Kandala, Lang, Klasen, Fahrmeir:

Semiparametric Analysis of the Socio-Demographic and Spatial Determinants of Undernutrition in Two African Countries

Sonderforschungsbereich 386, Paper 245 (2001)

Semiparametric Analysis of the Socio-Demographic and Spatial Determinants of Undernutrition in Two African Countries

Ngianga B. Kandala, Stefan Lang, Stephan Klasen and Ludwig Fahrmeir

University of Munich, Ludwigstr. 33 , Germany

Abstract

We estimate semiparametric regression models of chronic undernutrition (stunting) using the 1992 Demographic and Health Surveys (DHS) from Tanzania and Zambia. We focus particularly on the influence of the child's age, the mother's body mass index, and spatial influences on chronic undernutrition. Conventional parametric regression models are not flexible enough to cope with possibly nonlinear effects of the continuous covariates and cannot flexibly model spatial influences. We present a Bayesian semiparametric analysis of the effects of these two covariates on chronic undernutrition. Moreover, we investigate spatial determinants of undernutrition in these two countries. Compared to previous work with a simple fixed effects approach for the influence of provinces, we model small scale district specific effects using flexible spatial priors. Inference is fully Bayesian and uses recent Markov chain Monte Carlo techniques.

Keywords:
developing countries; semiparametric Bayesian inference; spatial models; undernutrition

1. Introduction

Acute and chronic undernutrition is considered to be one of the worst health problems in developing countries. As one of the most important indicators of deprivation, undernutrition is of intrinsic concern to policy-makers. In addition, it is also associated with other important development outcomes such as high mortality and poor labor productivity (Sen, 1999; UNICEF, 1998). In fact, some estimates claim that undernutrition is implicated in over 50 percent of deaths in developing countries (UNICEF, 1998).

Undernutrition among children is usually determined by assessing the anthropometric status of the child relative to a reference standard. Researchers distinguish between three types of undernutrition: wasting or insufficient weight for height indicating acute undernutrition; stunting or insufficient height for age indicating chronic undernutrition; and underweight or insufficient weight for age which could be a result of either. Wasting, stunting, and underweight for a child *i* are typically determined using a Z-score which is defined as:

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

where *AI* refers to the individual's anthropometric indicator (e.g. height at a certain age), *MAI* refers to the median of the reference population, and s refers to the standard deviation of the reference population. The reference standard typically used for the calculation is the NCHS-

CDC Growth Standard that has been recommended for international use by WHO (WHO, 1983; 1995).

Important determinants of undernutrition include the education, income, and nutritional situation of the parents, access to clean water and sanitation, and primary health care, and immunization facilities (UNICEF, 1998; Klasen, 1999; Nyovani et al., 1999). Some of these influences are likely to have non-linear effects on undernutrition. In particular, the impact of the nutritional situation of the parents, measured using the Body Mass Index (*BMI*, defined as the weight in kg divided by the square of height in meters) on the child's nutritional status is presumed to follow an inverse U-shape. Parents who exhibit a very low *BMI*, indicating their poor nourishment, are likely to have poorly nourished children. At the same time, parents with a very high *BMI* might also have poorly nourished children as the obesity associated with their high *BMI* indicates poor quality of nutrition and might therefore indicate poor quality of nutrition for their children.

Moreover, the development of undernutrition typically follows a pattern that is closely related to the age of the child. While some children are already born undernourished due to growth retardation *in utero*, the anthropometric status of children worsens considerably only after 4-6 months, when children are weaned and solid foods are introduced (WHO, 1995; Stephenson, 1999). This is due to the influence of poor quality nutrition that is replacing breastmilk as well as the onset of infectious diseases. These diseases are often related to unclean water and food which is replacing the breastmilk, and the child no longer profits from the mother's antibodies that were transmitted through the breastmilk (Stephenson, 1999). Initially, the worsening anthropometric status shows up as acute undernutrition. But then stunting develops and is held to worsen until about age 2-3. At that time, the body has, through reduced growth, adjusted to reduced nutritional intake and now needs fewer nutrients to maintain this smaller stature. In addition, the body has developed its immune system to fight the impact of infectious diseases more effectively (WHO, 1995; Moradi and Klasen, 2000).

Even after controlling for the impact from these well-known correlates, researchers have found important spatial differences in indicators such as undernutrition, or mortality in many developing countries (World Bank, 1995). They may be related to left out variables that have a distinct spatial pattern. Obvious examples of such variables are the disease environment in certain areas, the climate which may affect the quality of nutrition and the persistence of illness, access to important infrastructure (such as health centres, major roads or railroads), regional economic opportunities and constraints, etc. (Gallup and Sachs, 1998). To the extent that undernutrition is directly affected by the presence or absence of infectious diseases, such spatial patterns may also capture the spatial distribution of certain infectious epidemics.

In this paper, we model the determinants of stunting (i.e. chronic undernutrition) in Zambia and Tanzania. Stunting rates are high in both countries. Overall, 42 percent of Zambian children under age five are classified as stunted (Z score less than minus 2) and 18 percent as severely stunted (Z score less than minus 3). In Tanzania, some 43 percent of children under five are classified as stunted and 18 percent are severely stunted (Somerfelt and Stewart, 1994).

