UHASSELT

KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences

School for Information Technology

Master of Statistics

Masterthesis

Modelling The Socio-Demographic and Spatial Determinants of Undernutrition in Zambia: A Comparison of Full Bayesian with Penalised Structured Additive Regression

Modou Lamin Sey

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

Prof. dr. Christel FAES

2017
2018

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Faculty of Sciences

School for Information Technology

Master of Statistics

Masterthesis

Modelling The Socio-Demographic and Spatial Determinants of Undernutrition in Zambia: A Comparison of Full Bayesian with Penalised Structured Additive Regression

Modou Lamin Sey

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

Prof. dr. Christel FAES

Abstract

Full Bayes Markov Chain Multi Carlo (MCMC) and Penalised Structured Additive Regression (STAR) models were compared for an undernutrition (measured as stunting) study in Zambia. Spatial correlated effects were specified as a Markov random field prior, continuous covariates were modelled using Bayesian penalised splines and diffused priors were assigned to fixed effects.

A Bayesian Structured Additive Regression Model was developed for the Zambia data. Model estimation and inference was based on both fully Bayesian MCMC and Empirical Bayes (based on mixed method methodology). In a frequentist setting, EB inference is closely related to penalized likelihood estimation. (Approximate) restricted maximum likelihood are used to estimate Variance components which correspond to inverse smoothing parameters. Both inference procedures were then compared based on the results from the Zambia study and were found to be very similar.

The results indicate spatial variations in stunting among the districts of Zambia. Continuous covariates Age and BMI have a significant effect on stunting. There is also significant difference among the factors of all categorical variables except for mother's employment status where no difference was found in stunting between children of employed and unemployed mothers.

Keywords: MCMC, STAR, Full Bayes, Empirical Bayes, spatial, continuous, categorical, model estimation, inference.

Acknowledgment

I would first like to thank my thesis advisor Prof Dr Christel Faes of the Center for Statistics at Hasselt University. In addition to providing the thesis topic, she consistently allowed this paper to be my own work, but steered me in the right direction whenever she thought I needed it. I especially appreciate her patience and understanding during my periods of illnesses.

I would also like to thank all my professors at UHasselt and Censtat staff members, for giving me all the skills and valuable set of knowledge which will definitely shape my future career. My special thanks goes to The Government of The Gambia for providing me with funding during my first 2 years of studies and also to my brother, Alagie Sey, for supporting me during the period of extension of my study program.

Finally, I must express my very profound gratitude to my parents and siblings for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of writing this thesis. Thank you.

Modou Lamin Sey

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

A linear model for a transformed mean of an outcome variable that has distribution in the exponential family is known as a Generalised Linear Model (GLM) (Agretsi, 2002). In most practical regression situations at least one of these problems is encountered: it may not be appropriate to assume that for a continuous covariate the effect on the predictor is strictly linear; to model the interaction effects, complex interactions might be required; spatially and temporally correlated observations may be encountered; and covariates may not sufficiently explain heterogeneity among individuals. Structured Additive Regression (STAR) Models remedies these problems (Fahrmeir, Kneib and Lang, 2004).

In this thesis, spatio extension of generalised additive models are proposed for cross-sectional data with spatial information for each observation and inference is studied from a Bayesian perspective. In Bayesian inference, the effects of the fixed or non-linear functions are explained as random variables or random functions. It is assumed that the effects of correlated spatial effects follow a Gaussian random field prior or are modelled by two-dimensional splines (Fahrmeir, Kneib and Lang, 2004). Bayesian semiparametric regression modelling has progressed in the past years such as the area of the usage of adaptive knot selection for developing Uni-and Bivariate smoothers (Smith and Kohn, 1997).

Full Bayes (FB) or an Empirical Bayes (EB) are used to perform inference on STAR models. For Full Bayes, Markov Chain Monte Carlo (MCMC) techniques are used to estimate unknown functions and covariate effects together with unknown variance or smoothing parameters. For Empirical Bayes, Restricted Maximum Likelihood (REML) is used to estimate variance or smoothing parameters. EB inference and Penalised Likelihood estimation are closely related from a frequentist perspective (Fahrmeir and Knorr, 2000). The EB approach in this thesis is based on GLMM representations.

Finally both techniques (FB and EB) are applied in the 1992 Demographic Health Surveys (DHS) of Zambia undernutrition study and the results analysed and compared. Undernutrition is the outcome of insufficient intake of food and includes stunted (reduced height for one's age), wasted (extremely thin for one's height) and insufficient weight for one's age. Undernutrition is a major public health concern in developing countries like Zambia as it is a vital sign of deprivation. By some estimates undernutrition plays a role in almost 50 percent of death in the developing world (UNICEF, 1998). Vital factors implicated in undernutrition include nutritional status of the parents, access to clean water, sanitation, and primary health care, income and education levels of parents (Sen, Roy, and Mondal, 2009).

Some of these factors such as BMI of the parents and age of child are likely to have a nonlinear effect on undernutrition . BMI is likely to follow an inverse U shape because there is a greater likelihood of having an undernourished child for parents with very low BMI. Paradoxically, parents with very high BMI may have undernourished children, as their obesity could be due to eating low nutritious food. Even for children born undernourished, their anthropometric status only worsens around 4-6 months when solid food are introduced and the child weaned (WHO, 1995; Stephenson, 1999). This stunting continues to worsen till around 24-36 months when it plateaus as the body has considerably reduced in size and could now be supported with fewer nutrition (WHO, 1995). Spatial variations in undernutrition has also been observed in developing countries even after controlling for key indicators (World Bank, 1995).

1.2 Data Description

The Demographic Health Survey (DHS) of Zambia conducted in 1992, is used in this study. There are 4847 cases. Undernutrition is measured by stunting or insufficient height for age, indicating chronic undernutrition. Stunting for a child i is determined using a z-score defined as

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

where *AI* refers to the height at a certain age, while *MAI* and σ correspond to the median and the standard deviation in the reference population, respectively. Figure 1 displays a screenshot of the first 10 rows of the Zambia dataset.

	stunting	mbmi	agechild	district	emp	edu	locality	sex
1893	-0.6817840	22.33	29	12	1	0	1	1
1894	-1.5876900	22.33	57	12	1	0	1	0
1895	0.3110890	18.66	16	12	1	0	1	0
1896	0.0067044	18.66	46	12	1	0	1	1
1897	1.8765000	24.00	28	12	1	1	1	0
1898	0.1009190	24.22	9	12	1	1	1	1
1899	-0.7180200	25.58	5	12	1	0	1	0
1900	0.0067044	25.58	30	12	1	0	1	0
1901	1.2677300	25.58	56	12	1	0	1	1
1902	-0.7904930	26.61	25	12	1	1	1	0

Figure 1: Screenshot of the Zambia dataset

.

Table 1: Variables in the undernutrition dataset

Variable	**Description**
stunting	standardised Z-score for stunting
mbmi	body mass index of the mother
agechild	age of the child in months
district	district where the mother lives
emp	mothers employment status with categories **working** (= 1) and **not working** (= 0)
edu	mothers educational status with categories **complete primary** (edu = 1), **complete secondary or higher** (edu = 2) and **no education** or **incomplete primary** (edu = 0)
locality	locality of the domicile with categories **urban** (= 1) and **rural** (= 0)
sex	gender of the child with categories **male** (= 1) and **female** (=0)

Table 1 above displays all the variables in the Zambia study and their descriptions.

1.3 Objectives

1. To compare the methodologies of Full Bayes with Penalised/Empirical Bayes for structured additive regression models.

2. To compare the results of both inference methods based on analysis of the Zambia undernutrition study.

3. To determine the factors responsible for stunting in Zambia.

2 Methodology

2.1 Structured Additive Regression Model

In Generalised Linear Models (GLMs), the response variable Y_i, $r = 1, 2, \cdots, n$, with independent observations $(y_1, \cdots, y_n)$ from an exponential family distribution makes up the *random component*. The *systemic component* of a GLM relates a vector $(\eta_1, \cdots, \eta_n)$ to the predictor variables via a linear model. Let ψ_{iq} denote the value of predictor $q(q = 1, 2, \cdots, r)$ for subject i and γ denotes unknown parameters. Then

$$\eta_i = \sum_q \gamma_q \psi_{iq}. \tag{1}$$

This linear combination of predictor variables (1) is called the *linear predictor* (Agretsi, 2002). A link function connects the random component to the systemic components. Let $\mu_i = E(Y_i)$. A link function g is a monotonic differentiable function, and links μ_i to η_i by $\eta_i = g(\mu_i)$. Thus, the formula that links $E(Y_i)$ to the predictor variables is defined as

$$g(\mu_i) = \sum_q \gamma_q \psi_{iq}$$

However at least one of these problems in practical regression situations is encountered: the assumption of a linear predictor may not be ideal for a continuous covariate; observations may be spatially or temporally correlated; covariates may not adequately explain heterogeneity among individuals or units. In this thesis, in addition to the presence of a spatial effect, another problem encountered is that the continuous covariate(s) may not be linear (Fahrmeir, Kneib and Lang, 2004). To remedy these problems, the strictly linear predictor (1) is replaced by the following structured additive predictor.

$$\eta_i = f_1(x_{i1}) + \cdots + f_j(x_{ij}) + f_p(x_{ip}) + \mathbf{u}_i'\gamma \tag{2}$$

where x_{ij} denotes the $j(j = 1, \cdots, p)$ covariates of different types and dimensions for subject

$i(i = 1, \cdots, n)$, the u_i and γ denote the vectors of values of the Linear covariates for subject i and unknown regression parameters for the Linear covariates respectively.

f_j comprises of smooth functions of continuous; varying coefficient models; independent and identical random intercepts and slopes; time trends and seasonal effects; spatially or temporally correlated random effects; and two dimensional surfaces. From a frequentist perspective, these functions are deterministic while in the Bayesian paradigm they are considered as random functions realizations. P-splines are used to model the smooth functions. In this approach, it is assumed that a polynomial spline of degree l approximates an unknown smooth function of covariate x. Within the domain x, a set of equally spaced knots $x_{min} = \zeta_0 < \zeta_1 < \cdots < \zeta_{m-1} < \zeta_m = x_{max}$ defines the polynomial spline (Fahrmeir, Kneib and Lang, 2004). Another way of writing such a spline is using the linear combination of $K = m + l$ B-spline basis functions B_k, i.e,

$$f(x) = \sum_{k=1}^{K} \beta_k B_k(x)$$

Here $\beta = (\beta_1 \cdots, \beta_K)'$ denotes the vector of unknown regression coefficients. The $n \times K$ design matrix $\mathbf{X}$ with elements $X[i,k] = B_k(x_i)$ and a vector of unknown parameters β_j is used to express the vector of function evaluations $\mathbf{f} = (f_1(x_{i1}) + \cdots + f_j(x_{ij}))'$ of an unknown function f_j. i.e.,

$$\mathbf{f} = \mathbf{X}\beta$$

As a result the following is obtained for model (2),

$$\eta = \mathbf{X_1}\boldsymbol{\beta_1} + \cdots + \mathbf{X_p}\boldsymbol{\beta_p} + \mathbf{U}\gamma \tag{3}$$

where $\mathbf{U}$ is the design matrix for linear effects and γ is the vector of regression coefficients for linear effects.

