University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

1-1-2017

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

Hukum Chandra

Indian Agricultural Statistics Research Institute, hchandra@uow.edu.au

Nicola Salvati

University of Pisa, salvati@ec.unipi.it

Raymond L. Chambers

University of Wollongong, ray@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>

Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

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

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.

Disciplines

Engineering | Science and Technology Studies

Publication Details

Chandra, H., Salvati, N. & Chambers, R. (2017). Small area prediction of counts under a non-stationary spatial model. *Spatial Statistics*, 20 30-56.

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

Hukum Chandra^{a,*}, Nicola Salvati^b and Ray Chambers^c

^aIndian Agricultural Statistics Research Institute, Library Avenue, New Delhi, India

^bDipartimento di Economia e Management, University of Pisa, Italy

^cNational Institute for Applied Statistics Research Australia, University of Wollongong, Australia.

ABSTRACT

There is a growing need for current and reliable counts at small area level. The empirical predictor under a generalized 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.

Key words: Geographically weighted regression, Spatial nonstationarity, Count data, MSE estimation, Poverty mapping.

*Corresponding Author: Phone: +91-11-25843398, Fax: +91-11-25841564,
Email: hchandra12@gmail.com (Hukum Chandra)

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 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 categorized 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 generalized 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 SAE via its corresponding empirical predictor (NSEP). Throughout this article we assume that only aggregated level data are available, and consider SAE under an area level version of the small area model. The proposed approach is motivated by a study aimed at obtaining estimates of proportions of poor households within Districts in the State of Uttar Pradesh in India, using survey data from the Household Consumer Expenditure (HCE) Survey 2011-12 of the National Sample Survey Office (NSSO) 68th round and the Population Census 2011. Diagnostics applied to these data indicate the presence of spatial nonstationarity, and in this case the NSEP provides efficient estimates for the Districts.

Another area of statistical inference using small area counts is ecological regression on disease mapping, where data on disease incidence across a geographical region, typically numbers of cases recorded within administrative subdivisions, is combined with administrative data on risk factors with the aim of estimating disease risk within these subdivisions. This can be regarded as a special case of SAE (Rao, 2003, chapter 9), in which Bayesian and empirical Bayes methods are combined with a GLMM specification for the observed counts to provide estimates of risk by borrowing information across areas. Ecological regression on disease mapping mainly focuses on the estimation of risk in administrative subdivisions and on the analysis of the association between risk factors and disease incidence. Since these data usually exhibit over dispersion, Clayton and Kaldor (1987) propose the use of a Poisson-gamma model and an Empirical Bayes approach to estimating relative risks. This was generalized by Besag *et al.* (1991) using a Hierarchical Bayes approach that also allowed for spatial structure. More recently, Chambers *et al.* (2014) have proposed an alternative, and potentially more robust, approach to disease mapping based on negative binomial regression M-quantiles for count data. This approach is based on the M-quantile extension of the robust estimating equations for generalized linear models developed by Cantoni and Ronchetti (2001). Like these approaches, the methods developed in this paper can also be used for disease mapping. However it should

be noted that the majority of methods used in disease mapping employ the Bayesian paradigm while the model and predictors proposed here are motivated using frequentist arguments.

This paper is organized as follows. Section 2 is devoted to setting out the theoretical background of the area level version of the GLMM which is then used to define the empirical predictor (EP) for small areas. In Section 3 we describe the data from the Household Consumer Expenditure Survey 2011-12 of NSSO 68th round and the Population Census 2011 and report diagnostics aimed at assessing its spatial stationarity. The extension of the area level GLMM to spatially nonstationary data (the NSGLMM) is introduced in Section 4 and SAE using an empirical predictor (the NSEP) based on this model is described. Note that throughout this paper we focus on the area level GLMM with a logistic link function (i.e. the logistic-normal mixed model), applicable when the underlying area level data are proportions, and the GLMM with log link function (i.e. the general Poisson-normal mixed model), applicable when these area level data are counts. Our development can be easily generalised to other types of area level GLMMs. Analytic and bootstrap estimators of the mean squared error (MSE) of the NSEP are developed in Section 5. Simulation-based empirical results are discussed in Section 6 and application of the NSEP to poverty mapping in Uttar Pradesh is presented in Section 7. Section 8 summarizes the main conclusions and discusses further research areas.