A particular focus of our analysis is to use a flexible approach to model the impact of the child's age and the mother's BMI on undernutrition as well as consider spatial effects with the help of a semiparametric Bayesian modelling approach developed by Fahrmeir and Lang (2001a,b) and Lang and Brezger (2001). In a related paper (Kandala, Lang, and Klasen, 2001), spatial effects have been included by using simple fixed effects for provinces. In the current paper spatial random effects models are used to determine small scale regional (district-specific) effects. The results give refined insight to spatial effects on undernutrition. Inference is fully Bayesian and uses recent Markov chain Monte Carlo (MCMC) techniques.

2. Semiparametric Bayesian regression models

2.1 Observation model

Consider regression situations, where observations (y_i, x_i, w_i), $i=1,...,n$, on a metrical response y, a vector $x = (x_1,...x_p)$ of metrical covariates and a vector $w = (w_1,...w_r)$ of categorical covariates are given. We assume that y_i given the covariates and unknown parameters are independent and Gaussian with mean h_i and a common variance s^2 across subjects, i.e. $y_i \sim N(h_i, s^2)$. In our application on childhood undernutrition the response is stunting measured as a Z-score. Traditionally, the effect of the covariates on the response is modelled by a linear predictor

$$h_i = x'_i\, b + w'_i\, g. \qquad (1)$$

In this paper particular emphasis is on the effects of the two metrical covariates "age of the child" *AGC* and the "mother's body mass index" *BMI*, which are possibly nonlinear, and on regional effects of the district where the mother and child live. Thus, we replace the strictly linear predictor (1) by the more flexible semiparametric predictor

$$h_i = f_1(x_{i1}) + ... + f_p(x_{ip}) + f_{spat}(s_i) + w'_i\, g.$$

Here $f_1,...,f_p$ are nonlinear smooth effects of the metrical covariates and f_{spat} is the effect of district s_i Î $\{1,...,S\}$ where mother i lives. In a further step we may split up the spatial effect f_{spat} into a spatially correlated (structured) and an uncorrelated (unstructured) effect

$$f_{spat}(s_i) = f_{str}(s_i) + f_{unstr}(s_i).$$

A rationale is that a spatial effect is usually a surrogate of many unobserved influences, some of them may obey a strong spatial structure and others may be present only locally. By estimating a structured and an unstructured effect we aim at separating between the two kinds of factors. As a side effect we are able to assess to some extent the amount of spatial dependency in the data by observing which one of the two effects is larger. If the unstructured effect exceeds the structured effect, the spatial dependency is smaller and vice versa. Such models are common in spatial epidemiology, see e.g. Besag et al. (1991).

2.2 Prior assumptions

In a Bayesian approach unknown functions f_j, $j=1,...,p$, f_{str}, f_{unstr} and parameters γ as well as the variance parameter s^2 are considered as random variables and have to be supplemented with appropriate prior assumptions. In the absence of any prior knowledge we assume independent diffuse priors g_j µ $const$, $j=1,...,r$, for the parameters of fixed effects. Another common choice are highly dispersed Gaussian priors.

Several alternatives are available for the priors of the unknown (smooth) functions $f_1,...,f_p$. For the moment we may distinguish roughly two main approaches for Bayesian semiparametric modelling. These are base functions approaches with adaptive knot selection (e.g. Denison et al.,1998, Biller, 2000, and Smith and Kohn, 1996) and approaches based on smoothness priors. In the following we will focus on the latter one. Several alternatives have been proposed for specifying a smoothness prior for the effect f of a metrical covariate x. Among others, these are random walk priors (Fahrmeir and Lang, 2001a, Kandala, Lang and Klasen 2001), Bayesian smoothing splines (Hastie and Tibshirani, 2000) and Bayesian P-splines (Lang and Brezger, 2001). In this paper we focus on P-splines.

The basic assumption behind the P-splines approach is that an unknown smooth function f of a particular covariate x can be approximated by a spline of degree l defined on a set of equally spaced knots $z_0 = x_{min} < z_1 < ... < z_{r-1} < z_r = x_{min}$ within the domain of x. It is well known that such a spline can be written in terms of a linear combination of $m = r+l$ B-spline basis functions B_t, i.e.

$$f(x) = \sum_{t=1}^{m} b_t B_t(x)$$

The basis functions B_t are defined locally in the sense that they are nonzero only on a domain spanned by $2+l$ knots. It would be beyond the scope of this paper to go into the details of B-splines and their properties, see e.g. de Boor (1978). The vector $b = (b_1,..., b_m)$ is unknown and must be estimated from the data. In a simple regression spline approach the unknown regression coefficients are estimated using standard methods for fixed effects parameters. However, a crucial point with simple regression splines is the choice of the number and the position of knots. For a small number of knots the resulting spline space may be not flexible enough to capture the variability of the data. For a large number of knots estimated curves may tend to overfit the data. As a remedy to these problems Eilers and Marx (1996) suggest a moderately large number of knots (usually between 20 and 40) to ensure enough flexibility, and to define a roughness penalty based on differences of adjacent regression coefficients to guarantee sufficient smoothness of the fitted curves. In a Bayesian approach, we replace difference penalties by their stochastic analogues, i.e. first or second order random walk models for the regression coefficients

$$b_t = b_{t-1} + u_t, \qquad b_t = 2b_{t-1} - b_{t-2} + u_t,$$

with Gaussian errors $u_t \sim N(0,t^2)$ and diffuse priors b_1 μ $const$, or b_1 and b_2 μ $const$, for initial values, respectively. A first order random walk penalizes abrupt jumps $b_t - b_{t-1}$ between successive states and a second order random walk penalizes deviations from the linear trend $2b_{t-1}- b_{t-2}$. Random walk priors may be equivalently defined in a more symmetric form by specifying the conditional distributions of parameters b_t given its left *and* right neighbors, e.g. b_{t-1} and b_{t+1} in the case of a first order random walk. Then, random walk priors may be interpreted in terms of locally polynomial fits. A first order random walk corresponds to a locally linear and a second order random walk to a locally quadratic fit to the nearest neighbors, see e.g. Besag et al. (1995). The amount of smoothness is controlled by the additional variance parameter t^2, which corresponds to the smoothing parameter in a frequentist approach. The larger (smaller) the variance, the rougher (smoother) are the estimated functions.