2.2 Prior Assumptions

A diffuse prior $p(\gamma_j) \propto$ const will be considered for the fixed effects parameters γ. Priors for the unknown functions $f_1, \cdots, f_p$ in (2) depend on the type of the covariate and on the prior beliefs about smoothness.

A general form for representing the smoothness priors for the regression coefficients β_j is given by:

$$p(\beta_j/\tau_j^2) \propto \frac{1}{(\tau_j^2)^{rank(\mathbf{K}_j)/2}} exp(-\frac{1}{2\tau_j^2}\beta_j'\mathbf{k}_j\beta_j) \tag{4}$$

where $\mathbf{K}_j$ is a penalty matrix and the prior for β_j is partially improper due to $\mathbf{K}_j$ being rank deficient in most cases.

The variance parameter τ_j^2 controls the trade off between flexibility and smoothness and is equivalent to the inverse smoothing parameter in a frequentist approach.

The following subsections describe specific priors for continuous and spatial covariates and their functions f_j.

A form of quadratic penalties $\beta'\mathbf{P}(\boldsymbol{\lambda})\beta$, with $\mathbf{P}(\boldsymbol{\lambda})$ representing a penalty matrix, is specified on the regression coefficients so that overfitting is avoided. The form of most penalty matrices is $\mathbf{P}(\boldsymbol{\lambda}) = \lambda\mathbf{K}$ with λ denoting a scalar smoothing parameter and $\mathbf{K}$ denotes a penalty matrix.

Prior assumptions about the smoothness of f and the type and dimension of x determines the basis functions $(B_1, \cdots, B_k)'$ and penalty P($\boldsymbol{\lambda}$) to use (Fahrmeir, Kneib and Lang, 2004).

2.2.1 Priors for Continuous covariates

There are several alternatives for specifying smoothness priors for continuous covariates. The focus in this thesis will be on Random walks and bayesian P-splines.

P-splines

The B-spline basis functions evaluated at the observations gives the columns of the design matrix **X**. The number of knots to use requires making some important choices i.e. either the variability of the data might not be captured when using a small number of knots or overfitting of the data occurs when a large number of knots are used. To ensure enough flexibility and smoothness of the curve, Eilers and Marx (1996) recommends using a moderately large number of equally spaced knots (between 20 and 40). As a result the following Penalized likelihood estimation is obtained:

$$P(\gamma) = \frac{1}{2}\gamma \sum_{k=r+1}^{K} (\Delta^r \beta_k)^2, \qquad r = 1,2,$$

where γ is the smoothing parameter and Δ^r is the difference operator of order k. Suppose x represents an equally spaced ordered observations of a continuous covariate

$$x^{(1)} < x^{(2)} < \cdots < x^{(K)}.$$

where $K \leq n$ denotes the number of different observed values for x in the data set. In dynamic models, a common approach is to estimate one parameter β_k for each distinct $x^{(k)}$ i.e. $f(x^{(k)}) = \beta_k$ and then use random walk priors to penalize too abrupt jumps between successive parameters. In the Bayesian approach, for the regression coefficients, first or second order random walks are used as priors. First and second order random walk models are defined by

$$\beta_k = \beta_{k-1} + u_k \qquad \text{and} \qquad \beta_k = 2\beta_{k-1} - \beta_{k-2} + u_k \tag{5}$$

and for initial values, Gaussian errors $u_k \sim N(0, \tau^2)$ and diffuse priors $p(\beta_1) \propto const$, or $p(\beta_1)$ and $p(\beta_2)) \propto const$, are used respectively. A product of conditional densities defined by (5) easily computes the joint distribution of the regression parameters β_j and can be brought into the general form (4). The penalty matrix is of the form $\mathbf{K} = \mathbf{D}'\mathbf{D}$ where $\mathbf{D}$ is a first or second order difference matrix (Lang and Brezger, 2004). The order of the spline and the penalty determines the limiting behaviour $\lambda \to \infty$. A polynomial fit of degree $r-1$ is obtained in the limit if the order of the spline is equal to or higher than the order of the penalty. For example, the penalty matrix for a random walk of first order is:

$$\mathbf{K} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}$$

τ^2, the variance parameter, controls the amount of smoothness and the smaller (larger) the variance the smoother (rougher) are the estimated functions (Besag et al., 1995).

2.2.2 Priors for spatial effects

Assume that the index $s \in 1, \cdots, S$ is suppose to represent the location or site in connected geographical regions. A common way to introduce a spatially correlated effect is to assume that neighboring sites are more alike than arbitrary sites. Thus, a set of neighbors for each site s must be defined for a valid prior definition. For geographical data one usually assumes that two sites s and $s_{\prime}$ are neighbors if they share a common boundary. The simplest and most frequently used spatial smoothness prior for the function evaluations $f(s) = B_s$ is defined as

$$\beta_s|\beta_{s'}, s \neq s', r^2 \sim N\left(\frac{1}{N_s}\sum_{s' \in \delta_s} \beta_{s'}, \frac{\tau^2}{N_s}\right)$$

where N_s is the number of adjacent sites and $s' \in \delta_s$denotes that site s' is a neighbor of site s. The prior is called a Markov random field (MRF) and is a direct generalization of a first order random walk to two-dimensions (Fahrmeir, Kneib and Lang, 2004).

Unobserved covariates cause problems of heterogeneity among clusters of observations, and this may result in false standard error estimates and biased estimates for the remaining effects. Suppose $c \in 1, \cdots, C$ is a cluster variable denoting the cluster a particular observation belongs to. Problems of heterogeneity are remedied by introducing additional Gaussian i.i.d. effects $f(c) = \beta_c$ with

$$\beta_c \sim N(0, \tau^2), \qquad x = 1, \cdots, C. \tag{6}$$

A $n \times C$ 0/1 incidence matrix represents the design matrix $\mathbf{X}$ and the the identity matrix $(K = i)$ is the penalty matrix. For more complex modelling of spatial effects, the prior in (6) may be used. For example, in certain instances it is recommended to split up a spatial effect f_{spat} into a spatially correlated (smooth) part f_{str} and a spatially uncorrelated (unsmooth) part f_{unstr}, i.e., $f_{spat} = f_{str} + f_{unstr}$. Since surrogates of many unobserved influential factors defines a spatial effect, a rational for the split up is that some of them may obey a strong spatial structure whilst others may be present only locally (Fahrmeir, Kneib and Lang, 2004). Hence, the two kinds of influential factors are differentiated by estimating a structured and an unstructured component. As a result, by observing which of the two effects is larger it is possible to assess the degree of spatial dependency in the data. If the structured exceeds the unstructured effects, the spatial dependency is larger and vice versa (Besag York and Mollie, 1991). Markov random field priors or two dimensional surface smoothers are assumed as priors for the smooth spatial part and prior (6) is assumed for the uncorrelated part.

2.3 Mixed Model representation

This section shows how STAR models can be represented by generalized linear mixed models (GLMM) after appropriate reparameterization. Utilizing the structured additive predictor (3) provides the solution for the simultaneous estimation of the functions f_j, $j = 1, \cdots, p$, and the variance (or inverse smoothing) parameters τ_j^2 in an EB approach discussed in Section 1.6. the general model formulation again is crucial to rewrite the model as a GLMM.

Assume that the dimension $K_j \times 1$ and the corresponding penalty matrix $\mathbf{K}_j$ of the j-th coefficient vector has rank k_j. By decomposing the vectors of regression coefficients β_j, $j = 1, \cdots, p$ into an *unpenalized* and a *penalized* part, the decomposition is given by:

$$\beta_j = \mathbf{X}_j^{unp}\beta_j^{unp} + \mathbf{X}_j^{pen}\beta_j^{unp}, \tag{7}$$

where a basis of the nullspace of $\mathbf{K}_j$ is contained in the columns of the $K_j \times (K_j k_j)$ matrix $\mathbf{X}_j^{unp}$ and the orthogonal deviation from this nullspace is contained in $\mathbf{X}_j^{pen}$. the unpenalised part in j is separated from the penalised part by decomposition (7).

The following is obtained from the decomposition (7)

$$\frac{1}{\tau_j^2}\beta_j'\mathbf{K}_j\beta_j = \frac{1}{\tau_j^2}(\beta_j^{pen})'\beta_j^{pen}$$

and the following is obtained from the general prior (4):

$$p(\beta_{jm}^{unp}) \propto const, \qquad m = 1, \cdots, K_j - k_j$$

and

$$(\beta_j^{pen}) \sim N(0, \tau_j^2\mathbf{I}). \tag{8}$$

By defining the matrices $\tilde{\mathbf{U}}_j = \mathbf{X}\mathbf{X_j}^{unp}$ and $\tilde{\mathbf{X}}_j = \mathbf{X}\mathbf{X_j}^{pen}$, the predictor (3) can be rewritten as:

$$\eta = \sum_{j=1}^{p}\mathbf{X}_j\beta_j + \mathbf{U}\gamma$$

$$= \sum_{j=1}^{p}\tilde{\mathbf{U}}_j\beta_j^{unp} + \mathbf{U}\gamma$$

$$= \tilde{\mathbf{U}}\beta^{unp} + \tilde{\mathbf{X}}\beta^{pen}.$$

The matrices $\tilde{\mathbf{X}}_j$ and the vectors β_j^{pen} comprises the design matrix $\tilde{\mathbf{X}}$ and the vector β^{pen} respectively. i.e. $\tilde{\mathbf{X}} = \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \cdots, \tilde{\mathbf{X}}_p$ and $\beta^{pen} = ((\beta_1^{pen})', \cdots, (\beta_p^{pen})')'$. Similarly, $\tilde{\mathbf{U}} = (\tilde{\mathbf{U}}_1\tilde{\mathbf{U}}_2\cdots\tilde{\mathbf{U}}_p\mathbf{U})$ and $\beta^{unp} = ((\beta_1^{unp})', \cdots, (\beta_p^{unp})', \gamma')'$.

