Methodological aspect and application, ed. M. Moore, Springer-Verlag, New York*, 21–38.

Geys, H., G. Molenberghs, and L. Ryan (1997). Pseudo-likelihood inference for clustered binary data. *Communication in statistics- Theory and methods, 26*(11), 2743–2767.

Green, P. J. (1987). Penalized likelihood for general semi parametric regression models. *International statistical review, 55*(3), 245–259.

Green, P. J. and B. W. Silverman (1993). *Nonparametric regression and generalized linear models.* London: CRC Press.

Greenland, S., L. S. Arminda, and T. C. Waleska (1994). Alternative models for ordinal logistic regression. *Statistics in medecine, 13*(16), 1665–1677.

Grizzle, J., F. Starmer, and G. Koch (1969). Analysis of categorical data by linear models. *Biometrics, 25*(3), 489–504.

Gueorguieva, R. (2001). A multivariate generalized linear mixed for joint modelling of clustered outcomes in the exponential family. *Statistical modelling, 1*(3), 177–193.

Gutenbrunner, C., J. Jueckova, R. Koenker, and S. Portnoy (1993). Tests of linear hypotheses based on regression rank scores. *Journal of nonparametric statistics, 2*(4), 307–331.

Gutenbrunner, C. and J. Jureckova (1992). Regression rank scores and regression quantile. *Annals of statistics, 20*(1), 305–330.

Habyarimana, F., T. Zewotir, and S. Ramroop (2014). A proportional odds model with complex sampling design to identify key determinants of malnutrition of children under five years in rwanda. *Mediterranean Journal of Social Sciences, 5*(23), 1642–1648.

Habyarimana, F., T. Zewotir, and S. Ramroop (2015a). Analysis of demographic and health survey data to measure poverty of household in rwanda. *African population studies, 29*(1), 1472–1482.

Habyarimana, F., T. Zewotir, and S. Ramroop (2015b). Determinants of poverty of household in rwanda: An application of quantile regression. *Journal of Human Ecology, 50*(1), 19–30.

Habyarimana, F., T. Zewotir, and S. Ramroop (2015c). Determinants of poverty of households: Semiparametric analysis of demographic and health survey data from rwanda. *Journal of Economics and behavioral studies, 7*(3), 47–55.

Habyarimana, F., T. Zewotir, and S. Ramroop (2015d). Key determinants of malnutrition of children under five years in rwanda: Simultaneous measurement of three anthropometric indices. *African population studies, under review.*

Habyarimana, F., T. Zewotir, and S. Ramroop (2015e). Spatial distribution of key determinants of malnutrition of children under five years in rwanda: Simultaneous measurement of three anthropometric indices. *Journal of population studies, under review.*

Hakstian, A. R., W. D. Rogers, and R. B. Cattell (1982). The behaviour of numbers factors rules with simulated data. *Multivariate behavioral research, 17*(2), 193–219.

Hall, P. and S. Sheather (1988). On the distribution of the studentized quantile. *Journal of the royal statistical society, Serie B, 50*, 381–391.

Hampel, F. R., P. J. Rousseeuw, and W. A. Stable (1986). *Robust statistics , the approach based on influence functions*. New York: John Wiley & Sons.

Handcock, M. S. and M. L. Stein (1993). A bayesian analysis of kriging. *Technometrics, 35*(4), 403–410.

Hansen, M. H., W. N. Hurwitz, and W. G. Mallow (1953). *Sample survey methods and theory*. New York: Willey.

Hardle, W. (1999). *Applied nonparametric regression*. New York: Cambridge University Press.

Harville, D. A. (1974). Bayesian inference for variance components using only error-contrasts. *Biometrica, 61*(2), 383–385.

Harville, D. A. (1977). Maximum likelihood for general approaches to variance component estimation and related problems. *Journal of the american statistical association, 72*(358), 320–340.

Hastie, T. J. and R. J. Tibshirani (1990). *Generalized additive models*. New York: CRC press.

He, X. and F. Hu (2002). Markov chain marginal bootstrap. *Journal of american statistical association, 97*(1), 783–795.

