Semiparametric Analysis of Childhood Undernutrition in Developing Countries

N.B Kandala, S. Lang, S. Klasen , Ludwigstr. 33 D-80539, University of Munich.

July 6, 2001

Resume

La malnutrition pendant l'enfance est le resultat des effets synergiques repetees et mal traitees ainsi que d'un apport alimentaire inadequat, causes frequente d'une mortalite accrue(Pelletier *et al.* , 1993. Nous portons notre attention sur la modelistaion flexible des determinants de la malnutrition infantine avec de la methode Bayesienne de MCMC. L'etude est basee sur les donnees des Enquetes demographiques et sanitaire de 1992 de la Tanzanie et la Zambie respectivement.

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 and the mother's body mass index on chronic undernutrition. Conventional parametric regression models are not flexible enough to cope with possibly nonlinear effects of the two continuous covariates age of the child and mother's body mass index. We present a Bayesian semiparametric analysis of the effects of covariates on chronic undernutrition that is suitable for exploring unknown nonlinear effects of metrical covariates. In a second step we investigate possible interactions between categorical covariates and child's age and mother's body mass index using varying coefficients models. Inference is fully Bayesian and uses recent Markov chain Monte Carlo techniques.

Keywords Developing Countries; semiparametric Bayesian inference; undernutrition; varying coefficients models

1 Introduction

Acute and chronic undernutrition is considered to be one of the worst health problems in developing countries. Poor nutrition 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 measured by determining the anthropometric status of the child. 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:

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

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. 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 access to 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 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) 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 worsens 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).

In this paper, we model the determinants of stunting in Zambia and Tanzania. Stunting is very prevalent in these two 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, overall, 43 percent of Tanzanian children under five are classified as stunted and 18 percent are severely stunted (Somerfelt and Stewart, 1994).

A particular focus is to use a flexible approach to model the impact of age and the BMI of the mother on undernutrition with the help of a semiparametric Bayesian modelling approach developed by Fahrmeir and Lang (2000a, b). Inference is fully Bayesian and uses MCMC techniques.

2 Semiparametric Bayesian regression models

2.1 Observation model

Consider regression situations, where observations $(y_i, x_i, w_i),\ i = 1, \ldots, n$, on a metrical response y, a vector $x = (x_1, \ldots, x_p)$ of metrical covariates and a vector $w = (w_1, \ldots, w_r)$ of categorical covariates are given. We assume that y_i given the covariates and unknown parameters are independent and Gaussian with mean η_i and a common variance σ^2 across subjects, i.e. $y_i \sim N(\eta_i, \sigma^2)$. In our application on childhood undernutrition the response is stunting measured as a Z-score (multiplied by 100). Traditionally, the effect of the covariates on the response is modelled by a linear predictor

$$\eta_i = x_i'\beta + w_i'\gamma. \tag{1}$$

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. Thus, we replace the strictly linear predictor (**??**) by the more flexible semiparametric predictor

$$\eta_i = f_1(x_{i1}) + \cdots + f_p(x_{ip}) + w_i'\gamma. \tag{2}$$

Here, $f_1, \ldots, f_p$ are nonlinear smooth effects of the metrical covariates.

Models with the predictor (**??**) form the basis of our analysis. In a second step we investigate possible interactions between the metrical covariates AGC and BMI and the categorical covariates (e.g. education). A convenient way of modelling these kinds of interactions is through varying coefficients models introduced by Hastie and Tibshirani (1993). Here, an interaction between a metrical covariate x and a categorical covariate w is modelled by a predictor of the form

$$\eta_i = \cdots + f(x_i)w_i + \cdots, \tag{3}$$

where f is again a smooth function of x. Thus, the effect of w is no longer fixed but varies smoothly over the range of x. Covariate x is called the effect modifier of w.