Finally, a GLMM with fixed effects β^{unp} and random effects $\beta^{pen} \sim N(0, \Lambda)$ where $\Lambda = diag(\tau_1^2, \cdots, \tau_1^2, \cdots, \tau_p^2, \cdots$

is obtained (Fahrmeir, Kneib and Lang, 2004).

2.4 Full Bayesian Inference based on MCMC techniques

In full Bayesian inference, τ_j^2, the unknown variance parameters, are also considered as random variables consequently requiring suitable hyperprior assumptions. Highly dispersed and proper inverse Gamma priors $p(\tau_j^2) \sim IG(a_j, b_j)$ are assigned to the variances resulting probability density function:

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

Even though the priors for the β_j are partial improper, using proper priors for τ_j^2 (with $a_j > 0$ and $a_j > 0$) ensures propriety of the joint posterior. A common choice for the hyperparameters (also the default in *BayesX*) is $a_j = b_j = 0.001$.

The posterior of the model provides the basis for Bayesian inference and is defined as

$$p(\beta_1, \cdots, \beta_p, \tau_1^2, \cdots, \tau_p^2, \gamma | y) \propto L(y, \beta_1, \cdots, \beta_2, \gamma) \prod_{j=1}^{p} (p(\beta_j | \tau_j^2) p(\tau_j^2)),$$

where L(.) denotes the likelihood which is a product of the individual likelihood contributions assuming conditional independence.

The posterior distribution is numerically intractable in many practical situations (especially for structured additive regression models). Markov Chain Monte Carlo (MCMC) simulations method remedies this problem by drawing random samples from the posterior. Characteristics of the posterior such as posterior means, standard deviations or quantiles can be estimated by their empirical analogues. By using the posterior as stationary distribution, MCMC devices a way to construct a Markov chain. As a result, a sample of dependent random numbers are produced by converging the iterations of the transition kernel of this Markov chain to the posterior. Since some time is needed for the algorithms to converge the first part of the sample (the burn-in phase) is usually discarded. Furthermore, to minimize autocorrelations, some thinning is usually applied (Fahrmeir et al., 2004).

Sampling scheme based on Gaussian responses

For that the distribution of the response variable is gaussian i.e $y_i|\eta_i$, $\sigma^2 \sim N(\eta_i, \frac{\sigma^2}{c_i})$, $i = 1, \cdots, n$ or $\mathbf{y}|\eta$, $\sigma^2 \sim N(\eta, \sigma^2\mathbf{C}^{-1})$ where $\mathbf{C} = diag(c_1, \cdots, c_n)$ is a known weight matrix. A Gibbs sampler can be employed in this situation, since full conditionals for fixed effects as well as nonlinear functions f_j are multivariate Gaussian i.e. the mean for the full conditional $\gamma|.$ for fixed effects with diffuse priors is Gaussian and is defined as

$$E(\gamma|.) = (\mathbf{U}'\mathbf{C}\mathbf{U})^{-1}\mathbf{U}'\mathbf{C}(\mathbf{y} - \tilde{\eta}) \tag{9}$$

and covariance matrix

$$Cov(\gamma|.) = \sigma^2(\mathbf{U}'\mathbf{C}\mathbf{U})^{-1}) \tag{10}$$

where $\mathbf{U}$ denotes fixed effects designed matrix and $\tilde{\eta} = \eta - \mathbf{U}\gamma$ denotes the part of the additive predictor associated with the remaining effects in the model. Similarly, the mean for the full conditional for the regression coefficients β_j of a function f_j is Gaussian and is given by

$$\mathbf{m}_j = E(\beta_j|.) = (\frac{1}{\sigma^2}\mathbf{X}_j'\mathbf{C}\mathbf{X}_j + \frac{1}{\tau_j^2}\mathbf{K}_j)^{-1}\frac{1}{\sigma^2}\mathbf{X}_j'\mathbf{C}(\mathbf{y} - \eta_{-j}) \tag{11}$$

where $\eta_j = \eta - \mathbf{X_j}\beta_j$, and covariance matrix

$$Cov(\beta_j|.) = \mathbf{P}_j^{-1} = (\frac{1}{\sigma^2}\mathbf{X}_j'\mathbf{C}\mathbf{X}_j + \frac{1}{\tau_j^2}\mathbf{K}_j)^{-1} \tag{12}$$

Linear equation systems with a high dimensional precision matrix $\mathbf{P_j}$ must be solved in every iteration of the MCMC scheme, consequently drawing random samples in an efficient way is not simple. From Rue (2001), obtaining $(\beta_j|.)$ is used to obtain random numbers as follows: the Cholesky decomposition is computed $\mathbf{P}_j = \mathbf{L}\mathbf{L}'$ and $\mathbf{L}'\beta_j = \mathbf{z}$ solved, where $\mathbf{z}$ is a vector of independent standard Gaussians and $\beta_j \sim N(\mathbf{0}, \mathbf{P}_j^{-1})$. Then the mean $\mathbf{m}_j$ is computed by solving $\mathbf{P}_j\mathbf{m}_j = \frac{1}{\sigma^2}\mathbf{X}_j'\mathbf{C}(\mathbf{y} - \eta_{-j})$. Finally, $\beta_j \sim N(\mathbf{m}_j, \mathbf{P}_j^{-1})$ is produced by adding $\mathbf{m}_j$ to the previously simulated β_j (Lang and Brezger, 2004).

The full conditionals for the variance parameters τ_j^2, $j = 1, \cdots, p$, and σ^2 are all inverse Gamma distributions with parameters

$$a'_j = a_j + \frac{rank(\mathbf{k}_j)}{2} \quad \text{and} \quad b'_j = b_j + \frac{1}{2}\beta'_j \mathbf{K}_j \beta_j \tag{13}$$

for τ_j^2. For σ^2 we obtain

$$a'_\sigma = a_\sigma + \frac{n}{2} and \quad \text{and} \quad b'_\sigma = b_\sigma + \frac{1}{2}\varepsilon'\varepsilon \tag{14}$$

where ε is the usual vector of residuals.

Summary sampling scheme

1. *Initialization*

   Given fixed smoothing parameters $\lambda_j = \frac{\sigma^2}{\tau_j^2}$, *BayesX* uses $\lambda_j = 0.1$ as default, compute the posterior mode for $\beta_1, \cdots, \beta_p$ and γ . backfitting algortihm is used to compute the mode. These posterior mode estimates are used as the intial state β_j^2, $(\tau_j^2)^c$,γ^c of the chain.

2. *Update regression parameters γ*

   Draw from the Gaussian full conditional with mean and covariance matrix specified in (9) and (10) to update regression parameters γ .

3. *Update regression parameters β_j*

   Draw from the Gaussian full conditional with mean and covariance matrix specified in (11) and (12) to update β_j for $j = 1, \cdots, p$

4. *Update variance parameters τ_j^2 and σ^2*

   Draw from the inverse gamma full conditional with parameters specified in (13) and (14) to update variance parameters τ_j^2 and σ^2.

2.5 Penalised Likelihood/Empirical Bayes inference based on mixed model methodology

For the EB inference, where variances τ_j^2 are considered as constants, the variances τ_j^2 and the priors $p(\tau_j)^2$ have to be deleted. The posterior in terms of the GLMM representation of the model in section (2.4) is given by

$$p(\beta^{unp}, \beta^{pen}|\mathbf{y}) \propto L(\mathbf{y}, \beta^{unp}, \beta^{pen}) \prod_{j=1}^{p} \left(p(\beta_{\mathbf{j}}^{pen}|\tau_{\mathbf{j}}^{\mathbf{2}}) \right),$$

where $p(\beta^{pen}|\tau_j^2) \propto N(0, \tau_j^2 \mathbf{I})$

Regression and variance parameters can be estimated using iteratively wighted least squares (IWLS) and approximate marginal or restricted maximum likelihood (REML) developed for GLMMs. The following 2 iterative steps describe the Estimation process:

1. Given the current variance parameters as the solutions of the system of equations obtain updated estimates $\hat{\beta}^{unp}$ and $\hat{\beta}^{pen}$

$$\begin{pmatrix} \tilde{\mathbf{U}}'\mathbf{W}\tilde{\mathbf{U}} & \tilde{\mathbf{U}}'\mathbf{W}\tilde{\mathbf{X}} \\ \tilde{\mathbf{X}}'\mathbf{W}\tilde{\mathbf{U}} & \tilde{\mathbf{X}}'\mathbf{W}\tilde{\mathbf{X}} + \tilde{\Lambda}^{-1} \end{pmatrix} \begin{pmatrix} \beta^{unp} \\ \beta^{pen} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{U}}'\mathbf{W}\tilde{\mathbf{y}} \\ \tilde{\mathbf{U}}'\mathbf{W}\tilde{\mathbf{y}} \end{pmatrix}.$$

The usual working observations and weights in generalized linear models are defined by the $(n \times 1)$ vector $\tilde{\mathbf{y}}$ and the $n \times n$ diagonal matrix $\mathbf{W} = \mathrm{diag}(w_1, \cdots, w_n)$.

2. maximise the approximate marginal / restricted log likelihood to obtain updated estimates for the variance parameters $\hat{\tau}_j^2$

$$l^*(\tau_1^2 \cdots, \tau_p^2) = -\frac{1}{2} log(|\, \Sigma \,|) - \frac{1}{2} log(|\, \tilde{\mathbf{U}} \Sigma^{-1} \tilde{\mathbf{U}} \,|) - \frac{1}{2} (\tilde{\mathbf{y}} - \tilde{\mathbf{U}}\hat{\beta}^{unp})' \Sigma^{-1} (\tilde{\mathbf{y}} - \tilde{\mathbf{U}}\hat{\beta}^{unp}) \tag{15}$$

with respect to the variance parameters $\tau_1^2, \cdots, \tau_p^2$. An approximation to the marginal covariance matrix of $\tilde{\mathbf{y}} \mid \hat{\beta}^{pen}$ is given by $\Sigma = \mathbf{W}^{-1} + \tilde{\mathbf{X}}\Lambda\tilde{\mathbf{X}}'$. Iteration of the 2 steps are done until convergence is achieved. In BayesX, instead of the usual Fisher scoring iterations, the marginal likelihood (15) is maximized by a computationally efficient alternative (Fahrmeir et al., 2004).