Heeringa, S. G., B. T. West, and B. P. A. (2010). *Applied survey data analysis: Statistical in the social and behavioral sciences series*. New York: Chapman &Hall/CRC.

Hosmer, D. L., S. Lemeshow, and R. X. Sturdivant (2000). *Introduction to the logistic regression model,* 2^{nd} *ed.* New York: John Wiley &Sons.

Hotelling, H. (1933). Analysis of complex of statistical variables into principal components. *Journal of education psycholgy, 24*(6), 417–441.

Jiang, J. (2001). A non-standard χ^2-test with application to generalized linear model diagnostics. *Statistical and probability letters, 53*(1), 101–109.

Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. New York : Springer.

Jobson, J. D. (1992). *Applied multivariate data analysis*. New York: Springer-Verlag.

Johnson, R. A. and D. W. Wichern (2002). *Applied multivariate Statistical analysis, 4^{th} ed.* London: Prentice Hall.

Jolliffe, I. T. (1986). *Principal component analysis*. New York: Springer-Verlag.

Journel, A. G. (1978). *Mining geostatistics*. London:Academic Press.

Judkins, D. R. (1990). Fay's method for variance estimation. *Journal of official statistics, 6*(3), 223–239.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement, 20*(1), 141–151.

Kandala, N. B., T. P. Madungu, J. B. Emina, K. P. Nzita, and F. P. Cappuccio (2011a). Malnutriion among children under the age of five in democratic republic of congo (drc): Does geographic location matter? *BMC public health, 11*(1), 261–266.

Kandala, N. B., T. P. Madungu, J. B. Emina, K. P. Nzita, and F. P. Cappuccio (2011b). Models for discrete longitudinal data. springer series. *BMC public health, 11*(1), 261–266.

Karmakar, N. (1984). A new polynomial-time alogorithm for linear programming. *Combinatorica, 4*(4), 373–395.

Kim, J. O. and C. W. Mueller (1978). *Introduction to factor analysis: What it is and how to do it*. Beverly Hills,CA:Sage.

Kitanidis, P. K. (1985). Minimum variance unbiased quadratic estimation of covariances of regionalized variables. *Journal of international association for the* *mathematical geology, 17*(2), 195–208.

Koch, G. G., I. A. Amara, and J. M. Singer (1985). A two-stage procedure for the analysis of ordinal categorical data. *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences,ed.PK Sen. North Holland: Elservier Science*, 357–387.

Kocherginky, M., X. He, and F. Hu (2005). Practical confidence intervals for regression quantiles. *Journal of computational and graphical statistics, 14*(1), 41–55.

Koenker, R. (1994). Confidence intervals for regression quantiles. *Asymptotic statistics: Proceedings of the* 5^{th} *Prague symposium*, 349–359.

Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.

Koenker, R. and G. Basett (1978). Regression quantile. *Econometrica, 46*(1), 33–50.

Koenker, R. and G. Basett (1982a). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica, 50*(1), 43–61.

Koenker, R. and G. Basett (1982b). Test of linear hypothesis and l_1 estimation. *Econometrica, 50*(1), 1577–1584.

Koenker, R. and V. D'Orey (1993). Computing regression quantiles. *Applied statistics, 43*, 410–414.

Koenker, R. and A. Hallock (2000). *Quantile regression: An introduction*. University of Illinois, Urban-Campaign.

Koenker, R. and J. A. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association, 94*(448), 1296–1310.

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

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li (2005). *Applied linear statistical models,* 5^{th} *ed.* New York: McGraw-Hill Irwin.

Lehtone, R. and E. Pahkinen (2004). *Practical methods for design and analysis of complex surveys.* New York: John Willey & Sons.

Lewis-Beck, M. S. (1994). *Factor analysis and related techniques.* New London: Sage publicaion.

Lin, X. and D. Zhang (1999). Inference in generalized additive mixed models. *Journal of the royal statistical society, 61*(2), 381–400.

Lindsey, K. L. (1997). *Appliying generalized linear models.* New York: Springer-Verlag.

