Ludwig Fahrmeir & Khaled Khatab

Geoadditive Latent Variable Modelling of Child Morbidity and Malnutrition in Nigeria

Technical Report Number 020, 2008
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Geoadditive Latent Variable Modelling of Child Morbidity and Malnutrition in Nigeria

Ludwig Fahrmeir and Khaled Khatab

Department of Statistics, University of Munich

SUMMARY

Investigating the impact of important risk factors and geographical location on child morbidity and malnutrition is of high relevance for developing countries. Previous research has usually carried out separate regression analyses for certain diseases or types of malnutrition, neglecting possible association between them. Based on data from the Nigeria Demographic and Health Survey of 2003, we apply recently developed geoadditive latent variable models, taking cough, fever and diarrhea as well as stunting and underweight as observable indicators for the latent variables morbidity and mortality. This allows to study the common impact of risk factors and geographical location on these latent variables, thereby taking account of association within a joint model. Our analysis identifies socio-economic and public health factors, nonlinear effects of age and other continuous covariates as well as spatial effects jointly influencing morbidity and malnutrition.

Key words: Developing countries, geoadditive regression, latent variable models, child morbidity and mortality.

1 Introduction

Childhood morbidity and malnutrition are among the most serious health issues facing developing countries. Essential reduction of malnutrition rates and of disease prevalence is one of the central resolutions of the Millennium Summit (UN, 2000). Therefore, investigation of the effects of socio-economic, public health and individual factors as well as geographical location on morbidity and malnutrition is of high relevance.

Previous analyses are often based on Demographic and Health Surveys (DHS) as a well-established data source with reliable information on childhood diseases and undernutrition, and they rely on statistical inference with various forms of regression models. Because of methodological restraints, it is difficult to detect nonlinear covariate effects, for example of age, adequately, and it is impossible to recover small-scale, district-specific spatial effects with common linear regression or correlation analysis. Recent research has therefore applied geoadditive regression models (Fahrmeir and Lang, 2001; Fahrmeir, Kneib and Lang, 2004). These models can account for nonlinear covariate effects and geographical variation while simultaneously controlling for other important risk factors. They have been used in regression studies of risk factors for acute or chronic undernutrition (e.g., Kandala et al., 2001; Adebayo 2003; Khatab, 2007), for morbidity (Kandala, Magadi and Madise, 2006; Kandala et al., 2007; Khatab, 2007) and for mortality (Adebayo and Fahrmeir, 2005; Kandala and Ghilagaber, 2007).

However, in all these studies regression analyses are carried out separately for certain types of disease, such as cough or fever, or of undernutrition such as stunting, wasting or underweight, neglecting possible association among these response variables and without aiming at the detection of common latent risk factors. In this paper, we take a somewhat different point of view: We apply recently developed geoadditive latent variable models for mixed continuous and discrete responses (Fahrmeir and Raach, 2007), considering binary indicators for cough, fever and diarrhea as well as Z-scores for stunting and underweight as observable outcomes of latent health and nutrition status. This allows to simultaneously account for association between these indicators and to assess the common influence of certain risk factors, nonlinear effects of covariates such as age of child, and geographical variation on the latent variables morbidity and malnutrition. As a case study, we focus our empirical analysis on data from the 2003 DHS for Nigeria, but the methodological approach is, of course, appropriate for other countries as well.

Section 2 provides more descriptive and exploratory details of the data, and Section 3 explains the concepts of geoadditive latent variable modelling as needed in our analyses. Section 4 contains the results, including discussions.

2 The study data

The analyses in this paper are based on data available from the 2003 DHS for Nigeria. The DHS uses standard survey instruments to collect data on household members such as sex of child, age of child, mother's age, current employment status of mother, mother's educational attainment, exposure

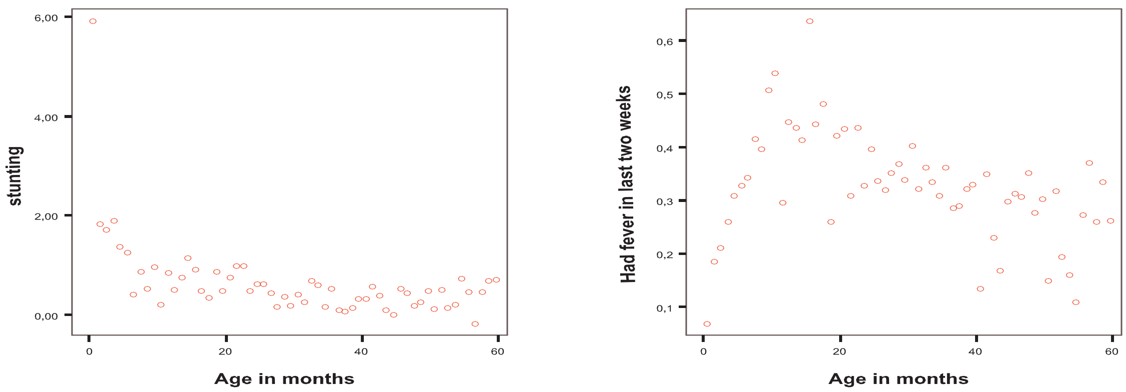

Figure 1: Non-linear effects from child's age on stunting (left) and Non-linear effects from child's age on fever (right).

to mass media, the type of toilet facility etc. It collects information on household living conditions such as housing characteristics, on childhood morbidity, malnutrition and child health from mothers in reproductive ages (15-49). The data is based on national samples that have been collected using questionnaires and allows for breakdowns by urban-rural and major regions. There are 6029 children's records in the 2003 survey of Nigeria.The NDHS samples are drawn through stratified clustered sampling with draws of 399 clusters in 37 states.