For the above algorithm, convergence problems may occur if one of the parameters τ_j^2 is small. In this situation, Fisher scoring fails in finding the marginal likelihood estimates $\hat{\tau}^2$ because the maximum of the marginal likelihood may be on the boundary of the parameter space. If criterion (16) is smaller than the user specified value **lowerlim**, the estimation of small variances τ_j^2 is stopped.

$$c(\tau_j^2) = \frac{\| \tilde{\mathbf{X}}_j \hat{\beta}_j^{pen}}{\| \hat{\mathbf{eta}}} \| \tag{16}$$

2.6 Full Bayesian MCMC and Empirical Bayes Models for The Zambia dataset

In the Zambia dataset, the response is stunting measured as a Z-score. Usually, the effect of the covariates on the response is modelled by a linear predictor. In this thesis, much emphasis is placed on the effects of the two continuous covariates *age of the child agc* and the *mother's body mass index bmi*, which could possibly be nonlinear, and on regional effects of the district where the mother lives. A second model will also be run where the two continuous covariates *agc* and *bmi* are treated as linear covariates and the two estimates compared.

The first model is given by

$$\begin{aligned}\eta &= \gamma_0 + \gamma_1 locality + \gamma_2 edu1 + \gamma_3 edu2 + \gamma_4 emp + \gamma_5 sex + f_1(bmi) + \\ &\quad f_2(agc) + f_{str}(district) + f_{unstr}(district)\end{aligned}$$

The second model is given by

$$\begin{aligned}\eta &= \gamma_0 + \gamma_1 locality + \gamma_2 edu1 + \gamma_3 edu2 + \gamma_4 emp + \gamma_5 sex + \gamma_6(bmi) + \\ &\quad \gamma_7(agc) + f_{str}(district) + f_{unstr}(district)\end{aligned}$$

The categorical covariates are modelled by independent diffuse priors i.e. $p(\gamma_j) \propto$ const. Since the two continuous covariates *agc* and *bmi* are assumed to have a possibly nonlinear effect on the Z-score, they are therefore modelled nonparametrically by cubic P-splines with 20 equidistant knots and a second order random walk prior. For their variance components, Inverse gamma prior with hyperparameters a=0.001 and b=0.001 are used.

For the spatially correlated effect $f_{str}(district)$, Markov random field priors are chosen, and for the variance component, Inverse gamma prior with hyperparameters a=0.001 and b=0.001 are used. While for the spatially uncorrelated effect $f_{unstr}(district)$ i.i.d. Gaussian random effects are chosen and for the variance component, Inverse gamma prior with hyperparameters a=0.001 and b=0.001 are used.

2.6.1 Full Bayesian MCMC Model

The function *bayesx* from the *R* software package "R2BayesX" is used to estimate the model:

zm = bayesx(stunting emp.f + edu.f + locality.f + sex.f + sx(mbmi, bs = "psplinerw2", knots = 20, degree = 3) + sx(agechild, bs = "psplinerw2", knots = 20, degree = 3) + sx(district, bs = "mrf", map = ZambiaBnd)+ sx(district, bs = "re"), family="gaussian", method="MCMC", predict = TRUE, hyp.prior = c(0.001, 0.001), iter = 32000, burnin = 2000, step = 15, data = zam)

Properties that define the MCMC algorithm such as iterations, burnin and step are utilised in the above model. Iterations=32000 defines the total number of MCMC iterations used in the model while burnin=2000 defines the number of burn in iterations used. From the above specifications, a sample of 30000 random numbers is obtained. Because these random numbers are correlated, the thinning parameter step is used to thin out the Markov chain. By specifying step=15 in the model, *BayesX* will only store every 15th sampled parameter resulting in a random sample of length 2000 for every parameter.

Convergence of the model will be assessed via autocorrelation plots and sampling paths.

2.6.2 Penalised likelihood/Empirical Bayes Model

The GLMM methodology used for EB inference in STAR models is enabled by the REML algorithms. The function *bayesx* from the *R* software package ”R2BayesX” is also used to estimate the EB model:

zm1 = bayesx(stunting emp.f + edu.f + locality.f + sex.f + sx(mbmi, bs = ”psplinerw2”, knots = 20, degree = 3) + sx(agechild, bs = ”psplinerw2”, knots = 20, degree = 3) + sx(district, bs = ”mrf”, map = ZambiaBnd) + sx(district, bs = ”re”), family=”gaussian”, method = ”REML”, lowerlim=0.001, eps=0.00001, data = zam)

The estimation process is controlled by options **lowerlim** and **eps**. The usual Fisher-scoring algorithm for determining small variances have to be modified because they are close to the boundary of their parameter space. The estimation of the variance of an effect is stopped if the fraction of the penalized part of an effect relative to the total effect is less than **lowerlim** and the estimator is defined to be the current value of the variance. The criteria for the termination of the estimation process is defined by **eps**. The estimation process is assumed to have converged if both the relative changes in the regression coefficients and the variance parameters are less than **eps**.

2.7 Software and Tools

All the statistical analysis are carried out using R version 3.3.1 and BayesX version 3.0.2 statistical packages.

3 Results

3.1 Exploratory Data Analysis

Table 2: Frequency table for variable sex

Sex	Observations	Freq	Cum
Female	2451	0.5057	0.5057
Male	2396	0.4943	1

Table 2 above displays the frequencey table for the sex variable, there were 2396 Male children and 2451 Female children. The number of employed mothers were 2657 while 2190 were unemployed (Table 7, Appendix I). For the residence of the children, 2102 live in urban areas whilst 2745 live in rural areas (Table 8, Appendix I). Table 9 in Appendix I, displays the different education levels of the mothers (2302 had no education, 2355 had primary education and 190 had secondary education).

Table 3: Discriptive statistics of BMI and Age of child

Variable	Mean	Median	Std	Min	Max
BMI	21.94	21.4	3.29	12.8	39.29
Age	26.67	25	17.11	0	59

Table 3 displays the descriptive statistics for the mother's BMI and the child's age. For the BMIs of the mothers, the minimum recorderd was 12.8, mean was 21.94 and maximum was 39.29. For the children age, the minimum recorderd was 0 months, median was 25 months and maximum was 60 months.

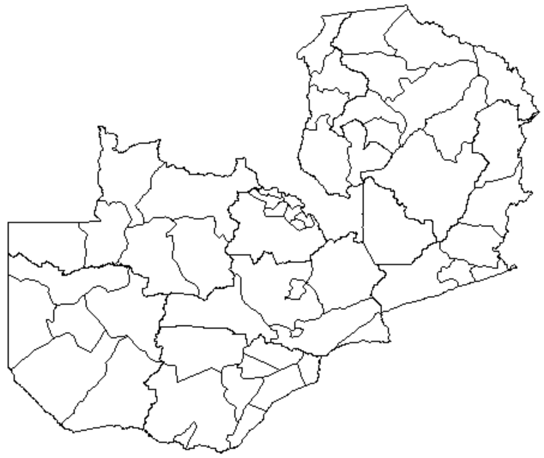

Figure 2: Map displaying the districts of Zambia

.

Figure 2 displays the map of Zambia and the 54 districts within it. The minimum number of neighbours is 1 and the maximum 9.