Let us now turn our attention to the spatial effects f_{str} and f_{unstr}. For the spatially correlated effect $f_{str}(s)$, $s=1,...,S$, we choose Markov random field priors common in spatial statistics (Besag, et al. 1991). These priors reflect spatial neighborhood relationships. For geographical data one usually assumes that two sites or regions s and r are neighbors if they share a common boundary. Then a spatial extension of random walk models leads to the conditional, spatially autoregressive specification

$$f_{str}(s) \mid f_{str}(r), r \neq s \sim N(\sum_{r \in \partial_s} f_{str}(r) / N_s, t^2 / N_s)$$

where N_s is the number of adjacent regions, and r Î ¶$_s$ denotes that region r is a neighbor of region s. Thus the (conditional) mean of $f_{str}(s)$ is an average of function evaluations $f_{str}(s)$ of neighboring regions. Again the variance t^2_{str} controls the degree of smoothness.

For a spatially uncorrelatated (unstructured) effect f_{unstr} a common assumption is that the parameters $f_{unstr}(s)$ are i.i.d. Gaussian

$$f_{unstr}(s) \mid t^2_{unstr} \sim N(0, t^2_{unstr}).$$

For a fully Bayesian analysis, variance or smoothness parameters t^2_j, $j=1,...,p, str, unstr$, are also considered as unknown and estimated simultaneously with corresponding unknown functions f_j. Therefore, hyperpriors are assigned to them in a second stage of the hierarchy by highly dispersed inverse gamma distributions $p(t^2_j) \sim IG(a_j,b_j)$ with known hyperparameters a_j and b_j.

2.3 Posterior inference

Bayesian inference is based on the posterior and is carried out using recent MCMC simulation techniques. Let a denote the vector of all unknown parameters in the model. Then, under usual conditional independence assumptions, the posterior is given by

$$P(a \mid y) \propto \prod_{i=1}^{n} L_i(y_i, h_i) \prod_{j=1}^{p} \{p(b_j \mid t_j^2) p(t_j^2)\} p(f_{str} \mid t^2_{str}) p(f_{unstr} \mid t^2_{unstr}) \prod_{j=1}^{r} p(g_j) p(s^2),$$

where b_j, $j=1,...,p$, are the vectors of regression coefficients corresponding to the functions f_j. The full conditionals for the parameter vectors $b_1,...,b_p$ as well as the full conditionals for f_{str}, f_{unstr} and fixed effects parameters g are multivariate Gaussian. For the variance components t^2_j, $j=1,...,p, str, unstr$, and s^2 the full conditionals are inverse gamma distributions. Thus, a Gibbs sampler can be used for MCMC simulation, drawing successively from the full conditionals for $b_1,...,b_p$, f_{str}, f_{unstr}, t^2_j, $j=1,...,p, str, unstr$, and s^2. Efficient sampling from the Gaussian full conditionals of nonlinear functions is guaranteed by Cholesky decompositions for band matrices. More details can be found in Rue (2001), Fahrmeir and Lang (2001b) and Lang and Brezger (2001).

3. Data and Results

The Demographic Health Surveys (DHS) of Tanzania and Zambia, both conducted in 1992, are used in this study. These surveys are produced jointly by Macro International, a USAID-funded firm specializing in demographic research, and the national statistical agency of the respective country. They draw a representative sample of women of reproductive age and then administer a questionnaire and an anthropometric assessment of themselves and their children that were born within the previous five years. The data set contains information on family planning, maternal and child health, child survival, HIV-AIDS, educational attainment, and household composition and characteristics. There are 8138 cases for Tanzania and 6299 for Zambia. The sample is drawn through stratified clustered sampling and draws, in the case of Zambia, 262 clusters from 53 districts in Zambia. In the case of Tanzania, we have data from 357 clusters drawn from 25 regions (which, to make them compatible with Zambia, we refer to as districts for our analysis). These districts can be grouped into nine provinces in Zambia and six provinces in Tanzania.

One cannot assume that the clusters selected in each district are fully representative of the districts in which they are located, as the surveys only attempted to generate a fully representative sample at the provincial level. Consequently, the spatial analysis will be affected by some random fluctuations. Some of this random variation can be reduced through the structured spatial effects as it includes neighboring observations in the analysis. It should, however, be pointed out that such a spatial analysis should preferably be applied to census data, the most important official demographic data source in most developing countries, where the precision of the spatial analysis would be much higher. Unfortunately, most censuses do not collect data on undernutrition and often the full dataset is not available for such analyses.