2. Small area estimation for discrete data

Let U denote a finite population of interest of size N and suppose that a sample s of size n is drawn from this population with a given survey design. The population is assumed to consist of m small areas or small domains (or simply areas or domains), which we index by $i = 1, \dots, m$. The set of population units in area i is denoted U_i , with known size N_i , so that $U = \bigcup_{i=1}^m U_i$ and $N = \sum_{i=1}^m N_i$. Following standard practice, the subscript s and r are used to denote units in the sample and non-sampled parts of the population. With n_i and N_i denoting the sample and population sizes in area i , respectively. The units making up the sample in area i are denoted by s_i , so $s = \bigcup_{i=1}^m s_i$ and $n = \sum_{i=1}^m n_i$. Let y_{ij} denote the

value of a target variable y for unit j in small area i . We assume that this variable is a "counting" variable, i.e. it takes values in the set of non-negative integers (with binary as a special case) and that the aim is to estimate the small area population counts $y_i = \sum_{j \in U_i} y_{ij}$, or equivalently the small area proportions $P_i = N_i^{-1} \left(\sum_{j \in U_i} y_{ij} \right)$. Let $y_{si} = \sum_{j \in s_i} y_{ij}$ denote the sample count in small area i . We assume that the available data consist of these sample aggregates, together with the values of area specific contextual covariates. That is, for area i we only observe the count y_{si} together with a p -vector of area-specific covariates $\mathbf{x}_i$ derived from secondary data sources (e.g. the census or administrative registers). Let $\mathbf{u} = (u_1, \dots, u_m)^T$ denote a vector of m independent area-specific random effects, each distributed as Gaussian with mean zero and variance σ_u^2 . The variable y_i , conditioned on the random effect u_i , is then assumed to belong to the exponential family of distributions, i.e. $y_i | u_i \sim f(\eta_i)$, where η_i is a function of u_i and

$$f(\eta_i) = \exp \left[\frac{y_i \eta_i - a(\eta_i)}{\theta_i} + b(y_i, \theta_i) \right]. \quad (1)$$

The distribution (1) is parameterized by the canonical parameter η_i , known scale parameter θ_i and functions $a(\cdot)$ and $b(\cdot)$, and includes well-known distributions like the Gaussian, Binomial and Poisson. Since our interest is in modelling counts, we focus on the Binomial and Poisson versions of (1) where the parameter η_i is assumed to have a linear mixed model specification. That is, following Johnson *et al.* (2010) and Chandra *et al.* (2011), we assume that the area level counts y_i can be modelled via a GLMM of the form

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i, \quad (2)$$

where $\mu_i = E(y_i | u_i) = g^{-1}(\eta_i)$. Here $g(\cdot)$ is a specified link function (we will assume either a logit or a log specification for this link) and $\boldsymbol{\beta}$ is a p -vector of regression coefficients, often referred to as the vector of fixed effect parameters. In this model, the area-specific random effect u_i is assumed to account for between area variability in the parameter η_i beyond that explained by the contextual covariates $\mathbf{x}_i$ included in the fixed part of the

model. This type of model is often referred to as ‘area-level’ model in SAE terminology, see for example (Rao and Molina, 2015; Fay and Herriot, 1979).