3.2 Full Bayesian MCMC

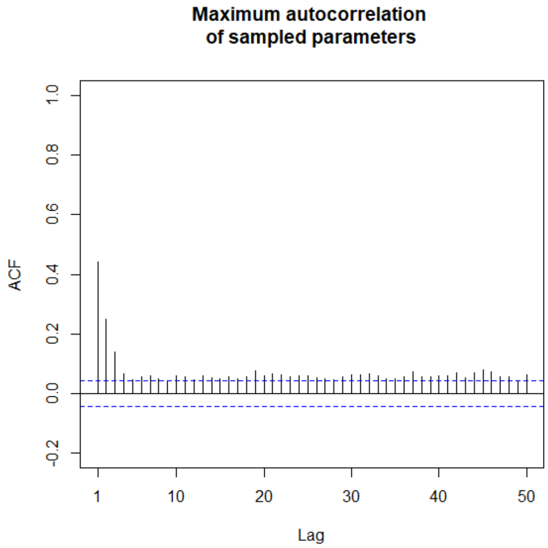

Figure 3: Maximum autocorrelation of all parameters of the model

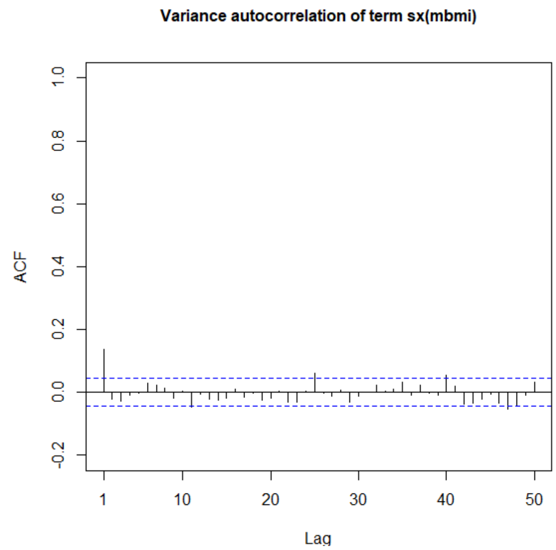

Figure 4: Autocorrelation function of the samples of the BMI variance parameter

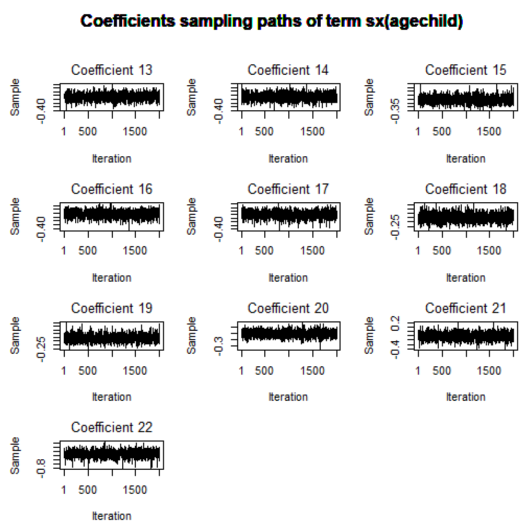

Figure 5: sampling paths of the last 10 coeffs of age

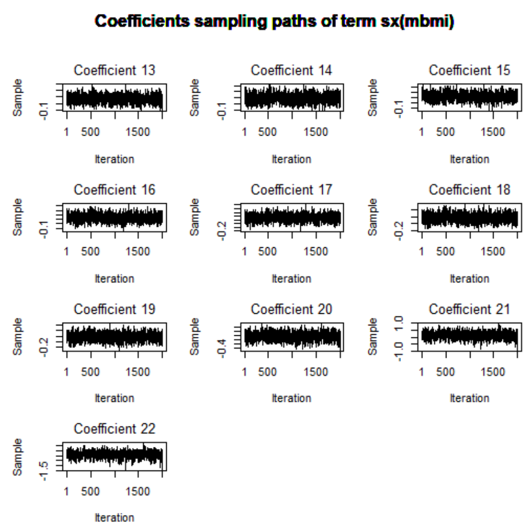

Figure 6: sampling paths of the last 10 coeffs of BMI

Convergence for the MCMC model is assessed via autocorrelation plots and sampling paths. Figures 3 and 4 display the autocorrelation function for all the parameters and the BMI variance parameter respectively. As can be observed from both plots, the autocorrelations for all the lags are very close to zero indicating convergence of our model. Autocorrelation for the lags of the other parameters were all very close to zero as displayed in Appendix II (Figures 14-17). Figures 5 and 6 of the sampled parameters show uncorrelated samples with no particular pattern further proving the convergence of our model.

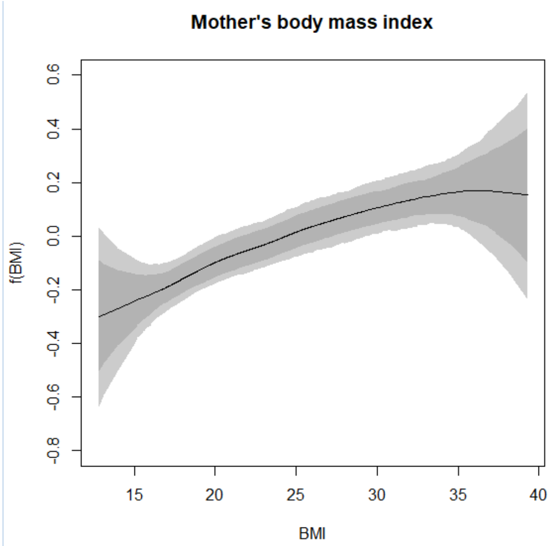

Figure 7: Non-linear Effect of 'bmi'

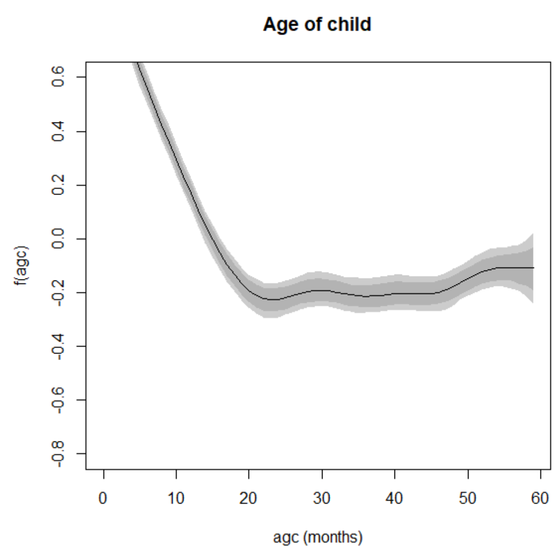

Figure 8: Non-linear Effect of 'agc'

Figures 7 and 8 show the nonlinear effects of childs age and the mothers BMI. From Figure 7 it can be observed that as BMI of the mother increases, the child's Z-score also increases meaning that children of mother's with higher BMI had less stunting compared to those with lower BMI. Figure 8 shows that as the child's age increases the Z-score reduces meaning that with increased age stunting also increases until at around 20 months when the z-score stabilises. It can also be seen from both figures that there is clear non-linearity of the curves (especially for the age variable) justifying the inclusion of both the age and BMI variables as non-linear covariates. This is also confirmed from the values of AIC=3821.88 and BIC=4108.43 of model I (treated both covariates as non-linear) versus values of AIC=4160.87 and BIC=4409.04 of model II (treated both as linear covariates). As a result, model I is preferred for the subsequent analysis.

Tables 4 and 5 show the posterior mean estimates of model I and II respectively. There is not much difference in the models in either the parameter estimates or their significance. From Table 4, the variances of the covariate BMI (0.0018) is less than that of age (0.0067) meaning that the curve of the former is smoother than that of the latter. The smoothing parameter is determined from the formula ($smoothingparameter = \frac{scale}{variance}$), hence the smaller the variance the smoother the curve.

Table 4: Parameters and Variances estimates of model I

Parameter	Mean	Sd	2.5%-Quant.	Median	97.5%-Quant.
intercept	-0.1137	0.0513	-0.2161	-0.1131	-0.0139
emp	0.0159	0.0273	-0.0374	0.0161	0.0687
edu-prim	0.1130	0.0293	0.0551	0.1130	0.1699
edu-sec	0.4072	0.0719	0.2655	0.4087	0.5508
urban	0.1804	0.0445	0.0928	0.1814	0.2659
sex-male	-0.1164	0.0260	-0.1669	-0.1161	-0.0660
variance parameter	Mean	Sd	2.5%-Quant.	Median	97.5%-Quant.
Var(age)	0.0065	0.0073	0.0013	0.0043	0.0238
var(mbmi)	0.0018	0.0022	0.0003	0.0011	0.0078
var(district-str)	0.0333	0.0177	0.0094	0.0294	0.0767
var(district-unstr)	0.0082	0.0060	0.0008	0.0069	0.0232
Scale estimate	mean	Sd	2.5%-Quant.	Median	97.5%-Quant.
Sigma2	0.8022	0.0165	0.7707	0.8021	0.8359

Table 5: Parameters and Variances estimates of model II

Parameter	Mean	Sd	2.5%-Quant.	Median	97.5%-Quant.
intercept	-0.2387	0.0943	-0.4173	-0.2346	-0.0621
emp	0.0156	0.0279	-0.0373	0.0161	0.0700
edu-prim	0.1044	0.0302	0.0491	0.1044	0.1648
edu-sec	0.3916	0.0741	0.2489	0.3937	0.5367
urban	0.1979	0.0453	0.1128	0.1966	0.2861
sex-male	-0.1189	0.0265	-0.1708	-0.1188	-0.0651
age	-0.0150	0.0008	-0.01651	-0.0151	-0.0137
mbmi	0.0230	0.0041	0.0151	0.0228	0.0314
variance parameter	Mean	Sd	2.5%-Quant.	Median	97.5%-Quant.
var(district-str)	0.0321	0.0189	0.0195	0.0272	0.0848
var(district-unstr)	0.0086	0.0063	0.0008	0.0074	0.0242
Scale estimate	mean	Sd	2.5%-Quant.	Median	97.5%-Quant.
Sigma2	0.8627	0.0181	0.8292	0.8628	0.9008

The posterior means of the parameters and their corresponding 95% credible intervals are shown in Table 4. Children of mothers who had both primary and secondary education had higher Z-score (implying less stunting) compared to those without education and the difference was significant. Being a male child was associated with more stunting; children who live in urban areas also had less stunting compared to those in rural areas; and there was no significant difference in stunting between children of working mothers and non-working mothers.

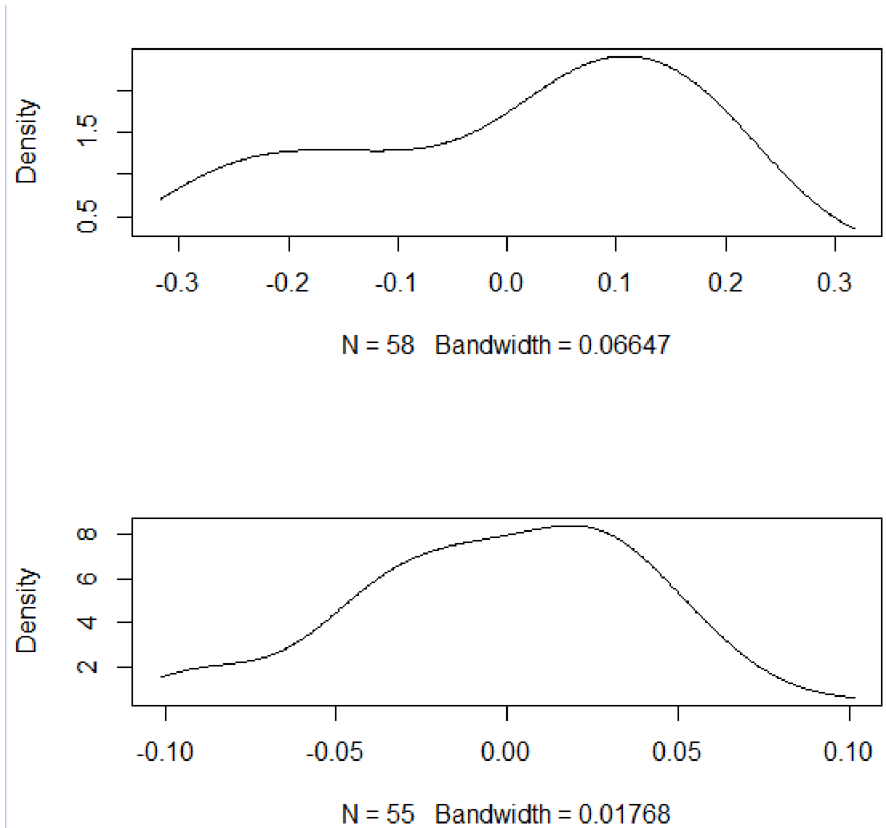

Figure 9: Kernel density estimates of the mean of the structured, top, and the unstructured spatial effect, bottom

.