We concentrate in the analysis on the flexible modeling of the effects of the child's age, the mother's *BMI* and the districts on chronic undernutrition (stunting), measured using the Z-score as described above. To avoid numerical difficulties we standardized the Z-score before estimation. In addition, we consider several categorical variables including the sex of the child, the education and employment situation of the mother, access to water (later omitted as it was found to have a negligible influence) and locality (urban and rural). All categorical variables are included in effect-coding (rather than as usual dummy variables) and in the tables we also report on the reference category. The education variable is coded in three categories called, respectively, 'no education and incomplete primary education' (reference category), 'complete primary education and incomplete secondary education', and 'complete secondary education and higher'. For the employment situation of the mother, we distinguish between working and not working. We estimate separate models for each country with predictor

$$h = g_0 + f_1(AGC) + f_2(BMI) + f_{spat}(s) + g'w$$

where w includes the categorical covariates in effect coding. The functions f_1 and f_2 are modeled by cubic P-splines with second order random walk penalty. For the spatial effect f_{spat} we experimented with different prior assumptions. For both countries we estimated models where either a structured or an unstructured effect was included as well as a model where both effects were included. Based on these results we found clear evidence for both countries of spatial correlation among neighboring districts. Hence, a spatially correlated effect f_{str} is included into the predictors of our final models. For Zambia, we additionally include an unstructured effect f_{unstr} because there is evidence of local extra variation in the highly urbanized areas of Zambia. For Tanzania an unstructured effect is excluded from the final model. All computations have been carried out with *BayesX*, a software package for Bayesian inference based on MCMC simulation techniques, see Lang and Brezger (2000).

Table 1 shows the results of the fixed effects parameters in Tanzania. Despite modeling the spatial effects differently here, the results for the (non-spatial) fixed effects are virtually identical to Kandala, Lang, and Klasen (2001).

The substantive findings are generally as expected. Children of highly educated mothers living in urban areas are better nourished than other children. Children of working mothers do slightly worse. Being female is also associated with reduced levels of stunting, a finding consistent with Svedberg (1996) and Klasen (1996).

The results are quite similar for Zambia (Table 2). The direction of influences are the same in both countries. The size of the coefficients differ slightly. In particular, both the effect of education and of residence (urban versus rural) is somewhat smaller in Zambia. Moreover, the 80% credible region for the mother's employment status now includes zero. Access to water was found to be insignificant in both countries and was therefore omitted.

Figures 1 through 4 show the nonlinear effects of child's age and the mother's *BMI*. Also here, the differences to Kandala, Lang, and Klasen (2001) which was based on a different prior are very minor. Moreover, the results are not sensitive to the additional inclusion of nonlinear regional effects, suggesting that the method applied is able to separately identify nonlinear covariate and regional effects. Figure 1 shows the effect of the *BMI* of the mother in Tanzania. Shown are the posterior means together with 80% pointwise credible intervals. As hypothesized, we find the influence to be in the form of an inverse U shape. While the inverse U looks nearly symmetric, the descending portion exhibits a much larger range in the credible region. This appears quite reasonable as obesity of the mother (possibly due to a poor quality diet) is likely to pose less of a risk for the nutritional status of the child as very low BMIs which suggest acute undernutrition of the mother. The Z-score is highest (and thus stunting lowest) at a *BMI* of around 30-35.

Figure 2 shows the effect of the child's age on its nutritional status in Tanzania. As suggested by the nutritional literature, we are able to discern the continuous worsening of the nutritional status up until about 20 months of age. This deterioration sets in right after birth and continues, more or less linearly, until 20 months. Such an immediate deterioration in nutritional status is not as expected as the literature typically suggests that the worsening is associated with weaning at around 4-6 months. One reason for this unexpected finding could be that, according to the surveys, most parents give their children liquids other than breastmilk already shortly after birth which might contribute to infections at these early ages.

After 20 months, stunting stabilizes at a low level. Through reduced growth and the waning impact of infections, children are apparently able to reach a low-level equilibrium that allows their nutritional status to stabilize.

We also see a sudden improvement of the Z-score around 24 months of age. This is picking up the effect of a change in the data set that makes up the reference standard. Until 24 months, the currently used international reference standard is based on white children in the US of high socioeconomic status, while after 24 months, it is based on a representative sample of all US children (WHO, 1995). Since the latter sample exhibits worse nutritional status, comparing the Tanzanian children to that sample leads to a sudden improvement of their nutritional status at 24 months. This anomaly of the reference standard is one reason for WHO's current efforts to construct a new reference standard (WHO, 1999).

Figure 3 shows the effect of mother's BMI on chronic undernutrition in Zambia. Also here we find a, somewhat less pronounced, inverse U-shape. The inverse U-shape is much more pronounced on the ascending left portion than on the descending right portion, which is only barely discernible. Again, this is consistent with the notion that acute undernutrition of the mother is more of a risk for the child than obesity. Figure 4 shows the impact of the child's age on stunting in Zambia. Here the deterioration in the nutritional status appears to be slightly longer. It only stabilizes at around 22-24 months. Since this stabilization coincides with the change in data set in the reference standard, it is not as easy to separate the two phenomena as it was in Tanzania so that the blip in the Z-score is hardly visible.

To explore district-specific spatial effects, Figures 5-11 explore the spatial effects of undernutrition in the two countries. As mentioned above, in Tanzania we report on the model that only includes structured effects, while in the case of Zambia we report on the model that includes both structured and unstructured effects. Figure 5 shows the structured random effects for Tanzania and Figure 6 indicates the significance of the observed spatial effects in the form of a posterior probability map. The levels correspond to significantly negative (black colored), significantly positive (white colored) and insignificant (grey colored). Two important observations emerge. First, there is a strong South-North gradient in these regional effects with a fairly sharp dividing line running through the center of the country. Over and above the impact of the fixed effects, there appear to be negative influences on undernutrition in the South that are quite general and affect most of the regions there. Given that the Southern districts all are at significantly lower elevation than the rest of the country, it is likely that climatic and associated disease factors are responsible for this pronounced regional pattern. Second, living in the capital Dar es Salaam is associated with significantly better nutrition despite being surrounded by areas with negative regional effects on undernutrition. Living in the capital must thus provide access to nutrition and health care that is superior in ways that have not been captured adequately in the fixed effects.