Throughout this paper, we make the standard assumption that sampling within areas is non-informative given the values of the contextual variables and the random area effects. Also, as noted earlier, this article focuses on proportions and on counts, and characterises their expected values using the popular logit and log link functions, respectively. Consequently, when the target of inference is a small area proportion, then the sample count y_{si} in area i will be assumed to follow a Binomial distribution with parameters n_i and π_i , i.e., $y_{si} \sim \text{Bin}(n_i, \pi_i)$, where π_i is the probability of occurrence of an event for a population unit in area i or the probability of prevalence in area i . The model (2) linking π_i with the covariates $\mathbf{x}_i$ is then the GLMM with logistic link function

$$\text{logit}(\pi_i) = \ln \left\{ \pi_i (1 - \pi_i)^{-1} \right\} = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i,$$

with

$$\pi_i = \exp(\eta_i) \{1 + \exp(\eta_i)\}^{-1} = \text{expit}(\mathbf{x}_i^T \boldsymbol{\beta} + u_i).$$

Under this model, the mean of y_{si} given u_i is $\mu_{si} = E(y_{si} | u_i) = n_i \text{expit}(\mathbf{x}_i^T \boldsymbol{\beta} + u_i)$. If the target of inference is a count, then $y_{si} \sim \text{Poisson}(n_i \mu_i)$ and the link function (2) is the log function

$$\log(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i,$$

with

$$\mu_i = \exp(\eta_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta} + u_i).$$

We then have $\mu_{si} = E(y_{si} | u_i) = n_i \exp(\mathbf{x}_i^T \boldsymbol{\beta} + u_i)$.

Estimation of the fixed effect parameter and the area specific random effects uses the data from all small areas. Without loss of generality we focus on the Binomial case and put $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)^T$ and $\mathbf{X} = (\mathbf{x}_1^T, \dots, \mathbf{x}_m^T)^T$. We assume that $\mathbf{u} = (u_1, \dots, u_m)^T$ is a $m \times 1$ vector of area random effects which has a Gaussian distribution with zero mean vector and

covariance matrix $\Sigma_u = \sigma_u^2 \mathbf{I}_m$. Here $\mathbf{I}_m$ is the identity matrix of order m . Collecting together the different area level models (2), we write the model at population level as

$$g(\boldsymbol{\pi}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (3)$$

where $g(\boldsymbol{\pi}) = (g(\pi_1), \dots, g(\pi_m))^T$ and $\boldsymbol{\eta} = (\eta_1, \dots, \eta_m)^T$. We adopt a Penalized Quasi-Likelihood (PQL) method of estimation for the parameters $\boldsymbol{\beta}$ and $\mathbf{u}$ in the GLMM (3), combined with restricted maximum likelihood estimation of the variance parameter σ_u^2 . See Saei and Chambers (2003) and Manteiga *et al.* (2007). Under (2), a plug-in empirical predictor (EP) of the population count y_i in area i is

$$\hat{y}_i^{EP} = y_{si} + \hat{\mu}_{ri} = y_{si} + (N_i - n_i) \hat{\pi}_i^{EP}, \quad (4)$$

where $\hat{\pi}_i^{EP} = \text{expit}(\hat{\eta}_i)$ for a binary target variable and $\hat{\pi}_i^{EP} = \exp(\hat{\eta}_i)$ for a count-valued target variable, with $\hat{\eta}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{u}_i$. An estimate of the corresponding proportion or rate in area i is obtained as $\hat{P}_i^{EP} = N_i^{-1} \hat{y}_i^{EP}$. Note that (4) differs from the empirical predictor $\hat{y}_i^B = N_i \hat{\pi}_i^{EP}$ considered by Boubeta *et al.* (2016) in that it includes rather than predicts observed y -values. Clearly, when sample sizes are small relative to population sizes there will be very little difference between these two predictors. Note also that (4) can be represented as a weighted combination of the direct estimator $\hat{y}_i^{dir} = N_i n_i^{-1} y_{si}$ and $\hat{y}_i^B$, i.e. $\hat{y}_i^{EP} = n_i N_i^{-1} \hat{y}_i^{dir} + (1 - n_i N_i^{-1}) \hat{y}_i^B$. For areas with zero sample sizes (i.e. non-sampled areas), the conventional approach to estimating area proportions or counts is synthetic estimation, based on a suitable GLMM fitted to the counts from the sampled areas (Chandra *et al.*, 2011). For non-sampled area i with associated vector of covariates $\mathbf{x}_{i,out}$, $\hat{\eta}_i = \mathbf{x}_{i,out}^T \hat{\boldsymbol{\beta}}$ and the synthetic estimator of y_i under (2) is $\hat{y}_i^{SYN} = N_i \hat{\pi}_i^{SYN}$, with $\hat{\pi}_i^{SYN} = \text{expit}(\hat{\eta}_i)$ for a binary target variable and $\hat{\pi}_i^{SYN} = \exp(\hat{\eta}_i)$ for a count-valued target variable. An alternative predictor has been proposed by Jiang (2003). Unfortunately, this predictor does not have a closed form and can only be computed via numerical approximation. This is generally not straightforward, and so many users tend to favour computation of analytic approximations like $\hat{P}_i^{EP}$.