With regard to measures of health and nutritional status for children under 5 years, the focus in this work and in the analysis will be on the following: (1) child morbidity indicated by prevalence of diarrhea, fever and cough with difficulty of breathing (a symptom of respiratory infection) and (2) child malnutrition measured through z-scores for stunting, wasting and underweight.

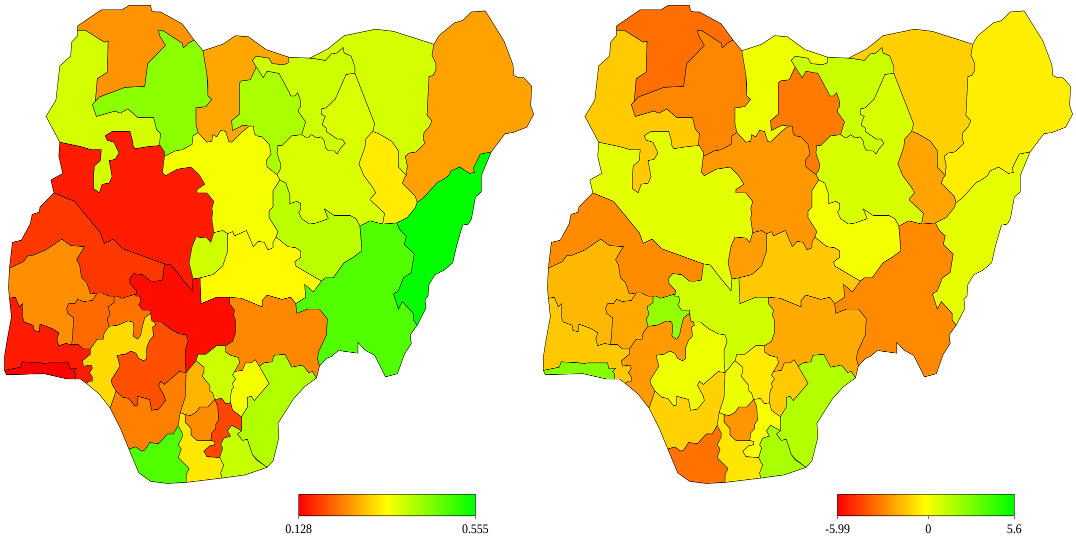

Figure 2: Distribution of fever (left) and distribution of stunting (right).

Childhood Disease

The diseases of children included in this work for Nigeria are diarrhea, cough, and fever. These diseases are still a major cause of morbidity and mortality among children in many developing countries, particularly in Sub-Saharan Africa. The success of health care intervention depends on a correct understanding of the socioeconomic, environmental and cultural factors that determine the occurrence of diseases, undernutrition and deaths. The mapping of variation in risk of child morbidity can help in improving the targeting of scarce resources for public health interventions. Bearing in mind that direct mapping of relevant environmental risk factors (which may vary considerably in both space and time) is difficult, this has led to investigations of environmental proxies (Kandala et al., 2007).

In our latent variable model approach, we consider the binary responses child had diarrhea (or not), child had cough (or not), and child had fever (or not) within two weeks before the interview as observable indicators for a child's health status.

Childhood Malnutrition

Childhood undernutrition is amongst the most serious health issues facing developing countries. It is an intrinsic indicator of well-being, but it is also associated with morbidity, mortality, impaired childhood development, and reduced labor productivity (Sen, 1999; UNICEF, 1998; Pelletier; 1998, Svedberg 1999).

To assess nutritional status, the 2003 DHS obtained measurements of height and weight for all children below five years of age. Researchers distinguish between three types of malnutrition: wasting or insufficient weight for height indicating acute malnutrition; stunting or insufficient height for age indicating chronic malnutrition; and underweight or insufficient weight for age which could be a result of both stunting and wasting.

These three anthropometric variables are measured through z-scores for wasting, stunting and underweight, defined by

$$Z_i = \frac{AI_i - MAI}{\sigma}, \tag{1}$$

where AI refers to the individual anthropometric indicator (e.g. height at a certain age), MAI refers to the median of a reference population, and σ refers to the standard deviation of the reference population.

Note that higher values of a Z-score indicate better nutrition and vice versa. Therefore, a decrease of Z-scores indicates an increase in malnutrition (and vice versa). This has to be taken into account when interpreting the results.

Based on preliminary analyses with separate geoadditive regression models (Khatab, 2007, chapters 5 and 6), the following covariates were considered to study childhood morbidity and malnutrition in Nigeria.

Metrical covariates

Chage: Child's age in months.

BMI: Mother's body mass index.

Mageb: Mother's age at birth.