To compare our district-specific nonlinear effects with our simple fixed effects for provinces which we used in Kandala, Lang, and Klasen (2001), Figure 7 presents a map that shows those provincial effects for the six provinces. Note that one can only distinguish five provinces as the effects for the Central Province and the Coastal Province are virtually identical. These crude provincial fixed effects miss most of the findings we discussed above. In particular, the sharp South-North divide present in the district analysis is now no longer visible as the Central and Coastal provinces include districts on both sides of that divide. Moreover, the positive effect of Dar es Salaam is simply averaged in with the Coastal province. Clearly, a lot is lost when relying on this crude strategy of modeling spatial effects.

Figure 8-10 shows the structured and the unstructured random effects for Zambia. The structured effects show a sizeable difference between significantly worse undernutrition in the Northern parts of the country (in particular the districts in Luapula and Northern province), and significantly better nutrition in the Central and South-Western parts. These regional patterns are similar, but not identical to analyses of poverty and deprivation undertaken by the World Bank (World Bank, 1995). In terms of income poverty, the World Bank found poverty to be lowest in the Central parts of the country. In addition, poverty was also much lower along the main trunk road and railroad lines even outside the central part of the country. In terms of deprivation (based on a mean score of various service items), the World Bank also found Luapula province among the worst off, while it surprisingly included the Central province and the Northwestern Province among the worst-off regions. While we also find Luapula province to be among the worst off in the country, our analysis shows a clearer geographic pattern with the North-East being worst off and the Central and South-Western districts being best off.

The unstructured random effects are mostly not significant. But they nevertheless point in interesting directions. In particular, they suggest a fair amount of variation over and above the structured effects. Particularly noteworthy is the fact that for some urban centers, the unstructured effects point to higher undernutrition, once the fixed effects (which include a positive effect of urban areas) and the structured effects are controlled for. This is particularly noteworthy for Kitwe in the Copperbelt, but also visible for Lusaka and Kabwe in the Central Part of the country. In contrast to Tanzania, it thus appears that some urban agglomerations are associated with worse nutrition. This may be related to the impact of economic decline and adjustment policies which have hit the Copperbelt and some other urban areas particularly hard (World Bank, 1995).

Figure 11 shows the provincial fixed effects used in Kandala, Lang, and Klasen (2001). While the overall spatial structure is more or less accurately reproduced, the effects of urban agglomerations on the structured and unstructured effects distorts the picture particularly for the Copperbelt and the Central Province where most of these urban agglomerations are.

In sum, the flexible modeling of the district-specific effects paints a much more nuanced picture than was presented by the regional fixed effects and thus give a better impression of the spatial variation of undernutrition. Moveover, the semi-parametric Bayesian approach used is able to identify subtle influences of the mother's *BMI*, the child's age on the nutritional status of the child.

These findings are not only relevant for analytical purposes but have considerable policy significance. In particular, the age effect points to considerable nutritional problems immediately after birth, possibly related to the use of unclean liquids. This is a subject that should be investigated further. Second, the nonlinear influence of the *BMI* indicates that not only parental undernutrition, but also parental malnutrition might also have negative effects on the nutritional status of children. Third, the regional influences on undernutrition also are of high policy significance. In particular, they suggest that in Tanzania inhabitants of the capital are much less affected by undernutrition, even if they suffer similar risk factors (as captured by the fixed effects). The same is, however, not true in Zambia, where some urban agglomerations are associated with higher undernutrition. Also, more emphasis must be placed upon the role of remoteness as well as climatic and geographic factors on undernutrition. The South-North divide in Tanzania and the regional effects in Zambia bear out the importance of such considerations.

4. Conclusion

In this paper, we have applied a semi-parametric Bayesian approach to model the determinants of chronic undernutrition (stunting) in Tanzania and Zambia. The fixed effects show the importance of mother's education, employment status, residence, and the sex of the child on chronic undernutrition. We also find that our methods are identifying subtle effects of the mother's *BMI*, the child's age, and regional influences on undernutrition. In particular, the effects of the *BMI* on the child's nutritional status appears to be in the form of an inverse U. Moreover, stunting appears to worsen until about 20-25 months and then stabilize at a low level equilibrium. Furthermore, we find sizeable regional effects. In both countries, we are able to pick up a distinct regional pattern of undernutrition that is not adequately captured by relying on provincial fixed effects.

Given the limitations of spatial analysis when the data base is a household survey, an important message emerging from this research is that it would be very worthwhile for census data and other official data sources to undertake such detailed spatial analyses. With such data sources, much more detailed and more precise spatial structures could be uncovered which would be highly relevant for both analytical as well as policy purposes.

Acknowledgement:
This research was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386 "Statistische Analyse diskreter Strukturen". We thank Macro International for providing the data and the maps of Tanzania and Zambia.

REFERENCES

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.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian Computation and Stochastic Systems (with discussion). Statist. Sci., 10, 3-66.

Biller, C. (2000). Adaptive Bayesian Regression Splines in Semiparametric Generalized Linear Models. J. Comp. Stat. and Graph. Stat., 9, 122-140.

De Boor, C. (1978). A Practical Guide to Splines. Springer-Verlag, New York.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998). Automatic Bayesian curve fitting. J. R. Statist. Soc.B, 60, 333-350.

