Spatial Statistics 20 (2017) 30–56

Contents lists available at ScienceDirect

Spatial Statistics

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

Small area prediction of counts under a non-stationary spatial model

Hukum Chandra [a,*], Nicola Salvati [b], Ray Chambers [c]

[a] ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi, India
[b] Dipartimento di Economia e Management, University of Pisa, Italy
[c] National Institute for Applied Statistics Research Australia, University of Wollongong, Australia

ARTICLE INFO

Article history:
Received 26 August 2016
Accepted 25 January 2017
Available online 7 February 2017

Keywords:
Geographically weighted regression
Spatial nonstationarity
Count data
MSE estimation
Poverty mapping

ABSTRACT

There is a growing need for current and reliable counts at small area level. The empirical predictor under a generalised linear mixed model (GLMM) is often used for small area estimation (SAE) of such counts. However, the fixed effect parameters of a GLMM are spatially invariant and do not account for the presence of spatial nonstationarity in the population of interest. A geographically weighted regression extension of the GLMM is developed, extending this model to allow for spatial nonstationarity, and SAE based on this spatially nonstationary model (NSGLMM) is described. The empirical predictor for small area counts (NSEP) under an area level NSGLMM is proposed. Analytic and bootstrap approaches to estimating the mean squared error of the NSEP are also developed, and a parametric approach to testing for spatial nonstationarity is described. The approach is illustrated by applying it to a study of poverty mapping using socio-economic survey data from India.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Sample surveys are generally conducted to produce estimates for populations, sub-populations or larger domains (e.g. province/state level). Accordingly, sample sizes are fixed in such a way that

* Corresponding author. Fax: +91 11 25841564.
E-mail address: hchandra12@gmail.com (H. Chandra).

http://dx.doi.org/10.1016/j.spasta.2017.01.004
2211-6753/© 2017 Elsevier B.V. All rights reserved.

the direct survey estimator (defined using domain specific survey data only) provides reliable estimates with a pre-determined level of precision for planned domains in these surveys. However, policy planners, researchers, government and public agencies often require estimates for unplanned domains. Such unplanned domains can be small geographic areas (e.g. municipalities, census divisions, blocks, tehsils, gram panchayats, etc.) or small demographic groups (e.g. age–sex–race groups within larger geographical areas) or a cross classification of both. The sample sizes for such unplanned domains in the available survey data may be very small or even zero. In the survey literature, a domain is regarded as small if the domain-specific sample size is not large enough to ensure that a direct survey estimator has adequate precision (Rao, 2003). In such cases it becomes necessary to employ indirect small area estimators that make use of the sample data from related areas or domains through linking models, thus increasing the effective sample size in the small areas. Such estimators can have significantly smaller coefficient of variation than direct estimators, provided the linking models are valid. The statistical methodology that tackles this problem of small sample sizes is often referred as small area estimation (SAE) theory in the survey literature, see Rao and Molina (2015). Based on the level of auxiliary information available, the models used in SAE are categorised as area level or unit level. Area-level modelling is typically used when unit-level data are unavailable, or, as is often the case, where model covariates (e.g. census variables) are only available in aggregate form. The Fay–Herriot model (Fay and Herriot, 1979) is a widely used area level model in SAE that assumes area-specific survey estimates are available, and that these follow an area level linear mixed model with independent area random effects. In economic, environmental and epidemiological applications, estimates for areas that are spatially close may be more alike than estimates for areas that are further apart. One approach to incorporating such spatial information in SAE modelling is to extend the random effects model to allow for spatially correlated area effects using, for example, a Simultaneous Autoregressive (SAR) model (Anselin, 1992; Cressie, 1993). Applications of SAR models in small area estimation have been considered by Singh et al. (2005), Pratesi and Salvati (2008), Pratesi and Salvati (2009), Molina et al. (2009), Marhuenda et al. (2013) and Porter et al. (2014).

Many small area applications are based on binary and count data. When the variable of interest is not continuous, the use of standard SAE methods based on linear mixed models becomes problematic. In this context, estimation of small area poverty ratios and other indicators related to poverty and food insecurity is often based on a generalised linear mixed model (GLMM). The most commonly used GLMMs are the logistic-normal mixed model (also referred as the logistic linear mixed model) for binary data and the general Poisson-normal mixed model (also referred as the log linear mixed model) for count data. When only area level data are available, an area level version of a GLMM can be used for SAE, see Johnson et al. (2010) and Chandra et al. (2011). This approach to SAE implicitly assumes that direct survey estimates from different small areas are uncorrelated. However the boundaries that define a small area are typically arbitrarily set, and there appears to be no good reason why neighbouring small areas should not be correlated. This can be the case, for example, with socio-economic, agricultural, environmental and epidemiological data. It is therefore reasonable to assume that the effects of neighbouring areas, defined via a contiguity criterion, are correlated. Saei and Chambers (2003) and Chandra and Salvati (2017) describe an extension of the area level version of GLMM that allows for spatially correlated random effects using a SAR model (SGLMM) and define an empirical predictor (SEP) for the small area proportion under this model. This model allows for spatial correlation in the error structure, while keeping the fixed effects parameters spatially invariant. A key feature of this approach is that it assumes that the parameters associated with the model covariates do not vary spatially.

There are situations, however, where this assumption is inappropriate (i.e. where fixed effects parameters are not the same everywhere in the study area), a phenomenon referred to as spatial nonstationarity, see for example Brunsdon et al. (1996) and the references therein. Geographical weighted regression (GWR) is an approach that is widely used for fitting data exhibiting spatial nonstationarity (Brunsdon et al., 1998; Fotheringham et al., 2002). Under GWR the data are assumed to follow a location specific regression function, with geographically defined weights used to estimate the parameters of this local regression function. Chandra et al. (2015) describe a spatial nonstationary extension of the Fay–Herriot model. In this paper we use the GWR concept to extend the GLMM to incorporate spatial nonstationarity, which we refer to as the NSGLMM, and then apply this model in

Download English Version:

https://daneshyari.com/en/article/5119011

Download Persian Version:

https://daneshyari.com/article/5119011

Daneshyari.com