Categorical covariates (in effect coding)

male: Child's sex : male or female (reference category).

educ: Mother's educational attainment: incomplete primary, complete primary, and incomplete secondary school; or complete secondary school and higher eduction (reference category).

trepr: Whether mother had treatment during pregnancy: yes or no (reference category).

anvis: Whether mother had antenatal care: yes or no (reference category).

water: Source of drinking water: controlled water or no (reference category).

toilet: Flush toilet at household, or not (reference category).

urban: Locality where respondent lives : urban or rural (reference category).

radio: Radio at household: yes or no (reference category).

elect: Electricity : yes or no (reference category).

work: Mother's current working status: working or not (reference category).

Spatial covariate

reg: state in Nigeria where the respondent resides (see figure 2).

Figure 1 shows scatter plots of child's age versus relative frequencies of fever per month (left) and age versus the average z-score of stunting per month (right). It is obvious that the (empirical) effect of age on prevalence of fever and on stunting, respectively, is highly nonlinear. Figure 2 displays maps of Nigeria, with relative frequencies of fever for the 37 states (left) and average z-score for stunting (right). Obviously, there is considerable geographical variation in prevalence of fever and in stunting.

Scatter plots and maps for the other disease and malnutrition indicators exhibit similar nonlinearity and spatial variability, but are not shown here.

3 Geoadditive latent variable models for mixed binary and continuous responses

Latent variable models are defined hierarchically: In the first stage, a measurement model relates observable responses or indicators to latent, unobservable variables and, possibly, to covariates influencing responses directly. The second stage assumes a structural model for the latent variables. In a Bayesian setting, unknown parameters or functions appearing in both stages are supplemented through appropriate priors.

3.1 Measurement models for observed responses

In latent variable models p responses y_j, $j = 1, .., p$, are observed, mostly together with covariates, for a sample of n individuals. The observed values y_{ij}, $j = 1, .., p$, of responses for each individual i are contained in the p-dimensional vector $y_i = (y_{i1}, .., y_{ip})'$. Responses may be of different type such as continuous or categorical. In our application we consider the continuous responses stunting and underweight related to malnutrition, and the binary disease indicators fever, cough and diarrhea related to morbidity. Therefore, we will focus on continuous and binary responses, and we assume that the response vector y_j consists of p_1 continuous and $p_2 = p - p_1$ binary components.

For a continuous component $y_{ij}, j = 1, .., p_1$, we assume a linear measurement model

$$y_{ij} = \alpha_j' u_i + \lambda_j' \upsilon_i + \varepsilon_{ij} \quad , i = 1, .., n \tag{2}$$

with Gaussian i.i.d.errors $\varepsilon_{ij} \sim N(0, \sigma_j^2)$. In (2), u_i is a vector of covariates with direct parametric effects α_j. Somewhat more generally, the covariates may also depend on j, but we do not make use of this extension in our application. The vector $\upsilon_i = (\upsilon_{i1}, ..., \upsilon_{iq})'$ contains the (unobserved) realization of the latent variables $(\upsilon_1, ..., \upsilon_q)'$, $q < p$, for individual i, and $\lambda_j = (\lambda_{j1}, .., \lambda_{jq})'$ are the effects of latent variables (or factor loadings) on the observable responses. In other words, (2) is a linear model augmented by adding the effects $\alpha_j' \upsilon_i$ of observable covariates. Equivalently, model (2) can be restated as

$$y_{ij} | \upsilon_i \sim N(\eta_{ij}, \sigma_j^2) \tag{3}$$

with linear predictor

$$\eta_{ij} = \alpha_j' u_i + \lambda_j' \upsilon_i = E(y_{ij} | \upsilon_i) \tag{4}$$

It is also assumed that errors ε_{ij} for different responses are independent. This implies that association among different responses is introduced through common latent variables. Thus, different responses are conditionally independent given latent variables, but they are correlated marginally.

Measurement models for binary responses $y_{ij}, j = p_1 + 1, ..., p$, are defined in complete analogy to (3): We assume that responses are conditionally independent and Bernoulli-distributed, i.e.

$$y_{ij}|\upsilon_i \sim Bin(1, \pi_{ij}), \quad j = p_1 + 1, .., p, \tag{5}$$

and follow a probit model

$$\pi_{ij} = P(y_{ij} = 1|\upsilon_i) = \phi(\alpha_j' u_i + \lambda_j' \upsilon_i), \tag{6}$$

with the same latent variables υ_i, as in (4) but with other effects α_j and factor loadings λ_j. We also assume that all responses $y_{ij}, j = 1, .., p$, are conditionally independent given the latent variables υ_i, so that association between responses is introduced through the common latent variables.

Latent variable probit models (5), (6) can also derived from the so-called underlying variable approach (UVA): The value $y_{ij} \in \{0, 1\}$ of binary response is generated by an underlying continuous variable y_{ij}^* through the threshold mechanism

$y_{ij} = 1 \Longleftrightarrow y_{ij}^* > 0$ and $y_{ij} = 0$ else.

Assuming a Gaussian linear measurement model for y_{ij}^* as in (2), with error variance $\sigma_j^2 = 1$ for identifiability reasons, leads to the probit model (5), (6). The UVA is not only useful to derive probit (and other categorical regression) models, see e.g. Fahrmeir and Tutz (2001), but also for constructing Gibbs samplers for Bayesian inference as in subsection 3.4.