Eilers, P. and Marx, B. (1996). Flexible Smoothing with P-splines and Penalized Likelihood (with comments and rejoinder). Statist. Sci. 11, 89-121.

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

Fahrmeir, L. and Lang, S. (2001b). Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. Ann. Inst. Statist. Math., 53, 11-30.

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

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient Models. J. R. Statist. Soc.B, 55, 757-796.

Hastie, T. and Tibshirani, R. (2000). Bayesian Backfitting. To appear in Statist. Sci..

Kandala, N.B., Lang, S. and Klasen, S. (2001). Semiparametric Analysis of Childhood Undernutrition in Developing Countries. To appear in ISBA 2000 Proceedings.

Klasen, S. (1996). Nutrition, Health, and Mortality in Sub-Saharan Africa: Is there a Gender Bias? Journal of Development Studies 32: 913-932.

Klasen, S. (1999). Malnourished and low mortality in South Asia, better nourished and dying young in Africa: What can explain this puzzle? SFB 386 Discussion Paper No. 214. University of Munich.

Lang, S. and Brezger, A. (2001). Bayesian P-Splines. SFB 386 Discussion Paper No. 236. University of Munich.

Lang, S. and Brezger, A. (2000). BayesX. Software for Bayesian Inference Based on Markov Chain Monte Carlo Simulation Techniques.
Available at http://www.stat.uni-muenchen.de/~lang/.

Moradi, A. and S. Klasen (2000). The Nutritional Status of Elites in India, Kenya, and Zambia: An Approprate Guide for Developing International Reference Standards for Undernutrition? SFB 386 Discussion Paper No. 217. University of Munich.

Nyovani, J.M., Z. Matthews, and B. Margetts (1999). Hetrogeneity of Child Nutritional Status between Households: A Comparison of six Sub-Saharan African Countries. Population Studies 53: 331-343.

Rue, H. (2001). Fast Sampling of Gaussian Markov Random Fields with Applications. To appear in J. R. Statist. Soc. B.

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

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

Sommerfelt, A.E. and Stewart, M.K. (1994). Children's Nutritional Status. Demographic and Health Surveys Comparative Studies No. 12, Macro International.

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

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 1998: Focus on Nutrition. New York: UNICEF.

WHO (1983). Measuring Change in Nutritional Status. Geneva: WHO.

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

WHO (1999). Infant and Young Child Growth: The WHO Multicentre Growth Reference Study. Executive Board: Implementation of Resolutions and Decisions EB105/Inf.doc/1. Geneva: WHO.

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

Table 1: Fixed effects for Tanzania (effect coding)

Variable	mean	10% quantile	90% quantile
Constant	0.29	0.17	0.41
Working	-0.02	-0.04	0
Not working	0.02	0	0.04
No edu. and incompl. prim. edu.	-0.26	-0.35	-0.17
Complete primary edu. and incomplete sec. Edu.	-0.18	-0.26	-0.09
Secondary edu. and higher	0.43	0.26	0.60
Urban	0.1	0.07	0.12
Rural	-0.1	-0.12	-0.07
Male	-0.04	-0.05	-0.02
Female	0.04	0.02	0.05

Table 2: Fixed effects for Zambia (effect coding)

Variable	mean	10% quantile	90% quantile
Constant	0.1	0.04	0.16
Working	0.01	-0.01	0.02
Not working	-0.01	- 0.02	0.01
No edu. and incompl. prim. edu.	-0.17	-0.21	-0.14
Complete primary edu. and incomplete sec. Edu.	-0.06	-0.09	-0.03
Secondary edu. and higher	0.24	0.18	0.30
Urban	0.09	0.06	0.12
Rural	-0.09	-0.12	-0.06
Male	-0.06	-0.07	-0.04
Female	0.06	0.04	0.07