Linn, R. L. (1968). A monte carlo approach to the number of factors problem. *Psychometrika, 33*, 37–71.

Littel, R. C., G. A. Milliken, W. W. Stroup, D. R. Wolfinger, and O. Schabenberger (2006). *SAS system for mixed models, 2^{nd} ed. Cary,NC:.* SAS Intitute Inc.

Liu, Q. and D. Pierce (1994). A note on gauss-hermite quadrature. *Biometrika, 81*, 624–629.

Liu, X. (2009). Ordinal regression analysis: fitting the proportional odds model using stata, sas and spss. *Journal of Modern Applied Statistical Methods, 8*(2), 631–645.

Liu, X. and H. Koirala (2013). Fitting proportinal odds models to educational data with complex sampling designs in ordinal logistic regression. *Journal of Modern Applied Statistical Methods, 12*(1), 235–248.

Lokosang, L., S. Ramroop, and T. Zewotir (2014). Indexing household resilience to food insecurity shocks: The case of south sudan. *Agrekon, 53*(2), 137–159.

Madsen, K. and H. B. Nielsen (1993). A finite smoothing algorithm for linear l_1 estimation. *SIAM ournal of optimization, 3*(2), 223–235.

Mahalanobis, P. C. (1939). A sample survey of the acreage under jute in bengal. *Sankhya, 4*, 511–531.

Mahalanobis, P. C. (1944). On a large-scale sample survey. *Philosophical transactions of the royal society of London, B 231*, 329–451.

Mahalanobis, P. C. (1946). Recent experiments in statitistical sampling in the india statistical institute. *Journal of the royal statistical society, 109*, 325–370.

Manly, B. F. J. (2005). *Multivariate statistical methods.* London: Chapman&Hall/CRC.

Mardia, K. V. and R. J. Marshall (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrics, 71*(1), 135–146.

Matheron, G. (1962). *Traite de geostatistique appliquee,TomeI.* Memoires de Bureau de recherches geologiques et minieres, 14 ,editions Technip, Paris.

Matheron, G. (1963). Principles of geostatistics. *Economics geology, 58*(8), 1246–1266.

McCarthy, P. J. (1969). Pseudo-replication: half-samples. *Review of the international statistical institute, 37*(3), 239–264.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal statistical society, Series B*(Methodological), 109–142.

McCullagh, P. and J. Nelder (1989). *Generalized regression models.* London: Chapman and Hall.

McCullagh, P. and J. A. Nelder (1983). *Generalized linear models.* New York : Chapman and Hall.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of American statistical association, 92*(437), 162–170.

McCulloch, C. E. (2001). *Generalized,linear, and mixed models.* New York: John Wiley &Sons.

Mery, N. S. (2009). Ordinal logistic regression in epidemiological studies. *Rev Saude Publica, 43*(1), 297–318.

Miller, R. G. J. (1974). The jackknife -a review. *Biometrika, 61*(1), 1–15.

Molenberghs, G. and G. Verbeke (2005). *Models for discrete longitudinal data.* New York: Springer-Verlag.

Moran, P. A. P. (1950). Note on continuous stochastic phenomena. *Biometrica, 37*(1/2), 17–23.

Newey, W. and J. L. Powell (1987). Asymmetric least-squares estimation and testing. *Econometrica, 55*(4), 819–847.

Nguefac-Tsague, G. and N. Dapi (2011). Multidimensional nature of undernutrition: Statistical approach. *Journal of medecine and medical scienses, 2*(2), 690–695.

Nguefac-Tsague, G., A. T. N. Kien, and C. N. Fokunang (2013). Using weight-for-age for predicting wasted children in cameroun. *Pan African medical journal, 14*(1), doi:10.11604/pamj.201314.14.96.1914.

NISR, ICF, and MoH (2012). *Rwanda Demographic and Health Survey 2010.* NISR.

O'Connel, A. A. (2006). *Logistic regression models for ordinal response variables.* Thousans Oaks, CA: SAGE.