3.2 Structural models for latent variables

Structural models relate latent variables to further covariates which have only indirect effects on the observable responses. Traditional linear structural models (e.g. Mostaki et al., 2004; see also Skrondal and Rabe-Hesketh, 2004) assume (latent) Gaussian linear models

$$\upsilon_{ir} = x_i' \beta_r + \delta_{ir}, \quad r = 1, .., q \tag{7}$$

with i.i.d Gaussian errors $\delta_{ir} \sim N(0, 1)$. Here x_i is the vector of covariates with direct effects on the latent variables. For identifiability reasons β_r must not contain an intercept term (it is included already in the measurement model), and also the error variance var (δ_{ir}) has to be fixed to 1.

The linear structural model (7) implies that the means of the latent variables are linearly dependent on the covariates x_i. This can be a severe restriction in real-life research settings as in our application: First, for continuous covariates, such as age of child, body mass index of mothers, and age of mothers at birth, the effect of a strictly linear effect on the mean may not be appropriate (compare figure 1). Second, the latent variables "malnutrition" and "morbidity" may be influenced by geographically varying effects (compare figure 2). To incorporate those issues we employ more versatile geoadditive structural models

$$\upsilon_{ir} = x_i'\beta_r + f_{r1}(z_{i1}) + ... + f_{rk}(z_{ik}) + f_{r,geo}(s_i) + \delta_{ir}, \quad r = 1,..,q \qquad (8)$$

where $f_{r1}(z_{i1}), ..., f_{rk}(z_{ik})$ are smooth, nonparametric functions (effects) of continuous covariates $z_1, .., z_k$ like age of child etc., and $f_{r,geo(s_i)}$ is the geographical (spatial) effect of location or geographical region $s_i \in \{1, .., S\}$, where individual i lives.

Such geoadditive models have been previously suggested for observable univariate responses y of different type, see e.g. Fahrmeir et al. (2004), and have been applied for analyzing malnutrition or disease indicators separately, see e.g. Kandala et al (2007) and Khatab (2007, chs 5,6). Geoadditive latent variable models, combining separate regression models to a joint multivariate model, have been suggested recently in Fahrmeir and Raach (2007) in a Bayesian framework. The next section shortly reviews modelling of the unknown function $f_1, .., f_k$ and f_{geo} and points out some identifiability issues.

3.3 Priors and identifiability

Priors for regression coefficients α_j and β_j are flat, i.e. $p(\alpha_j) \propto 1$, $p(\beta_j \propto 1)$, or weakly informative Gaussian, which is the standard prior in linear regression models. Similarly, inverse Gamma priors are usually chosen for error variances σ_j^2 in Gaussian measurement models.

Concerning factor loadings λ_j, we first have to deal with the well known identifiability problem in factor analysis and latent variables models: Any transformation from λ_j' to $\widetilde{\lambda_j'} = \lambda_j' V$ and υ_i to $\widetilde{\upsilon}_i = V'\upsilon_i$ with an orthogonal matrix V leads to the same predictor because $\widetilde{\lambda}'_j \widetilde{\upsilon}_i = \lambda'_j \upsilon_i$. To avoid this identifiability problem we choose the matrix $\Lambda = (\lambda'_1, .., \lambda'_p)'$ of factor loadings to be a lower block triangular matrix of full rank and positive diagonal elements as recommended by Geweke and Zhou (1996) and Aguilar and West (2000). To avoid so-called Heywood cases, we assume a standard normal prior for these factor loadings, which is a standard choice in applications.

The nonparametric effects $f_{r1}, .., f_{rk}$ for continuous covariates $z_1, .., z_k$ in the structural equations (8) are modelled as (Bayesian) P-splines: Dropping indices to simply the notation, a function f is approximated through a polynomial spline

$$f(z) = \Sigma_{c=1}^{d} \gamma_c B_c(z), \tag{9}$$

where $B_1(z), .., B_d(z)$ are B-spline basis functions. Smoothness of the function f is achieved by assuming a (second order) random walk model.

$$\gamma_{c=1} - 2\gamma_{c-1} + \gamma_{c-2} = u_c \sim N(0, \tau^2) \tag{10}$$

for the sequence of B-spline coefficients. The variance τ^2 controls the amount of smoothness and is estimated (together with all other parameters) by assuming an inverse Gamma prior.

More information about Bayesian P-spline regression is given in Lang and Brezger (2004) and Fahrmeir et al.(2004).

The geographical effects $f_{geo}(s)$ for regions $1, .., S$ are modelled through Markov random field priors, a popular model in disease mapping (Besag, York and Molie, 1991) and in spatial statistics (Rue and Held, 2005). The basic idea is that adjacent regions should have a similar impact on the latent variables, whereas two regions far apart from each other need not exhibit such a similarity. We assume the standard Markov random field prior

$$f_{geo}(s)|f_{geo}, s' \neq s \sim N\left(\Sigma_{s' \in N_s} \frac{f_{geo(s')}}{n_s}, \frac{\tau^2_{geo}}{n_s}\right), \tag{11}$$

where $N(s)$ is the set of neighboring regions s' of s, i.e. share a common boundary with region s, and n_s is the number of neighboring regions. Hence, the conditional mean of $f_{geo(s)}$ is an average of the spatial effects $f_{geo(s')}$ of all adjacent regions. As for P-splines, the variance τ^2_{geo} controls smoothness of geographical effects and, again, obeys an inverse Gamma prior.