2.2 Prior assumptions

In a Bayesian approach unknown functions f_j and parameters γ as well as the variance parameter σ^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 $\gamma_j \propto const$, $j = 1, \ldots, r$ 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_j, $j = 1, \ldots, p$. 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. Dennison 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, 2000a), Bayesian smoothing splines (Hastie and Tibshirani, 2000) and Bayesian P-splines (Lang and Brezger, 2000a). In this paper we focus on random walk priors. We also compared our results with Bayesian smoothing splines and P-splines but the estimated functions were more or less undistinguishable.

For the random walk prior, let us first consider the case of a metrical covariate x with *equally spaced observations* x_i, $i = 1, \ldots, m$, $m \leq n$. Suppose that $x_{(1)} < \ldots < x_{(t)} < \ldots < x_{(m)}$ is the ordered sequence of distinct covariate values. Define $f(t) := f(x_{(t)})$ and let $f = (f(1), \ldots, f(t), \ldots, f(m))'$ denote the vector of function evaluations. Then a first or second order random walk prior for f is defined by

$$f(t) = f(t-1) + u(t) \quad \text{or} \quad f(t) = 2f(t-1) - f(t-2) + u(t) \qquad (4)$$

with Gaussian errors $u(t) \sim N(0; \tau^2)$ and diffuse priors $f(1) \propto$ const, or $f(1)$ and $f(2) \propto$ const, for initial values, respectively. A first order random walk penalizes abrupt jumps $f(t) - f(t-1)$ between successive states and a second order random walk penalizes deviations from the linear trend $2f(t-1) - f(t-2)$. Random walk priors may be equivalently defined in a more symmetric form by specifying the conditional distributions of function evaluations $f(t)$ given its left *and* right neighbours, e.g. $f(t-1)$ and $f(t+1)$ in the case of a first order random walk. Thus, 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 neighbours. Of course, higher order autoregressions are possible but practical experience shows that the differences in results are negligible. For the case of *nonequally spaced observations* random walk priors must be modified to account for nonequal distances $\delta_t = x_{(t)} - x_{(t-1)}$ between observations. Random walks of first order are now specified by

$$f(t) = f(t-1) + u(t), \quad u(t) \sim N(0; \delta_t \tau^2), \tag{5}$$

i. e., by adjusting error variances from τ^2 to $\delta_t \tau^2$. Random walks of second order are defined by

$$f(t) = \left(1 + \frac{\delta_t}{\delta_{t-1}}\right) f(t-1) - \frac{\delta_t}{\delta_{t-1}} f(t-2) + u(t), \tag{6}$$

$u(t) \sim N(0; w_t \tau^2)$, where w_t is an appropriate weight. Several possibilities are conceivable for the weights, see Fahrmeir and Lang (2000a) for a discussion.

The trade off between flexibility and smoothness of f is controlled by the variance parameter τ^2. In our approach we want to estimate the variance parameter and the smooth function simultaneously. This is achieved by introducing an additional hyperprior for τ^2 in a further stage of the hierarchy. We choose a highly dispersed but proper inverse gamma prior $p(\tau^2) \sim IG(a; b)$ with $a = 1$ and $b = 0.005$. In analogy, we also define for the overall variance σ^2 a highly dispersed inverse gamma prior.

2.3 Posterior inference

Bayesian inference is based on the posterior and is carried out using recent MCMC simulation techniques. Without loss of generality, we restrict the presentation to models with predictor (**??**). Let $f = (f_1, \ldots, f_p)$ and $\tau^2 = (\tau_1^2, \ldots, \tau_p^2)$ denote parameter vectors for function evaluations and variances. Then, under usual conditional indepence assumptions, the posterior is given by

$$p(f, \tau^2, \gamma | y) \propto \prod_{i=1}^{n} L_i(y_i; \eta_i) \prod_{j=1}^{p} \{p(f_j | \tau_j^2) p(\tau_j^2)\} \prod_{k=1}^{r} p(\gamma_k) p(\sigma^2).$$

The full conditionals for unknown functions f_j, $j = 1, \ldots, p$, and fixed effects parameters γ are Gaussian and for variance components τ_j, $j = 1, \ldots, p$ and σ^2 the full conditionals are inverse gamma distributions. Thus, a simple gibbs sampler can be used for MCMC simulation, drawing successively from the full conditionals for f_j,τ_j^2, $j = 1, \ldots, p$ and σ^2. Efficient sampling from the Gaussian full conditionals of nonlinear functions is guaranteed by Cholesky decompositions for band matrices. More details can be found e.g. in Rue (2000) or Fahrmeir and Lang (2000b).