O'Connel, A. A. and X. Liu (2011). Model diagnostic for proportional and partial proportional odds models. *Journal of Modern Applied Statistical Methods, 10*(1), 139–175.

O'Connell, A. A. (2006). *Logistic regression models for ordinal response variables.* Thousand Oaks: SAGE.

Onis, M. D. (2000). Measuring nutritional status in relation to mortarity. *Bulletin of world Health Organization, 78*(10), 1271–1274.

Palmer, G. (2010). Relative poverty, absolute poverty and social exclusion. *The Poverty Site-p overty.org.uk 12.*

Parzen, E. (1979). Nonparametric statistical data modelling. *Journal of the american statistical association, 74*(365), 105–121.

Parzen, M. I., L. J. Wei, and Z. Ying (1994). A resampling method based on pivotal estimating equations. *Biometrika, 81*(2), 341–350.

Pearson, K. (1901). On lines and planes of closet fit to systems of points in space. *Phil. Mag., Series B, 2*, 559–572.

Peterson, B. and F. E. Harrel Jr (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics, 39*(2), 205–217.

Pfeiffer, D. U., T. P. Robinson, M. Stevenson, K. B. Stevens, D. J. Rogers, and A. C. A. Clements (2008). *Spatial analysis in epidemiology.* Oxford: Oxford university press.

Portnoy, S. and R. Koenker (1997). The gaussian hare and laplacian tortoise: Computation of square errors vs absolute-errors estimators. *Statistical Science, 12*(4), 279–300.

Powell, J. L. (1986). Censored regression quantile. *Journal of econometrics, 32*(1), 143–155.

Powers, D. A. and Y. Xie (2000). *Statistical models for categorical data analysis.* San Diego: Academica Press.

Quenouille, M. H. (1949). Problem in plane sampling. *Annals of mathematical statistics, 20*(3), 355–375.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika, 43*(3/4), 353–360.

Rao, C. R. (1979). Separation theorem for singular values of matrices and their application in multivariate analysis. *Journal of multivariate analysis, 9*(3), 362–377.

Rao, J. and J. Shao (1999). Modified balanced repeated relpication for complex survey data. *Biometrika, 86*, 403–415.

Rao, J. N. K. (1997). Development in sample survey theory: An appraisal. *Canadian journal of statistics, 25*(1), 1–21.

Rao, J. N. K. and J. Shao (1996). On balanced half-sample variance estimation in stratified sampling. *Journal of the american statistical association , 91*(433), 343–348.

Raudenbush, S. W., M. L. Yang, and M. Yosef (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximations. *Journal of computation and graphical statistics, 9*(1), 141–157.

Robinson, G. K. (1991). That blup is a good thing: the estimation of random effects. *Statistical Science, 6*(1), 15–32.

Rousseeuw, P. J. and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the Statistical Association, 88*(424), 1273–1283.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression.* Cambridge University Press.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of official statistics, 1*(4), 381–397.

Sahn, D. and D. Stifel (2003). Exploring alternative measures of welfare in the absence of expenditure data. *Cornell University*.

Sarndal, C. E., B. Swensson, and J. Wretman (1992). *Model assiated survey sampling.* New York: Springer series in statistics, Springer-Verlag.

SAS (2005). *The GLIMMIX procedure.*

Schaberger, O. and C. A. Gotway (2005). *Statistical methods for spatial data analysis.* New York: Chapman &Hall/CRC.

Schall, S. R. (1991). Estimation of generalized linear models with random effects. *Biometrika, 78*(4), 719–727.

Scott, S. C., M. S. Goldberg, and N. E. Mayo (1997). Statistical assessment of ordinal outcomes in comparative studies. *J. CLIN. Epidemiology, 50*(1), 45–55.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance components.* New York: John Willey & Sons.

Self, S. G. and K. Y. Liang (1987). Asymptotic proprtites of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of American Statistical Association, 82*(398), 605–610.

Shao, J. and D. Tu (1995). *Jackknife and bootstrap.* New York: Springer-Verlag.

Sheather, S. and J. Maritz (1983). An estimate of the asymptotic standard error of the sample median. *Australian journal of statistics, 25*(1), 109–122.