3. Data sources and model specification

In this Section we describe the basic sources of the data used in our application i.e. the Indian Household Consumer Expenditure Survey 2011-12 of NSSO 68th round and the Indian Population Census 2011. Data obtained from these sources were then used to estimate the proportion of poor households at District level in the Indian State of Uttar Pradesh. The sampling design used in the NSSO survey is stratified multi-stage random sampling with Districts as strata, villages as first stage units and households as second stage units. A total of 5916 households were surveyed from the 71 Districts of Uttar Pradesh. The District sample sizes range from 32 to 128 with average of 83. It is evident that these District level sample sizes are relatively small, with an average sampling fraction of 0.0002. As a consequence, it is difficult to derive reliable District level estimates with associated standard errors from this NSSO survey, and application of SAE methodology is advisable (Rao, 2003).

We note here that the target variable at the unit (household) level in the published survey data file is binary, corresponding to whether a household is poor or not. In our application however we focus on estimation where the available data are the corresponding counts of the number of poor households in sample in each District. In this context a household having monthly per capita consumer expenditure below the state's poverty line is defined as being poor. The poverty line used in this study (Rs. 768) is the same as that set by the Planning Commission, Govt. of India, for 2011-12. The parameter of interest is then the proportion of poor households within each District.

The auxiliary (covariates) variables used in our analysis were taken from the Indian Population Census 2011. These auxiliary variables are only available as counts at District level, and there are approximately 50 such covariates that are available for use in SAE analysis. We therefore carried out a preliminary data analysis in order to define appropriate covariates for SAE modelling, using Principal Component Analysis (PCA) to derive composite scores for selected groups of variables. In particular, we carried out PCA separately on three groups of variables, all measured at District level and identified as G1,

G2 and G3 below. The first group (G1) consisted of literacy rates by gender and proportions of worker population by gender. The first principal component for this group explained 51% of the variability in the G1 group, while adding the second principal component (G12) increased this to 85%. The second group (G2) consisted of the proportions of main worker by gender, proportions of main cultivator by gender and proportions of main agricultural labourer by gender. The first principal component (G21) for this second group explained 49% of the variability in the G2 group, while adding the second component (G22) increased this to 67%. Finally, the third group (G3) consisted of proportions of marginal cultivator by gender and proportions of marginal agriculture labourers by gender. The first principal component (G31) for this third group explained 61% of the variability in the G3 group, while adding the second component (G22) increased this to 78%. We then fitted a generalised linear model using direct survey estimates of proportions of poor households as the response variable and the six principal component scores G11, G12, G21, G22, G31 and G32 as potential covariates. The final selected model included the three covariates G11, G21 and G31, with residual deviance and AIC values of 327.18 and 636.89 respectively. This final model was then used to produce district wise estimates of poverty incidence, i.e. estimates of the head count ratio (HCR) used in poverty mapping.