Figure 1: Nonlinear effect of the mother's body mass index for Tanzania

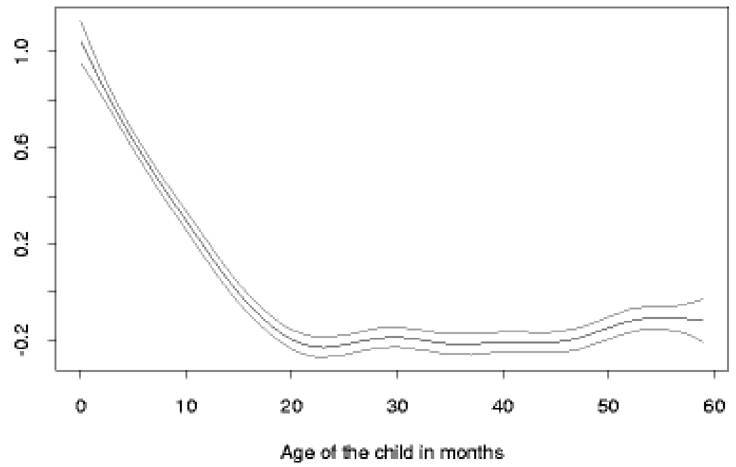

Figure 2: Nonlinear effect of child's age for Tanzania

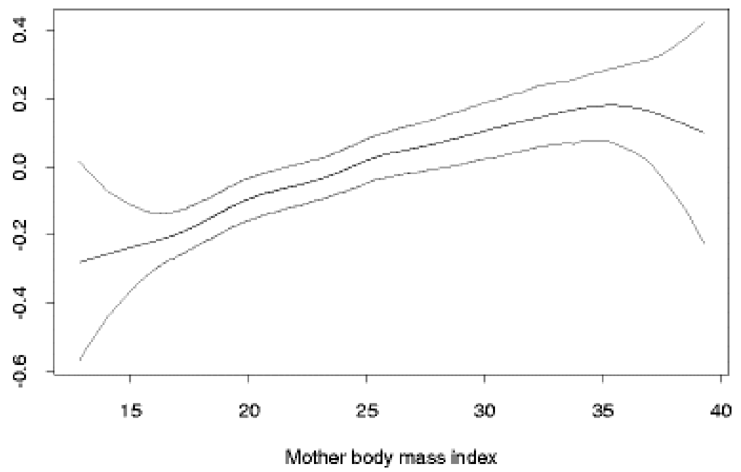

Figure 3: Nonlinear effect of the mother's body mass index for Zambia

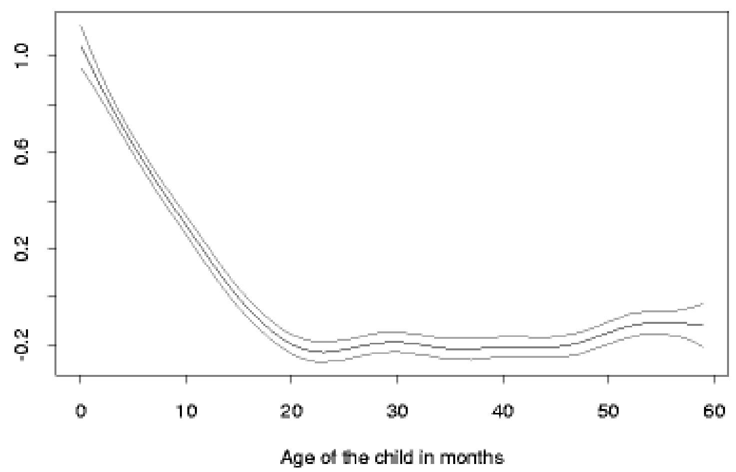

Figure 4: Nonlinear effect of the child's age for Zambia

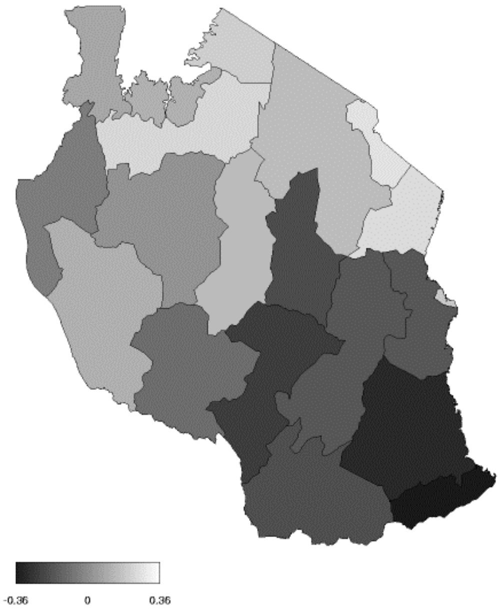

Figure 5: Posterior mean of the structured spatial effect for Tanzania

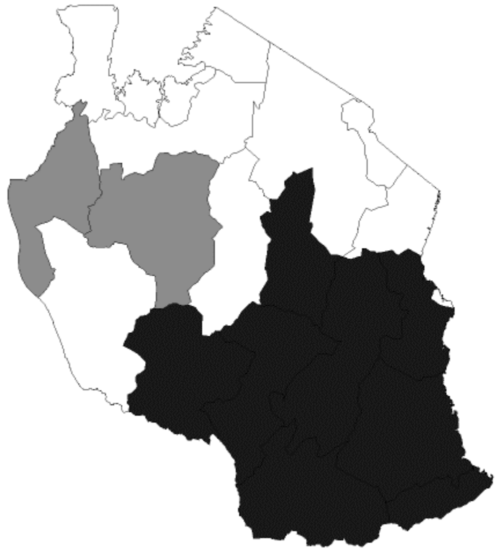

