Spatial and Spatio-temporal Epidemiology 13 (2015) 7–14

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Original Research

Spatial quantile regression using INLA with applications to childhood overweight in Malawi

Owen P.L. Mtambo [a,*], Salule J. Masangwi [b], Lawrence N.M. Kazembe [c]

[a] Mathematics and Statistics, Polytechnic of Namibia, Windhoek 264, Namibia
[b] Mathematics and Statistics, Polytechnic of Malawi, Blantyre 265, Malawi
[c] Mathematics and Statistics, University of Namibia, Windhoek 264, Namibia

ARTICLE INFO

Article history:
Received 3 November 2014
Revised 10 April 2015
Accepted 12 April 2015
Available online 18 April 2015

Keywords:
Fully Bayesian
Quantile regression
INLA
Intrinsic CAR
Structured spatial effects
Childhood overweight

ABSTRACT

Analyses of childhood overweight have mainly used mean regression. However, using quantile regression is more appropriate as it provides flexibility to analyse the determinants of overweight corresponding to quantiles of interest. The main objective of this study was to fit a Bayesian additive quantile regression model with structured spatial effects for childhood overweight in Malawi using the 2010 Malawi DHS data. Inference was fully Bayesian using R-INLA package. The significant determinants of childhood overweight ranged from socio-demographic factors such as type of residence to child and maternal factors such as child age and maternal BMI. We observed significant positive structured spatial effects on childhood overweight in some districts of Malawi. We recommended that the childhood malnutrition policy makers should consider timely interventions based on risk factors as identified in this paper including spatial targets of interventions.

1. Introduction

Childhood malnutrition has serious adverse effects on a child, a family and the development of a country. A malnourished child is more likely to be sick and die (Tomkins and Watson, 1989). In sub-Sahara Africa, it leads to more than 30% of deaths in children below 5 years (UNICEF, 2009). Malnutrition can lead to retarded growth (Martorell et al., 1994), impaired cognitive and behaviour development (UNICEF WFP, 2006), poor school performance, lower working capacity and lower income (UNICEF, 1998). It can slow down economic growth and increase level of poverty. Furthermore, it can prevent a society from meeting its full potential through loss in productivity, cognitive capacity and increased cost in health care (UNICEF WFP, 2006). The indicators of malnutrition range from stunting, wasting and underweight to overweight and obesity.

Childhood overweight and obesity rates are steadily increasing in sub-Sahara Africa. The prevalence of childhood overweight and obesity in Africa in 2010 was estimated at 8.5% and is expected to reach 12.7% by 2020 (Onis et al., 2010). In particular, childhood overweight in Malawi significantly increased from 4% in 2000 to 9.2% in 2012 (WHO, 2013). Assuming that the trend has remained the same, the childhood overweight prevalence in Malawi should be over 10% at present. The consequences of overweight can be more devastating than those for undernutrition because it leads to chronic failure problems which call for more medical expenditure (UNICEF, 1998; UNICEF WFP, 2006).

The reduction of childhood malnutrition is reflected in the United Nations Millennium Development Goal (MDG) number 1, aiming at halving the proportion of children suffering from hunger by 2015. In addition, it also has direct impact on the MDG number 4 which aims at reducing

* Corresponding author.
E-mail address: owen.mtambo@yahoo.com (O.P.L. Mtambo).

http://dx.doi.org/10.1016/j.sste.2015.04.001
1877-5845/

the under-five mortality rate by two-thirds by 2015. In order to attain both MDG1 and MDG4, a first 1000 child-days project was launched in Malawi in 2010. However, no evaluation has been made to statistically understand the socio-demographic determinants and spatial variation of childhood overweight. For this reason, only childhood overweight was analysed in this study in order to assess socio-demographic determinants and geographical variation of childhood overweight prevalence in Malawi.

Spatial models have previously been used to appropriately analyse the childhood malnutrition in most sub-Sahara African countries including Malawi (Kandala et al., 2001; Khatab and Fahrmeir, 2008). Unfortunately, these have emphasized on modelling mean regression instead of quantile regression. Modelling malnutrition using quantile regression is more appropriate than using mean regression with extensive literature examples in (Koenker and Portnoy, 1994; Koenker and Hallock, 1999; Kneib et al., 2009; Fenske et al., 2009; Kneib, 2013; Rigby et al., 2013; Koenker, 2013; Harvey, 2013; Green, 2013), in that it provides flexibility to analyse the determinants of malnutrition corresponding to quantiles of interest either in the lower tail (say 5% or 10%) or upper tail (say 90% or 95%) or even median (50%) of the distribution rather than only analysing the determinants of mean distribution. The fact is that when modelling malnutrition, it makes more sense to model severe responses rather mean responses (Koenker and Portnoy, 1994; Koenker and Hallock, 1999; Kneib et al., 2009; Fenske et al., 2009; Kneib, 2013; Rigby et al., 2013; Koenker, 2013; Harvey, 2013; Green, 2013). For instance, it is more sensible to model severe stunting or severe overweight/obesity than to model mean stunting or mean overweight/obesity which corresponds to the lower and upper tails of the distribution of the same appropriate anthropometric measure. In this study, childhood overweight was of primary interest and for this reason, the tau parameter was fixed at 0.85 ($\tau = 0.85$) which corresponded to a CBMIZ >2 (the cut-point for childhood overweight according to WHO standards) (WHO, 2013). If we were also interested in childhood obesity, we would simply fix the tau parameter at 0.92 ($\tau = 0.92$) which corresponded to a CBMIZ >3 (the cut-point for childhood obesity according to WHO standards) (WHO, 2013).