Kernel density estimates or shaded maps are used to display the posterior means for the structured and unstructured spatial effects of the district covariates. In Figure 9, the Kernel density are assumed to follow a Gaussian distribution and the range of the structured spatial effects is much larger than the range of the unstructured spatial effect. This implies that there is greater spatial dependency, i.e. global unobserved factors dominate locally unobserved factors.

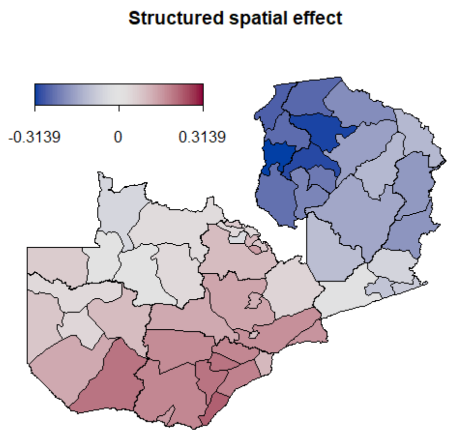

Figure 10: Map of Structured spatial random effects

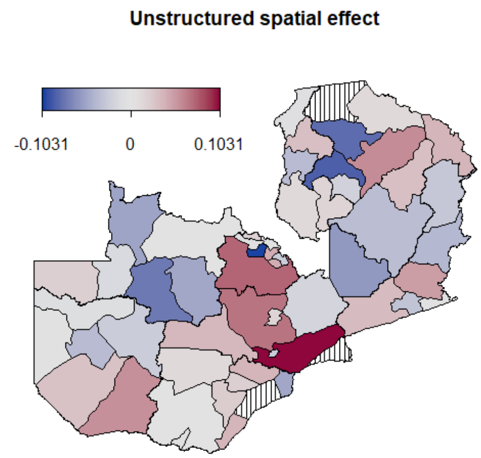

Figure 11: Map of Unstructured spatial random effects

The maps in Figures 10 and 11 display the structured and unstructured random effects respectively and the colours correspond to significantly negative (Blue colored), significantly positive (Pink colored) and insignificant (Grey colored). The structured spatial effects are dominant over the unstructured as confirmed by the former having wider range of effects (-0.3139 to 0.3139 vs -0.1031 to 1.1031). Thus, there is presence of a very large spatial dependency for all nearby districts in each of the 3 regions (North, Centre and South). From Figure 10, it can be observed that there is a strong South-North difference in these regional effects with the center of the country dividing the two regions. It can be inferred that children living in districts in the central and southern regions have less stunting than those living in the north with children living in districts of the southern region suffering the least stunting by far.

After controlling for the structured spatial and other model effects, it can be observed from Figure 11 that there are variations in the degree of stunting for neighbouring districts in each region. For example, one of the districts in the south-central region (dark-pink coloured district) corresponds to major urban areas such as the capital city, has considerably less stunting than nearby districts in the central region. But it can also be seen in one of the districts in the north-central region (small dark-blue coloured district) that there is presence of high degree of stunting compared to its neighbouring central districts. These observations of both districts indicate that there are local unobserved variables at play.

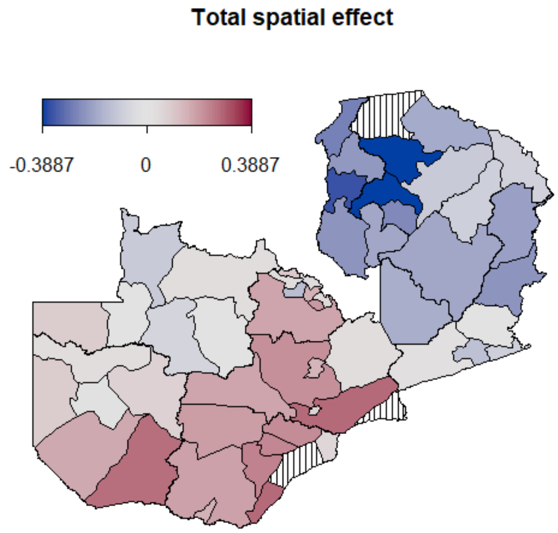

Figure 12: Map of the sum of spatial random effects

.

The map of Figure 12 shows the sum of the structured and unstructured spatial effects for the district covariate. It can be seen that it has very similar pattern to the structured spatial effects of Figure 10 indicating the large dominance of the structured spatial effects over the unstructured effects.

3.3 Sensitivity Analysis

Table 6: Summary of the sensitivity analysis of the choice of hyperparameters for Model I

Parameter	$\frac{\mathbf{a=0.001}}{\mathbf{b=0.001}}$	$\frac{\mathbf{a=0.01}}{\mathbf{b=0.01}}$	$\frac{\mathbf{a=0.5}}{\mathbf{b=0.0005}}$	$\frac{\mathbf{a=1}}{\mathbf{b=0.005}}$
*spatial effects**				
$\tau_{\mathbf{str}}{}^2$	0.0333	0.0368	0.0352	0.0359
	(0.0094-0.0767)	(0.0111-0.0821)	(0.0105-0.0823)	(0.0102-0.0073)
$\tau_{\mathbf{unstr}}{}^2$	0.0082	0.0075	0.0080	0.0076
	(0.0008-0.0232)	(0.0007-0.0223)	(0.0008-0.0229)	(0.0007-0.0224)
*smooth functions***				
$\tau_{\mathbf{age}}{}^2$	0.0065	0.0060	0.0064	0.0065
	(0.0013-0.0238)	(0.0012-0.0212)	(0.0012-0.0239)	(0.0013-0.0258)
$\tau_{\mathbf{BMI}}{}^2$	0.0018	0.0018	0.0020	0.0017
	(0.0003-0.0078)	(0.0003-0.0082)	(0.0003-0.0096)	(0.0003-0.0073)

*Variance components and 95% credible intervals for the spatially structured and unstructured effects; **Variance components and 95% credible intervals for the nonlinear smooth functions.

The estimated regression parameters depend on the choice of hyperparameters, as a consequence Model I was ran again to investigate the results sensitivity to different choices of hyperparameters. The different hyper-parameters investigated are: IG (a = 0.01, b = 0.01), IG (a = 0.5, b = 0.0005) and IG (a = 1, b = 0.005) respectively. The four selected hyperparameters produced similar results for the posterior means of the fixed parameters as well as the variance components of the spatial effects and smooth functions. This implies that the model is less sensitive to different hyperparameters. Table 6 summarises the sensitivity analysis of the choices of hyper-parameters for Model I and the values of their respective variance components.

3.4 Penalised likelihood/Empirical Bayes

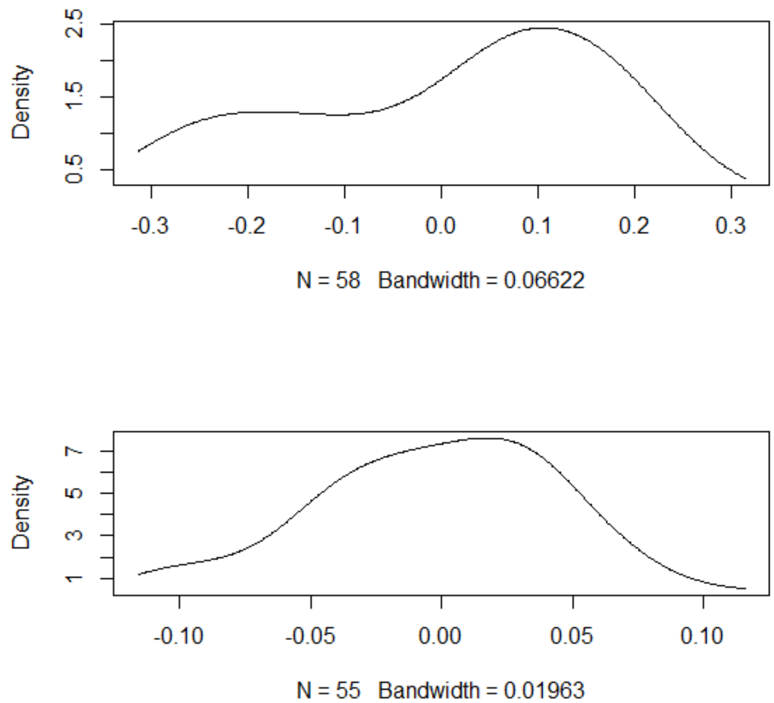

Figure 13: Kernel density estimates of the mean of the structured, top, and the unstructured spatial effect, bottom

.

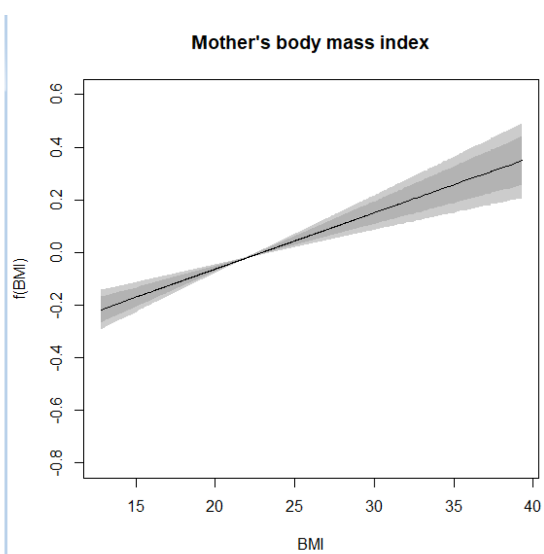

Figure 14: Non-linear Effect of 'bmi'

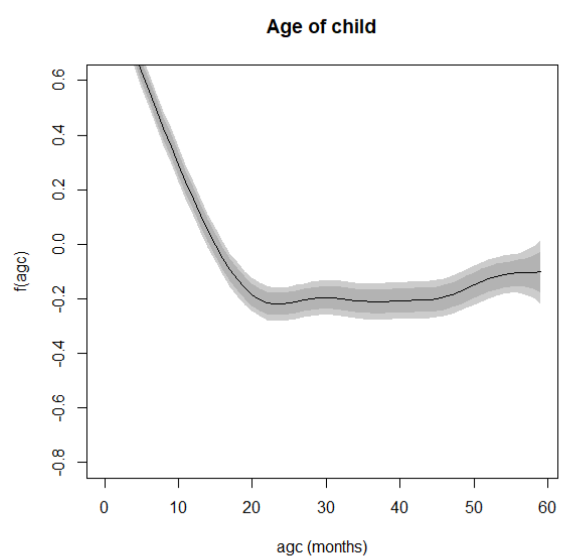

Figure 15: Non-linear Effect of 'agc'

Similar to the previous FB MCMC analysis, Figure 13 shows that the range of mean effects of structured spatial effect is much larger than unstructured effects. Figures 14 and 15 also show similar patterns to the FB MCMC for the BMI and age variables respectively. The linear covariate parameters also have similar estimates and significance to those of the FB MCMC analysis as observed in Table 7. A difference observed is that the EB inference has lower variances than the FB MCMC for both the non-linear covariates and spatial effects. As a result, EB inference produces smoother curves than FB MCMC analysis.