3.4 Bayesian MCMC inference

Gathering all regression coefficients and spatial effects into vectors α, β, υ and f_{geo}, variances in vectors σ^2, τ^2, factor loadings in λ and latent variables in $\upsilon = \upsilon\{\upsilon_{ir}, i = 1,..,n \quad r = 1,..,q\}$. Bayesian inference is based on the posterior $p(\alpha, \beta, \gamma, \sigma^2, \tau^2, \lambda, \upsilon|y)$ given the observable data y. Deriving full conditionals for Gibbs sampling can be carried out as described in detail in Fahrmeir and Raach (2007) to estimate parameters, functions, factor loadings and latent variables through their empirical posterior means together with credible intervals. In particular, data augmentation, through UVA is used here as an essential computational tool.

4 Statistical Analyses and Results

In this section, we investigate how the indicators of diseases and undernutrition can be interpreted as indicators of latent variables for morbidity and malnutrition and how much of the variation of latent variables can be explained through common predictors. This concept does not only allow us to analyze the impact of covariates on health and nutritional status, but also allows us to account for correlation among the indicators. In order to decide which of the covariates should be included in the measurement model as direct parametric effects, or in the structural equation as indirect effects via their impact on the latent variables, the results in Khatab (2007, chapter 5 and chapter 6), are taken into account.

Based on these analyses, the binary covariates *male, educ, radio* and *water* are included in the covariate vector u with direct parametric effects $\alpha_j, j = 1, ..., 5$, on the continuous and binary indicators in the measurement models (1) and (5) respectively. The remaining binary covariates *urban, work, trepr, anvis, toilet* and *elect*, are included in the vector x of covariates with indirect linear effects in the geoadditive structural model (7), while the effects of the continuous covariates chage (child's age in months), BMI (of mother) and mageb (mother's age at birth) are modelled through smooth functions

Parameter	Mean	Std	2.5%	97.5%
		Factor Loadings		
1. Fever λ_{11}	0.821^*	0.081	0.682	0.989
2. Cough λ_{21}	0.651^*	0.063	0.538	0.781
3. Diarrhea λ_{31}	0.896^*	0.084	0.741	1.087
4.Stunting λ_{41}	-0.262^*	0.046	-0.348	-0.171
5. Underweight λ_{51}	-0.21^*	0.028	-0.246	-0.136
		Parametric Indirect Effects		
urban	-0.179^*	0.079	-0.326	-0.017
work	0.004	0.070	-0.126	0.147
trepr	0.204^*	0.074	0.053	0.331
anvis	-0.039	0.085	-0.204	0.126
toilet	-0.111	0.100	-0.325	0.078
elect	-0.018	0.077	-0.171	0.127

Table 1: Estimates of factor loadings, and parametric indirect effects of the LVM with one latent variable.

f_1, f_2 and f_3, and the effect of geographical location reg (state in Nigeria) is described through the geographical effect f_{geo} in (7).

4.1 Model Estimation with One Latent Variable

In a first attempt we analyzed the data with a LVM with the five observables fever, cough, diarrhea, stunting and underweight as indicators for one common latent variable "morbidity and malnutrition". Looking at the estimated mean factor loadings in Table 1 we can draw the following conclusions: First, the latent variable has significant effect on all five indicators. Second, as we expected, disease and malnutrition indicators are positively associated. (Strictly speaking, disease indicators and z-scores for stunting and underweight are negatively correlated, because by definition z-scores for stunting and underweight decrease with increasing undernutrition.). Third, however, the latent variable loads much higher onto the disease indicators than onto the malnutrition indicators. Therefore, we reanalyze the data with a LVM with two latent variables in the next subsection. Yet, for comparison, we show estimation results for parametric indirect effects in Table 1, and nonlinear as well as spatial effects in Figure 3. Because the latent variable loads mainly on the disease indicators, these results are comparably close to the ones obtained with the results for the LVM with two latent variables, so we defer interpretation to the following subsection.

4.2 Model Estimation with Two Latent Variables

In a model with the same covariates as before, but with two latent variables υ_1 and υ_2, the predictors in the measurement models are now

$$\alpha_j' u_i + \lambda_{j1}' \upsilon_{i1} + \lambda_{j2}' \upsilon_{i2} \quad , j = 1, .., 5 \tag{12}$$

and there are two structural models

$$\upsilon_{i1} = x_i' \beta_1 + f_{11}(chage) + f_{12}(BMI) + f_{13}(mageb) + f_{1,geo}(reg) + \delta_{i1} \tag{13}$$

$$\upsilon_{i2} = x_i' \beta_2 + f_{21}(chage) + f_{22}(BMI) + f_{23}(mageb) + f_{2,geo}(reg) + \delta_{i2} \tag{14}$$

(Note that $\lambda_{12} \equiv 0$ for identifiability reasons as mentioned in subsection 3.3)

The factor loadings estimates (table 2) show that the first latent variable loads onto the first three indicators, whilst indicators four and five are explained by the second latent variable. This was to be expected, because the two different sets of indicators are supposed to measure two different latent constructs. Both factor loadings and coefficients of the parametric indirect covariates of the first latent factor are very similar to the estimates of the single latent factor model in 4.1.

The factor loadings estimate of the second latent variable, it has a higher influence on the *stunting* with factor loading of 1.16. Obviously, the first latent variable can be interpreted as "morbidity", and the second one as "malnutrition".