3 Data and Results

The Demographic Health Surveys (DHS) of Tanzania and Zambia, both conducted in 1992, are used in this study. These surveys 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 sets 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.

We concentrate in the analysis on the flexible modeling of the effects of the child's age and the mother's BMI on chronic undernutrition (stunting), measured using the Z-score (multiplied by 100) as described above. The response variable stunting described above has been further standardized in this analysis for computational purposes. 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), locality (urban and rural) and the province in which the household is located. 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. There are six provinces for Tanzania and nine provinces for Zambia as shown in Table 1. We estimate separate models for each countries with predictor

$$\eta = \gamma_0 + f_1(AGC) + f_2(BMI) + \gamma' w$$

where w includes the categorical covariates in effect coding. The functions f_1 and f_2 are modelled by second order random walk priors defined in (**??**) and (**??**). All computations have been carried out with *BayesX*, a software package for Bayesian inference basd on MCMC simulation techniques, see Lang and Brezger (2000b).

Table 1 shows the results of the fixed effects parameters in Tanzania. The results are generally as expected. Children of highly educated working mothers living in urban areas are better nourished than other children. Being female is also associated with reduced levels of stunting, a finding consistent with Svedberg (1996) and Klasen (1996). There are also sizeable regional fixed effects, which are highly significant (i.e. credible intervals are either strictly positive or strictly negative).

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 and in Zambia the the 80 % credible region for the mother's employment status includes zero. The regional effects are also large and significant. Access to water was found to be insignificant in both countries and was therefore omitted.

The left panel of Figure 1 shows the effect of the BMI of the mother in Tanzania. Shown are the posterior means together with 80 % pointwise credible intervals. For comparison a regression line (dashed) obtained by a linear fit is added to the plot. 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 months.

The right panel of Figure 1 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 set in right after birth and continues, more or less linearly, until 20 months. Such an immediate deterioration in nutritional status is not quite 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.

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 blip 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).

The left panel of Figure 2 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. The right panel of Figure 2 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 possible to distinguish between the two phenomena as it was in Tanzania.

We experimented with several interactions. First, we ran separate models for males and females (not reported here) but found them to be very similar. Second, we estimated a varying coefficients model where we interact the effect of mother's education with the age of the child. This interaction only had a significant effect in Zambia. Figure 3 shows that the effect of mother's education is negligible at birth and then rises, more or less linearly, to become quite pronounced for older children. Higher education which is an indicator also of greater parental resources and better care practices thus appears to matter more for older than for younger children. Since most infants are breastfed and greater parental discretion only enters at and after weaning, this finding suggests that this has an important impact on the effect of education on stunting.

Figure 4 shows the region fixed effects as shown in table 1 and table 2 in effect coding.

Quite clearly, the semi-parametric Bayesian approach used is able to identify subtle influences of the mother's BMI and the child's age on the nutritional status of the child. The linear fits included in the figures show the inadequacy of such an approach. We also experimented with cubic splines to model the impact of the child's age and the mother's BMI more flexibly, but found that this approach was particularly unreliable in regions where there are relatively few observations.

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. We find that our methods are identifying subtle relationships between the mother's BMI and the child's age 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. There also appears to be an interaction between mother's education and age, with mother's education having an increasing impact on stunting among older children. While some of these effects have been identified in univariate analysis, this study is able to show that these subtle influences remain in a multivariate context, controlling for a range of fixed effects and using a flexible approach to modelling these influences.

The fixed effects show the importance of mother's education, employment status, residence, the household size, and the sex of the child on chronic undernutrition. There are sizeable regional effects which need to be scrutinized in further work.