Moreover, spatial regression is most appropriate for modelling malnutrition in that it takes into account the spatially correlated (district-specific) effects onto malnutrition response variable. The main purpose of this study was to fit a modern spatial quantile model that would better explain variability in childhood overweight, at a relatively small area level, in Malawi.

The rest of this paper is structured as follows. Section 2 describes the methods used in this study. The results of this study are given in Section 3. Finally, the discussion and conclusion are presented in Sections 4 and 5, respectively.

2. Methods

This section summarises the conceptual framework of the Bayesian structured additive quantile regression models, the data sources, and data analysis procedures used in this study.

2.1. The Model

2.1.1. Quantile regression

In general, quantile regression is all about describing conditional quantiles of the response variable in terms of covariates instead of the mean. The general additive conditional quantile model is given by:

$$Q_{Y_i|x_i z_i}(\tau|x_i, z_i) = x_i^T \beta\tau + \sum_{j=1}^{q} g_{\tau j}(z_{ij}) \quad (1)$$

where $Q_{\boldsymbol{Y}_i|\boldsymbol{x}_i,\boldsymbol{z}_i}(\tau|\boldsymbol{x}_i, \boldsymbol{z}_i)$ is the conditional τth quantile response given $\boldsymbol{x_i}$ and $\boldsymbol{z_i}$, η_{τ_i} is the semi-parametric predictor, $\tau\epsilon(0,1)$ is the τth quantile of the response e.g. $\tau = 0.5$ for the median response regression, $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ is the vector of p categorical covariates (assumed to have fixed effects) for each individual i, $\boldsymbol{z_i} = (z_{i1}, z_{i2}, \ldots, z_{iq})^T$ is the vector of q metric/spatial covariates, $\boldsymbol{\beta_\tau} = (\beta_{\tau 1}, \beta_{\tau 2}, \ldots, \beta_{\tau p})^T$ is the vector of p coefficients for categorical covariates at a given τ, $\boldsymbol{g_\tau} = (g_{\tau 1}, g_{\tau 2}, \ldots, g_{\tau q})^T$ is the vector of q smoothing functions for metric/spatial covariates at a given τ (Koenker and Portnoy, 1994; Yu et al., 2003; Kneib, 2013).

It is worthy to note that quantile regression duplicates the roles of quartile, quintile, decile, and percentile regressions. This is achieved by selecting appropriate values of τ in the conditional quantile regression model where $\tau\epsilon(0,1)$.

The two unknowns, $\boldsymbol{\beta}_\tau$ and $\boldsymbol{g}_\tau$ are estimated via the minimisation rule given by:

$$\left(\min_{\boldsymbol{\beta}_\tau, \boldsymbol{g}_\tau}\right) \sum \rho_\tau(\eta_{\tau_i}) + \lambda_0 ||\boldsymbol{\beta}_\tau||_1 + \sum_{j=1}^{q} \lambda_j \vee (\nabla \boldsymbol{g}_{\tau_j}) \quad (2)$$

where ρ_τ is the check function (appropriate loss function) evaluated at a given τ, λ_0 is the zeroth (initial) tuning parameter for controlling the smoothness of the estimated function, λj is the jth tuning parameter for controlling the smoothness of the estimated function, $||\boldsymbol{\beta}_\tau||_1 = \sum_{k=1}^{k} |\boldsymbol{\beta}_{\tau_k}|$ and $\vee(\nabla g_{\tau_j})$ denotes the total variation of the derivative on the gradient of the function g_{τ_j} Koenker and Portnoy, 1994.

Bayesian inference requires likelihood. We need an assumption on data distribution for Bayesian quantile inference because the classical quantile regression has no such restriction. A possible parametric link between the minimisation problem and the maximum likelihood theory is the asymmetric Laplace density (ALD). This skewed distribution is defined in Koenker and Hallock (1999), Yu and Moyeed (2001), Kneib (2013).

2.1.2. Prior distributions

In fully Bayesian framework, all unknown functions $\{g\}'s$ for both metric and spatial covariates, all parameters $\{\beta\}'s$ for categorical covariates, and all variance parameters $\{\sigma^2\}'s$ are considered as random variables and have to be supplemented by appropriate prior distributions.