The estimation results for parametric indirect effects can be interpreted as follows: Living in an urban area significantly lowers the risk of morbidity, but it has no significant effect on nutritional status. Mothers' working status has only negligible effect on morbidity, but there is evidence that the nutritional status is better with a working mother. Treatment during pregnancy significantly increases the risk for diseases. At first view this seems to be strange. A possible explanation is that treatment during pregnancy was necessary because of diseases of the mother, and that children "inherit" higher risk of morbidity from their mother. For a more thorough investigation, it would be useful if DHS data would contain information about the reasons for treatment during pregnancy. On the other hand there is some evidence that the same covariate has positive effect on nutritional status!

While the remaining covariates antenatal visit, having a flush toilet and, having electricity do not show significant effects of lowering the risk of morbidity, they have the expected positive effect on nutritional status, as also reported in previous analyses with separate regression models.

Concerning covariates with direct effects on the indicators, only education and male have significant effects. Again, the results confirm previous analyses: boys are significantly more affected from stunting (an indicator for chronic malnutrition) than girls. The positive effect (0.115) on diarrhea for boys is also highly evident, which is in line with the association between stunting and diarrhoea reported in Adebayo (2003 Ch. 4). Similarly, good education of the mother decreases the risk for diarrhea and for stunting.

The left panel of Figure 4 displays the nonlinear effects of child's age, mother's BMI and mother's age at birth on the latent factor morbidity. Shown are the posterior means together with 80% pointwise credible intervals. Comparing with the corresponding plots in Figure 3 for the analysis with one latent variable, we see that these effects are quite similar. During the first 8-10 months, the effect of child's age turns quickly from negative to positive. Afterwards, the effect decreases slowly but continuously. At about 40 months confidence intervals widen because there are less observations, caused by child mortality. The effects of mother's BMI and mother's age at birth exhibit much less linearity and do not have pronounced impact on morbidity.

The right panel shows the effects of the same covariates on the latent variable malnutrition. At first glance, the nonlinear effect of child's age seems to be inverse to the effect on morbidity. However, because decreasing z-scores indicate increasing malnutrition, the effects of child's age on malnutrition and on morbidity have a similar nonlinear pattern. The effect of mother's BMI displays the inverse U-shape often reported in studies on undernutrition. The effect of mother's age at birth seems to indicate an improvement of nutritional status when mothers become older.

The estimated spatial effects on morbidity (left map in Figure 5) reasonable to the corresponding map in Figure 3. It is obvious that there still remains a considerable amount of residual geographical variability, even after controlling for the other covariates. The maps confirm the spatial patterns found in Kandala et al. (2007) and in Khatab (2007, chapter 5) with separate geoadditive probit models for the three disease indicators.

The spatial pattern of effects on malnutrition (right map in Figure 5) indicates that malnutrition is high in the northeastern part of the country compared to regions in the southeast and some regions in the south.

Both maps provide evidence for unobserved heterogeneity and motivate the search for omitted covariates, not contained in DHS data, which might partly explain the spatial effects shown in the maps.

5 Conclusion

Geoadditive latent variable models offer new opportunities and concepts to analyze child morbidity and malnutrition in developing countries within a joint modelling framework. In our case study for Nigeria we found strong support for flexibly modelling the effect of some covariates that have non-linear influences and for including a spatial analysis. The spatial pattern points to the influence of omitted variables with a strong spatial structure. The maps could be used for targeting regional development efforts and they may highlight unexpected relationships that would be overlooked in analyses with standard regression or latent variable models.

Acknowledgment

Part of this research was financially supported by the Munich Center of Health Sciences.

References

Adebayo S.B. (2003). Semiparametric Bayesian Regression for Multivariate Responses. Ph.D Thesis, Hieronymus Verlag, Munich, Germany.

Adebayo, S. and Fahrmeir, L. (2005). Analysing Child Mortality in Nigeria with Geoadditive Survival Models. Statistics in Medicine 24, 709-728.

Aguilar, O., West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portofolio allocaion. Journal of Business and Economic Statistics 18, 338-357.

Bartholomew, D.J. (1987). Latent variable models and factor analysis. Charles Grif- fin and Company Ltd., London, Oxford University

