Spatial and Spatio-temporal Epidemiology 6 (2013) 71–84

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

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

An additive regression model for investigating the relationship between childhood health and socio-economic status

Lawrence N. Kazembe *

Statistics Department University of Namibia, P/Bag 13301, Windhoek, Namibia

ARTICLE INFO

Article history:
Received 27 July 2012
Revised 11 July 2013
Accepted 12 July 2013
Available online 25 July 2013

Keywords:
Childhood health
Socio-economic status
Structured additive regression models
Space-varying coefficient models
Bayesian inference

ABSTRACT

Health risks associated with socio-economic status (SES) have extensively been studied in epidemiology. It is not uncommon that data used exhibit spatial correlation, nonlinear effects, overdispersion and heterogeneity, and structured additive regression (STAR) models permit incorporating these features in a single analytical framework. Nevertheless, most STAR models assume constant spatial effects. However, due to social or disease transmission processes, covariates may be space-varying. We explore this feature by fitting a multinomial logistic model on a joint response variable constructed from four health indicators (i.e. a child having fever, diarrhoea, or being stunted and underweight) and examine its relationship with SES estimated as a space-varying coefficient (SVC) variable. Implementation of the model follow a Bayesian framework. Our comparison with models that assume constant spatial effects shows superiority of the SVC model as well as confirms the fact that SES varies in space.

1. Introduction

The study of socio-economic inequality has a wide range of interest in human sciences with considerable research emphasis in public health (Wagstaff, 2000; Braveman and Tarimo, 2002; Wagstaff, 2000; Black et al., 2003). It is now an established fact, in public health, that individuals at low socio-economic position have a higher disease burden than those at higher position. Education attainment, income levels, race and ethnicity among others all have exceedingly great impact on health (Braveman and Tarimo, 2002; Wagstaff, 2000; Black et al., 2003). Studies commissioned by World Health Organization (WHO) on socio-economic determinants of health underscore the importance of social disparities in health, and its impact on socio-economic development (Wagstaff, 2000; Zere and McIntyre, 2003).

The monitoring and reporting of socio-economic position has become a common place in characterizing health statistics. Indeed, common among developed countries for example in Americas and Europe, is that routine public health data are reported stratified by socio-economic variables, mainly by income, race and ethnicity, or rurality (Braveman and Tarimo, 2002). These have allowed comparison across social classes, facilitated monitoring of socio-economic disparities in health, and informed planning of interventions. In contrast, in developing countries, such studies are relatively few and new, and where these have been considered there are no explicit investigation of socio-economic patterning (Fotso and Kuate-Defo, 2005; Hong, 2007).

In this article, we aim at modelling the linkage of childhood health and socio-economic status (SES). A standard approach in epidemiology is to use regression models to study the relationship between a response and covariates, often using a linear model if the response is normally distributed or generalized linear models where the response variable is non-normal, for example categorical, skewed or count outcomes. Existence of nonlinear effects in continuous covariates have been handled using generalized additive models (Fahrmeir and Lang, 2001). In many of these

* Tel.: +264 61 206 4515; fax: +264 61206 3446.
E-mail address: lkazembe@unam.na

1877-5845/$ - see front matter
http://dx.doi.org/10.1016/j.sste.2013.07.002

applications, a strong assumption of independence of the responses is used, a fact which is flawed when data are clustered (Steele et al., 2004; Fahrmeir and Raach, 2007). A common feature in epidemiological data is spatial clustering (Kandala, 2006; Kandala et al., 2006), a phenomenon which is best handled by permitting spatially correlated random effects (Kandala et al., 2006). In addition, modelling becomes complicated when data exhibit nonlinearity, spatial autocorrelation, unobserved heterogeneity and overdispersion. Structured additive regression (STAR) models have been suggested for analysing such data features in a single regression framework (Kneib and Fahrmeir, 2006).

We use STAR models to jointly analyse the geographical distribution of the four leading causes (diarrhoea, fever, stunting and underweight) of child morbidity, and investigate their association with SES. One challenge is that the outcome-determinants relationship is not constant in any given geographical area, thus inducing spatially varying association. In this case, standard regression methodology will lead to biased estimates and inference. Appropriate analysis entails using spatial and space-varying coefficient models to examine the association of childhood health to socio-economic variables. In our analysis, we propose fitting SES as a spatially varying covariate, while controlling for other risk factors on the four causes. Space-varying effects control for spatial correlation and allow regression coefficients to vary in space, i.e. they are considered as smooth functions of geographical space (Fotheringham et al., 2002; Gamerman et al., 2003; Gelfand et al., 2006). In our literature search, comparatively few studies have mapped socio-economic inequalities in public health by adopting space-varying coefficient models (Congdon, 2003), and none of these in an African setting.