Siddiqui, M. (1960). Distribution of quantiles from a bivariate population. *Journal of research of national bureau of standards, 64B*, 145–150.

Silvapulle, M. J. and P. Silvapulle (1995). A score test against one-sided alternatives. *Journal of American Statistical Association, 90*(429), 342–349.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non parametric regression curve fitting. *Journal of the Royal statistical society, 47*(1), 43–44.

Skinner, C. J., D. Holt, and T. M. F. Smith (1989). *Analysis of complex survey.* New York: John Wiley & Sons.

Slootbeek, G. T. (1998). Bias correction in the balanced half-sample method if the number of sampled units in some strata is odd. *Journal of official statistics, 14*(2), 181–188.

Stevens, J. (1986). *Applied multivariate statistics for the social sciences.* London: Lawrence Erlbaum.

Stram, D. O. and J. W. Lee (1994). Variance components testing in the longitudinal mixed effects models. *Biometrics, 50*(4), 1171–1177.

Tierny, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association, 81*(393), 82–86.

Tucker, L. R., R. F. Koopman, and R. L. Linn (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika, 34*, 421–459.

Turkey, J. W. (1965). Which part of the sample contains the information. *Proceedings of the national academy of sciences, 53*(1), 127–134.

Vaida, F. and S. Blanchard (2005). Conditional akaike information for mixed effects models. *Biometrika, 92*(2), 351–370.

Verbeke, G. and G. Molenberghs (2000). *Linear mixed models for longitudinal data*. New York: Springer-Verlag.

Verbeke, G. and G. Molenberghs (2003). The use of score tests for inference of variance components. *Biometrics, 59*(2), 254–262.

Verbyla, A. P., B. R. Cullis, M. G. Kenward, and S. J. Welham (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical society: Series C (Applied Statistics) 48*(3), 269–311.

Vittingoff, E., D. V. Glidden, S. C. Shiboski, and C. E. McCulloch (2005). *Regression methods in Biostatistics: linear, logistic, survival and repeated measures models*. New York: Springer.

Vyas, S. and L. Kumaranayake (2006). Constructin socio-economic status indices: how to use principal components analysis. *Health Policy and Planning, 21*(6), 459–468.

Walter, S. J., C. M. J., and R. Lall (2001). Design and analysis of trials with life as an outcome: a practical guide. *J. Biophramaceutical Statistics, 11*(3), 155–176.

Wang, Y. (1998). Mixed effects smoothing spline analysis of variance. *Journal of royal statistical society, 60*(1), 159–174.

Whaba, G. (1985). A comparison of g c v and gml for choosing the smoothing parameter in generalized spline smoothing problem. *Annals of Statistics, 13*(4), 1378–1402.

WHO (1995). *Physical status: The use and interpretation of anthropometry*. Report of the expert committe.

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

Wolfinger, R., R. Tobias, and J. Sall (1994). Computing gaussian likelihood and their derivatives for generalized linear mixed models. *SIAM Journal of scientific and Statistical computing 15*(6), 1294–1310.

Wolfinger, R. D. and M. O'Connell (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of statistical computation and simulation, 48*(3-4), 233–243.

Wolter, K. (1985). *Introduction to variance estimation.* New York: Springer.

Wolter, K. M. (2007). *Introduction to variance estimation, 2^{nd} ed.* New York: Springer Science & Business Media.

Wood, S. (2006). *Generalized additive models: an introduction with R.* Boca Raton: CRC press.

Yu, K., Z. Lu, and J. Stander (2003). Quantile regression: Applications and current research areas. *Journal of the royal statistical society, 52*(3), 331–350.

Yung, W. and J. N. K. Rao (2000). Jackknife variance estimation under imputation for estimators using post-stratification information. *Journal of American statistical association, 95*(451), 903–915.

Zhang, D., X. Lin, J. Raz, and M. Sowers (1998). Semi-parametric stochastic mixed models for longitudinal data. *Journal of american statistical association, 93*(442), 710–719.