Parameter	Mean	Std	2.5%	97.5%
Factor Loadings of First Latent Variable				
1. Fever λ_{11}	0.957*	0.083	0.821	1.146
2. Cough λ_{21}	1.032*	0.091	0.868	1.208
3. Diarrhea λ_{31}	0.77*	0.065	0.665	0.901
4. Stuntingλ_{41}	-0.025	0.048	-0.118	0.073
5. Underweight λ_{51}	-0.155*	0.037	-0.226	-0.090
Factor Loadings of Second Latent Variable				
1. Fever λ_{12}	0.000	0.000	0.000	0.000
2. Cough λ_{22}	0.253*	0.045	0.164	0.336
3. Diarrhea λ_{32}	-0.088*	0.0355	-0.1588	-0.014
4. Stunting λ_{42}	1.165*	0.028	1.109	1.224
5. Underweightλ_{52}	0.958*	0.0238	0.910	1.006
Parametric Indirect Effects of First LV				
urban	-0.144*	0.067	-0.277	-0.020
work	0.010	0.068	-0.108	0.160
trepr	0.243*	0.074	0.091	0.380
anvis	-0.037	0.076	-0.171	0.111
toilet	-0.075	0.099	-0.287	0.109
elect	-0.023	0.074	-0.170	0.121
Parametric Indirect Effects of Second LV				
urban	0.021	0.060	-0.103	0.141
work	0.105	0.060	-0.014	0.219
trepr	0.093	0.0613	-0.030	0.218
anvis	0.359*	0.063	0.234	0.474
toilet	0.143	0.080	-0.012	0.304
elect	0.159*	0.064	0.029	0.287
Parametric Direct Effects of Both LV				
male(a_{11})	-0.0073	0.066	-0.135	0.1230
educ (a_{12})	-0.039	0.078	-0.186	0.130
radio(a_{13})	-0.052	0.042	-0.137	0.034
water(a_{14})	0.041	0.103	-0.168	0.245
male(a_{21})	0.032	0.074	-0.104	0.168
educ (a_{22})	0.068	0.087	-0.100	0.242
radio(a_{23})	-0.024	0.048	-0.115	0.068
water(a_{24})	-0.056	0.110	-0.261	0.185
male(a_{31})	0.115	0.0707	-0.023	0.256
educ (a_{32})	-0.154*	0.081	-0.312	-0.006
radio(a_{33})	-0.066	0.038	-0.142	0.011
water(a_{34})	-0.06	0.102	-0.266	0.137
male(a_{41})	-0.242*	0.063	-0.374	-0.119
educ (a_{42})	0.185*	0.066	0.0565	0.326
radio(a_{43})	0.028	0.039	-0.052	0.105
water(a_{44})	0.072	0.0897	-0.102	0.230
male(a_{51})	-0.061	0.047	-0.1472	0.042
educ (a_{52})	0.048	0.054	-0.052	0.153
radio(a_{53})	0.027	0.029	-0.030	0.082
water(a_{54})	-0.005	0.072	-0.144	0.139

Table 2: Results of LVMM using 2 latent variable.

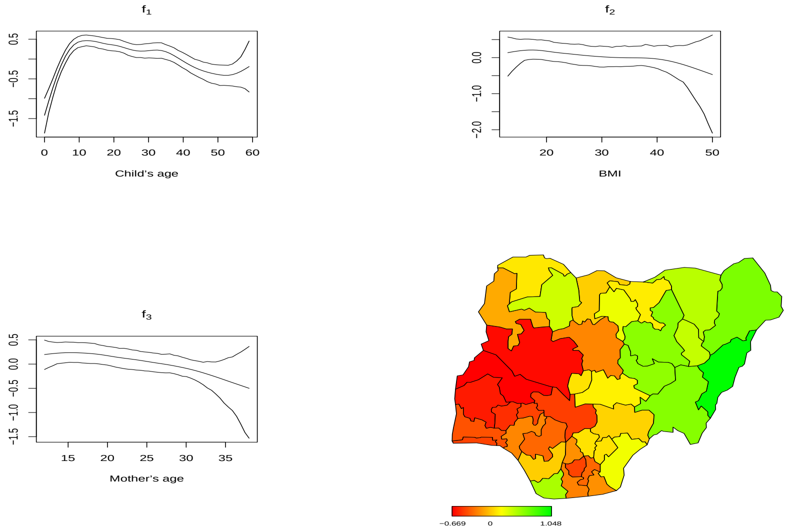

Figure 3: Non-linear effects from top to bottom: child's age, mother's BMI, mother's age at birth and spatial effects (for model LVMM using 5 indicators), on the latent variable "morbidity and malnutrition" of children for Nigeria using only one latent variable.

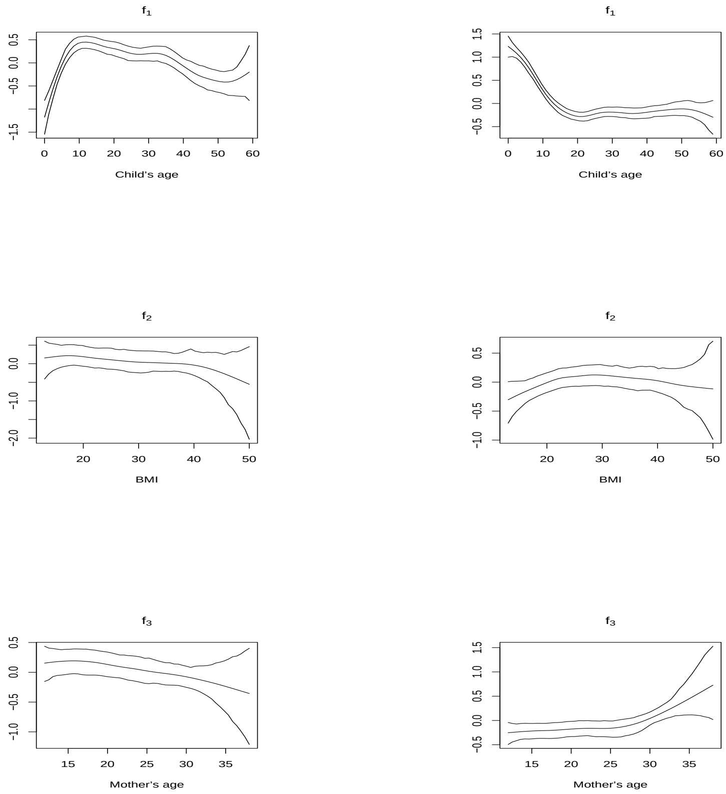

Figure 4: Estimates of nonparametric effect of nonlinear covariates from top to bottom: child's age, mother's BMI, and mother's age at birth for the first (left) and second latent variable for Nigeria.