Table 7: Parameter and Variance estimates

Parametric coefficients				
Parameter	Mode	SE	t value	P-value
intercept	-0.1529	0.0340	-4.2589	2e-16
emp	0.0153	0.0273	0.5618	0.5743
edu-prim	0.1139	0.0291	3.9175	0.0001
edu-sec	0.4093	0.0717	5.7083	<2e-16
urban	0.1808	0.0438	4.1303	<2e-16
sex-male	-0.1171	0.0259	-4.5298	<2e-16

variance parameter	Variance	Smooth par
Var(age)	0.0032	249.0010
var(mbmi)	0.0000	69862.1000
var(district-str)	0.0294	27.2828
var(district-unstr)	0.0081	99.4393
Scale estimate: 0.8021		

Figures 16 and 17 show the maps of the structured and unstructured effects respectively. The trends observed in Figures 16 and 17 are similar to those of the FB analysis. As can be seen in Figure

16 there is a strong regional divide in stunting with the districts in the southern region having the least stunting followed by the districts in the central region and the districts in the Northern regions having the most stunting. The structured spatial effects also dominates over the unstructured spatial effects.

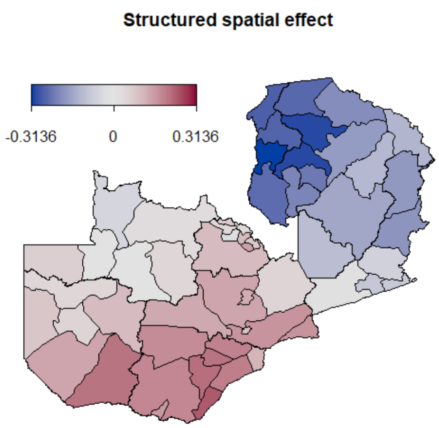

Figure 16: Map of Structured spatial random effects

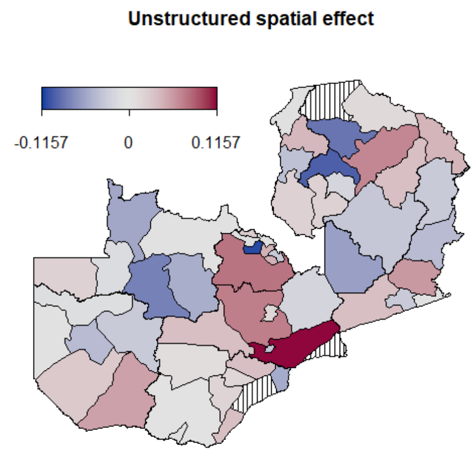

Figure 17: Map of Unstructured spatial random effects

4 Discussion and Conclusion

Mother's BMI had a positive linear relations to the z-score for stunting i.e. as mothers BMI increase stunting reduced sgnificantly in their children. Several studies (Tiga and Sen, 2016; Dharmalingam and Krishnakumar, 2010; Borah and Agarwalla, 2016) have shown that higher BMI was associated with better nutrition in a child and lowered stunting. In developing countries like Zambia, higher BMI is associated with better socio-economic status and access to better living conditions and more nutritious food choices (Tigga et al., 2015).

Increasing age of the child was found to be associated with stunting. The anthropometric status of children was found to worsen around 4-6 months age as this is the period in which solid food is introduced. This increase in stunting could be due to poor quality food replacing breastmilk as well as other factors such as poor living conditions leading to infectious diseases and parasites (WHO, 1995; Stephenson,1999).

Significant spatial variability was found in the incidence of stunting among the districts of Zambia. Districts in the Central and Southern regions had less stunting compared to the Northern Regions. This could be due to several factors prominent among them is that both the Central and Southern regions have higher concentration of urban areas such as the capital city and other major cities providing access to better socioeconomic and health facilities. Other factors causing the spatial variability may include poor climatic conditions leading to low agricultural yield, presence of infectious diseases in certain districts and poor infrastructure hindering access to health services etc (Gallup and Sachs, 1998).

In this thesis, EB estimation inference was developed using GLMM and then compared to FB MCMC inferences using a spatio-additive model. EB approach is a promising alternative to FB analysis due to its computationally efficient modification of the usual version of REML estimation of smoothing parameters. Based on the estimation results obtained herein, the two inference methods produce similar results.

5 Limitations and Recommendations

In this study, the structured spatial effects were much larger than the unstructured spatial effects as indicated by the three regional divides in the degrees of stunting. Thus, there was more global than local unobserved risk factors. Future studies should focus on unobserved global factors that led to these regional discrepancies. For example factors such as GDP, Primary health facility availability, employment opportunities, level of infrastructure development, climatic conditions, etc, for the different districts of these regions could be included in future studies to determine the cause of this sharp regional divides.

There is also presence of some unstructured spatial effects especially in one of the south central districts (having a very low stunting) and another north central district (having very high stunting) compared to their corresponding neighbours after controlling for all other factors. Policy makers should focus on local unobserved factors that could have caused such anomalies in these two districts.

6 References

1. Agretsi, A. (2002). Categorical Data Analysis. 2nd edition. Wiley and Sons, New Jersey.

2. Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). Ann. Inst. Statist. Math. 43, 1-59.

3. Dharmalingam, A., Navaneetham, K., & Krishnakumar, C. S.(2010).Nutritional status of mothers and low birth weight in India. Maternal and Child Health Journal, 14(2),290298.

4. Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing using B-splines and penalties. Statist. Sci. 11, 89-121.

5. Fahrmeir,L., Kneib,Th., & Lang,S. (2004). Penalized additive regression for spacetime data:a Bayesian perspective. Statistica Sinica 14, 715745.

6. Fahrmeir, L. and Knorr-Held, L. (2000). Dynamic and semiparametric models. In Smoothing and Regression: Approaches, Computation and Application (Edited by M. Schimek). Wiley, New York.

7. Gallup, J. and J. Sachs (1998). Geography and Economic Growth. Proceedings of the Annual Bank Conference on Development Economics. Washington DC: The World Bank.

8. Lang, S., Brezger, A. (2004). Bayesian P-Splines Journal of Computational and Graphical Statistics, 13, 183212.

9. Sen, J., Roy A., & Mondal, N. (2009).Association of maternal nutritional status, body composition and socio-economic variables with low birth weight in India. Journal of Tropical Pediatrics, 56(4), 254259.

10. Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. Journal of Econometrics, 75, 317-343.

11. Stephenson, C. (1999). Burden of Infection on Growth Failure. Journal of Nutrition, Supplement, 534S-538S.

12. Tigga, P.L, & Sen, J (2016). Maternal Body Mass Index Is Strongly Associated with Children Z-Scores for Height and BMI. Journal of Anthropology, 2016(6538235). doi.org/10.1155/2016/6538235.

13. Tigga, P.L, Sen, J., & Mondal, N (2015). Association of some socio-economic and socio-demographic variables with wasting among pre-school children of north Bengal, India. Ethiopian Journal of Health Sciences, 25(1), 6372.

14. UNICEF (1998). The State of the World's Children 1998: Focus on Nutrition. New York: UNICEF.

15. WHO (1995). Physical Status: The Use and Interpretation of Anthropometry. WHO Technical Report Series No. 854. Geneva: WHO.

16. World Bank (1995). Zambia Poverty Assessment. Washington, DC: The World Bank.

7 Appendix

Appendex I: Exploratory Data Analysis Tables

Table 8: Frequency table for Mother's employment status

Emp Status	Observations	Freq	Cum
No	2190	0.4518	0.4518
Yes	2657	0.5482	1

Table 9: Frequency table for type of locality

Locality	Observations	Freq	Cum
Rural	2745	0.5663	05663
Urban	2102	0.4337	1

Table 10: Frequency table for Education level

Level	Observations	Freq	Cum
No	2302	0.4749	0.4749
Primary	2355	0.4859	0.9608
Secondary	190	0.0392	1

Appendex II: Autocorrelation plots

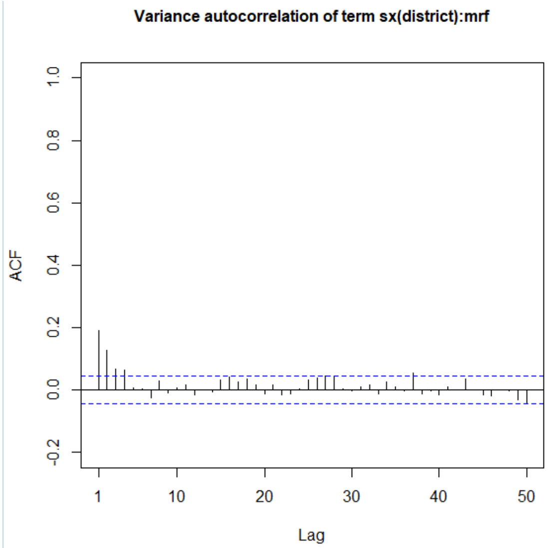

Figure 18: Autocorrelation function of the samples of the structured variance parameter