Figure 6: Posterior probabilities of the structured spatial effect for Tanzania

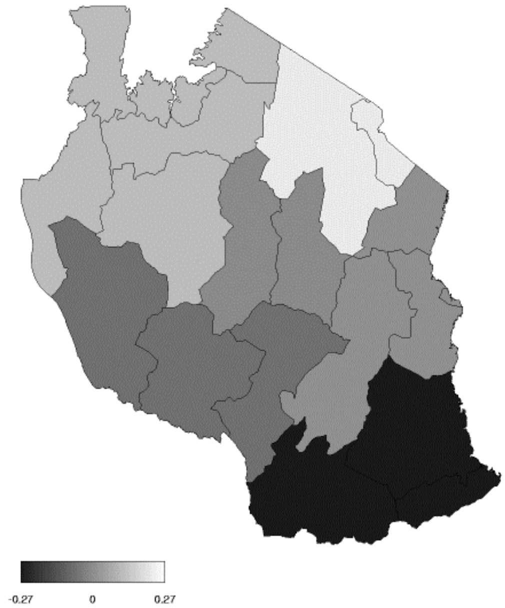

Figure 7:Regional effects for Tanzania

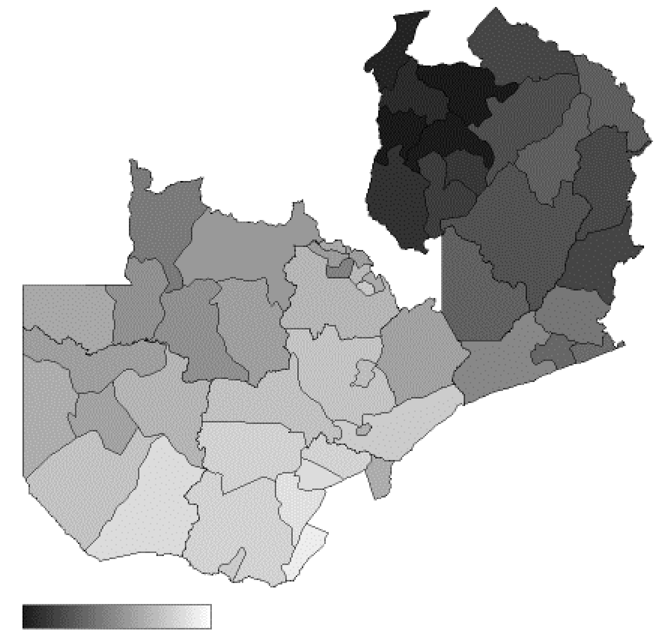

Figure 8: Posterior mean of the structured spatial effect for Zambia

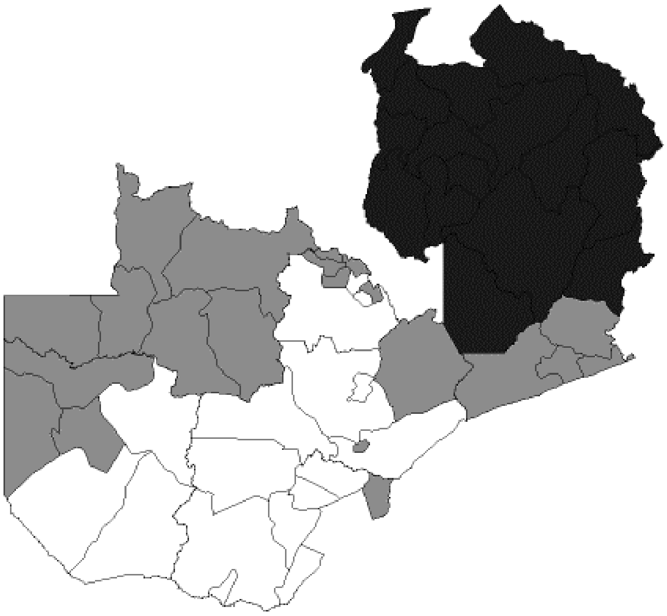

Figure 9: Posterior probabilities of the structured spatial effect for Zambia

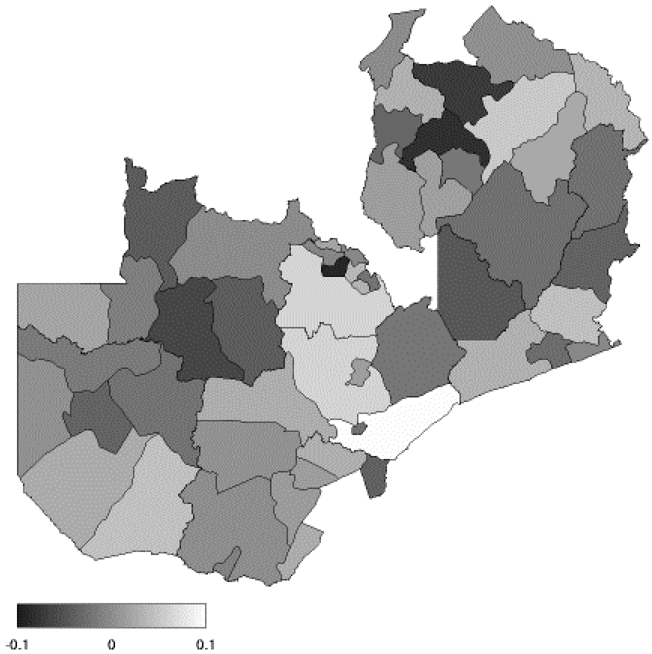

Figure 10: Posterior mean of the unstructured spatial effect for Zambia

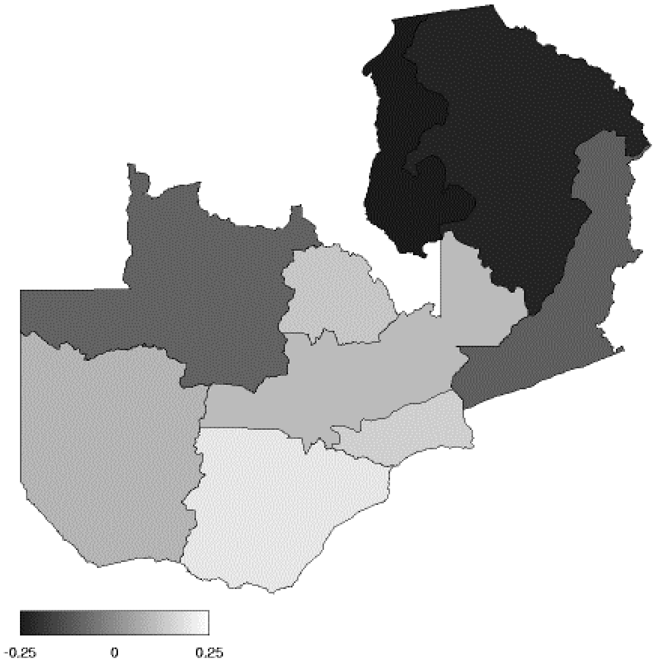

Figure 11: Regional effects for Zambia