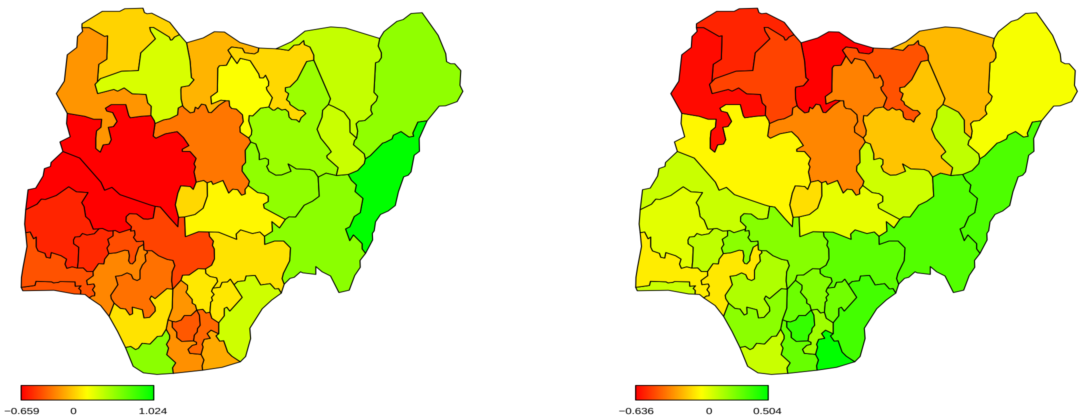

Figure 5: Estimates of the nonparametric effect of spatial covariate for the first (left) and second (right).

Press, New York.2nd edition: Bartholomew, D.J., Knott, M.(1999). Arnold, London.

BayesX (2005). Software for Bayesian inference by Brezger, A., Kneib, T., Lang, S.

Besag, J., York, Y. and Mollie, A. (1991). Bayesian Image Restoration with Two Applications in Spatial Statistics (with Discussion). Ann. Inst. Statist. Math., 43, 1-59.

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

Demographic and Health Survey (NDHS) for Nigeria (2003).

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

Fahrmeir, L., Kneib, T.and Lang, S. (2004). Penalized structured additive regression of space-time data:a Bayesian perspective. Statistica Sinica 14, 731-761.

Fahrmeir, L., and Raach, A. (2007). A Bayesian semiparametric latent variable model for mixed responses. Psychometrika, 72,327-346.

Fahrmeir, L., Tutz, G. (2001). Multivariate statistical modelling based on generalized linear models. Springer, New York.

Geweke, J., and Zhoiu, G. (1996). Measurnig the pricing error of the arbitrage pricing theory. Review of Financial Studies, 9, 557-587

Kandala, N.B.(2002). Spatial Modelling of Socio-economic and Demographic Determinants of Childhood Undernutrition and Mortality in Africa. Ph.D Thesis, Shaker Verlag , Munich, Germany.

Kandala, N.B., Chen Ji, Nigel Stallard, Saverio Stranges, and Francesco P.Cappuccio.Spatail Analysis of Risk Factors for Childhood Morbidity in Nigeria. Journal of Tropical Medicine, 77(4), 2007, pp.770-778.

Kandala, N.B., Lang, S., Klasen, S. and Fahrmeir, L. 2001. Semiparametric Analysis of the Socio-Demographic Determinants of Undernutrition in Two African Countries. Research in Official Statistics, Vol. 4 No.1:81-100.217.

Kandala N-B, Gebrenegus Ghilagabar (2006) 'A Geo-additive Bayesian Discrete time Survival Model and its application to Spatial Analysis of Childhood Mortality in Malawi.', Quality And Quantity, 935 - 957.

Khatab, K. (2007). Analysis of childhood diseases and malnutrition in developing countries of Africa. Ph.D thesis, Dr.Hut Verlag , Munich, Germany.

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

Lang, S.; Brezger, A.; Kneib, T (2005). BayesX: Analysing Bayesian structured additive regression models. Journal of Statistical Software, 14 (2005), Heft 11.

Moustaki, I., Jeoreskog, K.G. and Mavridis, D.(2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables:a comparison of LISREL and IRT approaches. Structural Equation Modeling 11, 487-513.

Pelletier, D. (1998). Malnutrition, Morbidity, and Child Mortality in Developing Countries, In: United Nations (eds.) Too Yuong too Die: Genes or Gender? New York: United Nations.

Raach, A.W. (2005). A Bayesian semiparametric latent variable model for binary, ordinal and continuous response. Dissertation, available from edoc.ub.uni-muenchen.de

Rue, H.and Held, L. (2005). Gaussian Markov Random Fields: Theory and Applications.

Sen, A. (1999). Development as Freedom. Oxford: Oxford University Press.

Skrondal, A., Rabe-Hesketh,S.(2004). Generalized latent variable modeling. Chapman and Hall.

Svedberg, P. (1996). Gender Bias in Sub-Saharan Africa: Reply and Further Evidence. Journal of Development Studies 32: 933-943.

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