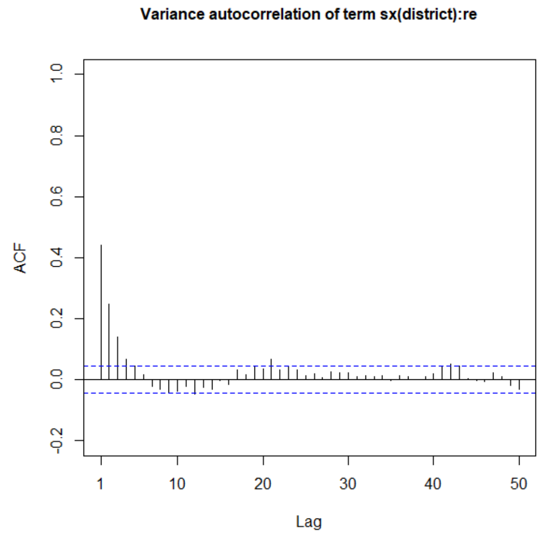

Figure 19: Autocorrelation function of the samples of the unstructured variance parameter

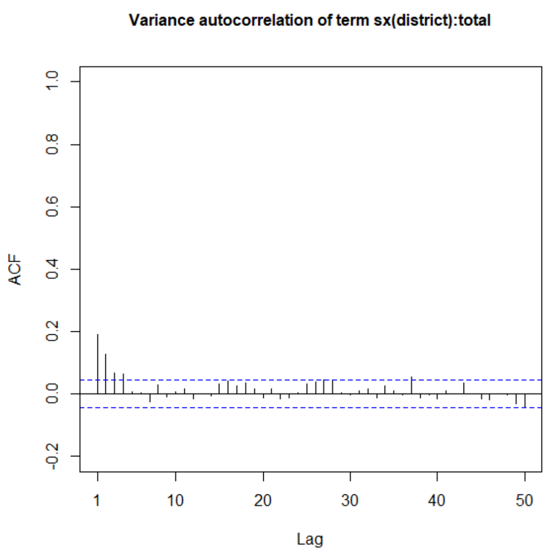

Figure 20: Autocorrelation function of the samples of the total spatial variance parameter

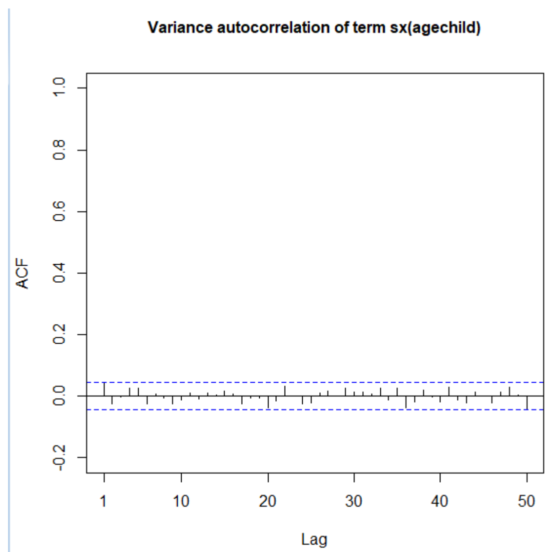

Figure 21: Autocorrelation function of the samples of the age variance parameter

Appendex III: codes

```
library("BayesX")
library("BayesXsrc")
library("R2BayesX")


## load ZambiaBnd and plot it
data("ZambiaBnd")
plotmap(ZambiaBnd)

data("ZambiaNutrition")
str(ZambiaNutrition)

####print first 10 rows of ZambiaNutrition
head(ZambiaNutrition, n=10)

zam1=ZambiaNutrition
zam1$emp=NULL
zam1$emp[zam1$memployment=="yes"]=1
zam1$emp[zam1$memployment=="no"]=0

zam2=ZambiaNutrition
zam2$edu=NULL
zam2$edu[zam2$meducation=="no"]=0
zam2$edu[zam2$meducation=="primary"]=1
zam2$edu[zam2$meducation=="secondary"]=2
```

```
zam3=ZambiaNutrition
zam3$locality=NULL
zam3$locality[zam3$urban=="no"]=0
zam3$locality[zam3$urban=="yes"]=1

zam4=ZambiaNutrition
zam4$sex=NULL
zam4$sex[zam4$gender=="female"]=0
zam4$sex[zam4$gender=="male"]=1

zam=cbind(ZambiaNutrition,"emp"=zam1$emp,"edu"=zam2$edu,
"locality"=zam3$locality,"sex"=zam4$sex)


zam=cbind(ZambiaNutrition,"emp"=zam1$emp,"edu"=zam2$edu,
"locality"=zam3$locality,"sex"=zam4$sex)


zam$memployment <- NULL
zam$urban  <- NULL
zam$gender  <- NULL
zam$ meducation <- NULL

zam$emp.f <- factor(zam$emp)
zam$edu.f <- factor(zam$edu)
zam$locality.f <- factor(zam$locality)
```

```
zam$sex.f <- factor(zam$sex)


####MODEL I######
zm <- bayesx(stunting ~ emp.f + edu.f + locality.f + sex.f +
sx(mbmi,bs = "psplinerw2", knots = 20, degree = 3) +
sx(agechild, bs = "psplinerw2", knots = 20, degree = 3) +
sx(district, bs = "mrf", map = ZambiaBnd)+ sx(district, bs = "re")
,family="gaussian", method="MCMC", hyp.prior = c(0.001, 0.001),
iter = 32000, burnin = 2000, step = 15, data = zam)
summary(zm)
AIC(zm)
####plots
plot(zm, term = c("sx(district):mrf", "sx(district):re"),main = "")
plot(zm, which = "max-acf")
plot(zm, term = "sx(mbmi)", which = "var-samples", acf = TRUE)
plot(zm, term = "sx(district):total", which = "var-samples",
acf = TRUE)
plot(zm, term = "sx(district)", which = "coef-samples")
plot(zm, term = "sx(agechild)", which = "coef-samples")
plot(zm, term = "sx(district):mrf", map = ZambiaBnd, main =
 "Structured spatial effect", digits = 4, pos = "topleft")
plot(zm, term = "sx(district):re", map = ZambiaBnd, main =
"Unstructured spatial effect", digits = 4,pos = "topleft")
plot(zm, term = "sx(mbmi)", main = "Mother's body mass index",
 xlab = "BMI",ylab = "f(BMI)", ylim = c(-0.8, 0.6), rug = FALSE)
plot(zm, term = "sx(agechild)", main = "Age of child", xlab =
```

```
"agc (months)",ylab = "f(agc)", ylim = c(-0.8, 0.6), rug = FALSE)
plot(zm, term = "sx(district):total", map = ZambiaBnd, main =
"Total spatial effect", pos = "topleft", digits = 4)

##### same model with REML
zm1 <- bayesx(stunting ~ emp.f + edu.f + locality.f + sex.f  +
sx(mbmi,bs = "psplinerw2", knots = 20, degree = 3) + sx(agechild,
 bs = "psplinerw2", knots = 20, degree = 3) +
sx(district, bs = "mrf", map = ZambiaBnd) + sx(district, bs = "re"),
lowerlim=0.001, eps=0.00001, family="gaussian",
predict =TRUE, method = "REML", data = zam)
summary(zm1)
DIC(zm1)

plot(zm1, term = c("sx(district):mrf", "sx(district):re"),
main = "")
plot(zm1, term = "sx(district):mrf", map = ZambiaBnd, main =
"Structured spatial effect", digits = 4, pos = "topleft")
plot(zm1, term = "sx(district):re", map = ZambiaBnd, main =
"Unstructured spatial effect", digits = 4, pos = "topleft")
plot(zm1, term = "sx(mbmi)", main = "Mother's body mass index",
xlab = "BMI",ylab = "f(BMI)", ylim = c(-0.8, 0.6), rug = FALSE)
plot(zm1, term = "sx(agechild)", main = "Age of child", xlab =
"agc (months)",ylab = "f(agc)", ylim = c(-0.8, 0.6), rug = FALSE)
plot(zm1, term = "sx(district):total", map = ZambiaBnd, main =
"Total spatial effect", digits = 4,pos = "topleft")
```

```
### model II########
zm2 <- bayesx(stunting ~ emp.f + edu.f + locality.f +
sex.f + mbmi+ agechild + sx(district, bs = "mrf", map = ZambiaBnd)
+ sx(district, bs = "re"), family="gaussian", hyp.prior =
 c(0.001, 0.001), iter = 32000, burnin = 2000, step = 15, data = zam)
summary(zm2)

### model II REML
zm2a <- bayesx(stunting ~ emp.f + edu.f + locality.f
+ sex.f + mbmi+ agechild + sx(district, bs = "mrf",
map = ZambiaBnd)
+ sx(district, bs = "re"),lowerlim=0.001, eps=0.00001,
 predict =TRUE,family="gaussian",
method = "REML", data = zam)
summary(zm2a)

####plots
plot(zm, term = "sx(district):mrf", map = ZambiaBnd,
pos = "topleft")



#####sensitivity analysis#######
zma <- bayesx(stunting ~ emp.f + edu.f + locality.f + sex.f
 + sx(mbmi,bs = "psplinerw2", knots = 20, degree = 3) +
```

```
sx(agechild, bs = "psplinerw2", knots = 20, degree = 3)
+ sx(district, bs = "mrf", map = ZambiaBnd)+ sx(district, bs = "re")
,family="gaussian", method="MCMC", hyp.prior = c(1, 0.005)
, iter = 32000, burnin = 2000, step = 15, data = zam)
summary(zma)

zmb <- bayesx(stunting ~ emp.f + edu.f + locality.f + sex.f
 + sx(mbmi,bs = "psplinerw2", knots = 20, degree = 3) +
sx(agechild, bs = "psplinerw2", knots = 20, degree = 3) +
sx(district, bs = "mrf", map = ZambiaBnd)+ sx(district, bs = "re")
,family="gaussian", method="MCMC", hyp.prior = c(0.5, 0.0005),
 iter = 32000, burnin = 2000, step = 15, data = zam)
summary(zmb)

zmc <- bayesx(stunting ~ emp.f + edu.f + locality.f + sex.f +
 sx(mbmi,bs = "psplinerw2", knots = 20, degree = 3) +
sx(agechild, bs = "psplinerw2", knots = 20, degree = 3) +
sx(district, bs = "mrf", map = ZambiaBnd)+ sx(district, bs = "re")
,family="gaussian", method="MCMC", hyp.prior = c(0.01, 0.01),
iter = 32000, burnin = 2000, step = 15, data = zam)
summary(zmc)
```