A multinomial model is applied to analyse spatial patterns of childhood comorbidity in Malawi, in what is called a multi-categorical response model (Kneib and Fahrmeir, 2006), where we consider a joint occurrence of: (i) diarrhoea and fever; (ii) diarrhoea and stunting; (iii) fever and stunting and (iv) stunting and underweight as a response category of the multinomial random variable. Note that such responses may be modelled in different ways. Multivariate models are such an alternative, see for example Fahrmeir and Raach (2007) who introduced a latent class model to analyse multiple indicators. However, multicategorical models such as the multinomial logistic model are widely used in the social sciences, public health and epidemiology, as either choice or classification models, for instance in demographic analysis of life-course events, assessment of disease stages or cause of death to name a few (Steele et al., 2004; Fahrmeir and Tutz, 2001).

The joint analysis approach considered here recognises the fact that several health outcomes occur simultaneously, largely because of common risk factors, and probably due to overlap between multiple risk factors, or that one disorder creates an increased risk for the other (Kazembe and Namangale, 2007; Fenn et al., 2005). In many sub-Saharan African countries, diarrhoea, malaria, and malnutrition cause and inflict the largest burden (Black et al., 2003; Fenn et al., 2005), and they are often common forms of comorbidities (Källander et al., 2004). Indeed, their co-existence is largely blamed for expediting early and high childhood mortality (Fenn et al., 2005; Källander et al., 2004; Black et al., 2003). Modelling and inference is implemented using full Bayesian inference (Fahrmeir and Lang, 2001). Alternative approaches to model fitting include approximate nested Laplace approximation and mixed methodology, and details are given elsewhere (Rue et al., 2009; Fahrmeir and Kneib, 2006).

Now the rest of this paper is structured as follows. Section 2 describes model development, while Section 3 gives details of model fitting. In Section 4, we apply the techniques to real data from 2006 Malawi Multiple Indicator Cluster Survey. Section 4 gives the results. The final section is the conclusion.

2. Additive regression model

2.1. The multinomial model

A multinomial random variable is applied where an event, Y, ends up with three or more outcomes $1, \ldots, J$ $(J > 2)$. Specifically suppose Y has unordered categories, we assume

$$Y \sim \text{multinomial}(1, p(v_i, \alpha)) \text{ for } i = 1, \ldots, n,$$

such that $p(v_i, \alpha) = (p_1(v_i, \alpha), \cdots, p_J(v_i, \alpha))'$, and $P(y_i = j|\alpha) = p_j(v_i, \alpha)$, for some covariates $v = (v_1, \ldots, v_p)'$ and corresponding parameter set α. The most common approach to estimate multinomial probabilities is through the logistic model

$$p(v_i, \alpha) = P(y_i = j|\alpha) = \begin{cases} \frac{\exp(\eta_{ij})}{1+\sum_{h=1}^{J-1}\exp(\eta_{ih})} & j = 1, \cdots, J-1 \\ \frac{1}{1+\sum_{h=1}^{J-1}\exp(\eta_{ih})} & j = J \end{cases} \tag{1}$$

where $\eta_{ij} = v\prime\alpha_j$ is the linear predictor. The last category J is considered as a reference classification outcome. The likelihood L takes the form

$$L = \prod_{i=1}^{n}\prod_{j=1}^{J}[p(v_i, \alpha)]^{y_{ij}}$$

with log-likelihood

$$\log L = \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij} \log[p(v_i, \alpha)].$$

In the classical multinomial logit model all covariates are assumed to be independent of the category while effects are category-specific. Extensions of the classic model allows for the inclusion of category-specific covariates w_{J-1} leading to the predictor $\eta_{ij} = v\prime\alpha_j + w_j\theta$.

When observations are associated with location of residence, it is desirable to account for spatial correlation and heterogeneity. Modelling of heterogeneity and spatially structured variation are obtained by introducing random effects. Similarly, nonlinear effects are introduced in the model through smoothing functions. The predictor (1) is expanded to include all possible explanatory variables like