This semi-parametric approach thus appears to be able to discern subtle influences on undernutrition. It could also be of value for a flexible modeling of other determinants of undernutrition in developing countries, a subject to be addressed in future work.

References

[1] Biller, C. (2000).

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

[2] 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.

[3] Fahrmeir, L. and Lang, S. (2000a). *Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. To appear in Applied Statistics (JRSS C).*

[4] Fahrmeir, L. and Lang, S. (2000b). *Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. To appear in Ann. Inst. Statist. Math..*

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

[6] Hastie, T. and Tibshirani, R. (2000). *Bayesian Backfitting. To appear in Statistical Science.*

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

[8] 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.*

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

Appendix

Table 1: Region Fixed effects for tanzania

Variable	mean	10% quant.	90%quant.
constant	0.49	0.39	0.59
Urban	0.06	0.04	0.09
Rural	-0.06	-0.09	-0.04
Male	-0.04	-0.05	-0.02
Female	0.04	0.02	0.05
Not working	0.02	0	0.03
Working	-0.02	-0.03	0
No edu. and incompl. prim. edu.	-0.23	-0.32	-0.14
Compl.primary edu. and incompl. sec. edu	-0.16	-0.24	-0.07
Secondary edu. and higher	0.39	0.21	0.56
Coastal	0.01	-0.03	0.05
Northern Highlands	0.19	0.14	0.23
Lake	0.1	0.07	0.13
Central	0.01	-0.03	0.06
Southern Highlands	-0.07	-0.11	-0.03
South	-0.24	-0.28	-0.2

Table 2: Region Fixed effects for Zambia

Variable	mean	10% quant.	90%quant.
constant	-1.82	-9.57	3.3
Urban	0.01	-0.02	0.04
Rural	-0.01	-0.04	0.02
Male	-0.06	-0.08	-0.04
Female	0.06	0.04	0.08
Not working	-0.01	-0.02	0.01
Working	0.01	-0.01	0.02
No edu. and incompl. prim. edu.	-0.13	-0.17	-0.09
Compl.primary edu. and incompl. sec. edu	-0.04	-0.08	-0.01
Secondary edu. and higher	0.18	0.12	0.24
Central	0.09	0.03	0.14
Copperbelt	0.08	0.03	0.13
Eastern	-0.08	-0.14	-0.03
Luapula	-0.25	-0.3	-0.2
Lusaka	0.13	0.08	0.18
Northern	-0.22	-0.27	-0.17
North-Western	-0.1	-0.17	-0.04
Southern	0.22	0.17	0.26
Western	0.14	0.09	0.21

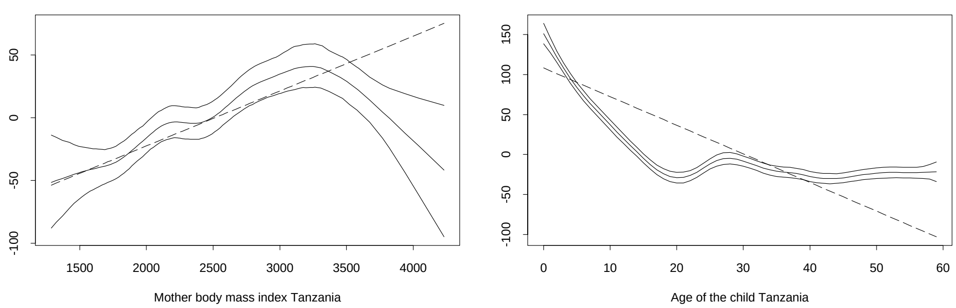

Figure 1. Nonlinear effects of body mass index and child's age for Tanzania.

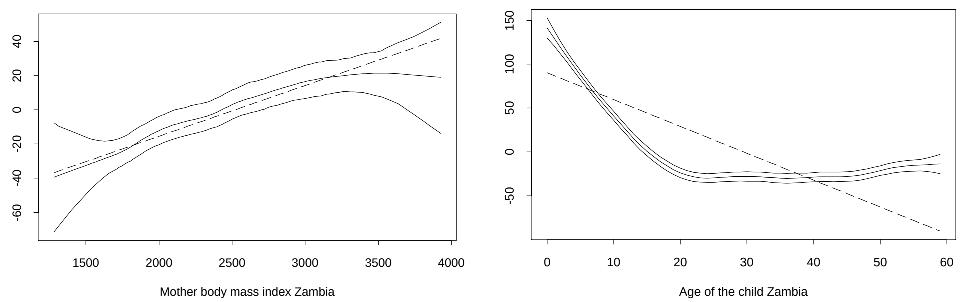

Figure 2. Nonlinear effects of body mass index and child's age for Zambia.

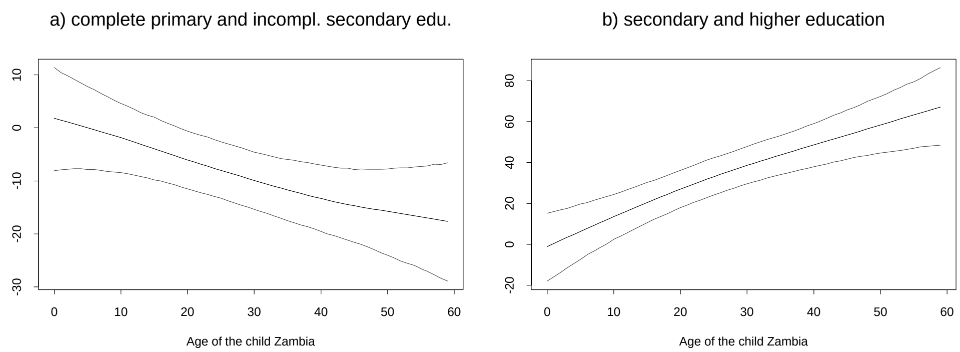

Figure 3. Varying effects of education with respect to child's age for Zambia.

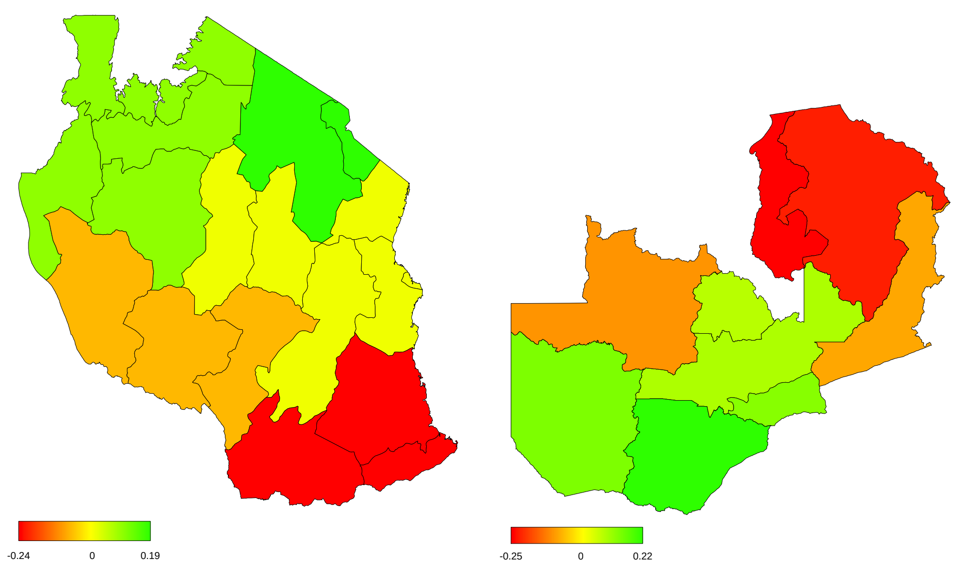

Figure 4. Region Fixed effects for Tanzania (left) and Zambia (right) (Effects coding).