4. SAE under a NSGLMM

The vector of fixed effect parameters $\boldsymbol{\beta}$ in (4) is spatially invariant. However there are situations where this assumption is inappropriate, and the parameters associated with the model covariates vary spatially. In this section we define a spatial nonstationary extension of (4) that can be used in such situations. Let d_i denote the coordinates of an arbitrary spatial location (longitude and latitude) in area i . Typically, this will be its centroid. Let $\mathbf{d} = (d_1, \dots, d_m)^T$ denote the corresponding m -vector of such spatial locations, and let $\pi_i(d_i)$ be the probability of occurrence of a characteristic of interest in area i , defined relative to the location d_i . A model for a nonstationary GLMM (NSGLMM) for $\pi_i(d_i)$ is then

$$g(\pi_i(d_i)) = \eta_i(d_i) = \mathbf{x}_i^T \boldsymbol{\beta}(d_i) + u_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \boldsymbol{\gamma}(d_i) + u_i, \quad (5)$$

where we assume that nonstationarity is characterised by an area-specific vector of fixed effects $\beta(d_i) = \beta + \gamma(d_i)$; u_i is the area-specific random effect, assumed to follow a Gaussian distribution with zero mean and variance ϕ ; and $\gamma(d) = (\gamma_1(d), \dots, \gamma_p(d))^T$ is a spatially correlated vector-valued random process with $E(\gamma(d_i)) = \mathbf{0}$ and such that

$$\text{cov}(\gamma_k(d_i), \gamma_l(d_j)) = c_{kl} (1 + L(d_i, d_j))^{-1}. \quad (6)$$

Here $L(d_i, d_j)$ is the spatial distance between locations l_i and l_j and $\mathbf{c} = (c_j)$ is a p -vector of unknown positive constants that satisfies the conditions for the $pm \times pm$ matrix $\Sigma_\gamma = \mathbf{\Omega} \otimes (\mathbf{c}\mathbf{c}^T)$ to be a covariance matrix, with $\mathbf{\Omega} = \left[(1 + L(l_i, l_j))^{-1} \right]$ defining the matrix of distances between the sample areas, and where $\otimes$ denotes Kronecker product. We can write the population level version of (5) as

$$g(\pi(d)) = \eta(d) = \mathbf{X}\beta + \mathbf{Z}\gamma(d) + \mathbf{u} \quad (7)$$

where $\pi(d) = g^{-1}(\eta(d))$, with $\eta(d) = (\eta_1(d), \dots, \eta_m(d))^T$, $\mathbf{Z} = \{\text{diag}(\mathbf{x}_1), \dots, \text{diag}(\mathbf{x}_m)\}$ is a $m \times pm$ matrix and $\gamma(d)$ is a $pm \times 1$ vector of spatial Gaussian random effects that capture the spatial nonstationarity in the data. See (3) for definitions of all other quantities. We assume that $\gamma(d)$ has a zero mean vector and a covariance matrix $\Sigma_\gamma = \mathbf{\Omega} \otimes (\mathbf{c}\mathbf{c}^T)$. In general, the only constraint on the vector $\mathbf{c}$ is that $\Sigma_\gamma = \mathbf{\Omega} \otimes (\mathbf{c}\mathbf{c}^T)$ is symmetric and non-negative definite. In practice ϕ and $\mathbf{c}$ are unknown and have to be estimated from the data. We restrict ourselves to the simple specification $\mathbf{c} = \sqrt{\lambda} \mathbf{1}_p$ so that $\text{cov}(\gamma_k(d_i), \gamma_l(d_j)) = \lambda (1 + L(d_i, d_j))^{-1}$, where $\lambda \geq 0$ and $\mathbf{1}_p$ denotes the unit vector of order p . In this case, we assume that the distance metric used to define $L(d_i, d_j)$ is such that the matrix $\mathbf{\Omega} \otimes (\mathbf{1}_p \mathbf{1}_p^T)$ is positive semidefinite, with the parameter λ then reflecting the 'intensity' of spatial clustering in the data, so $\lambda = 0$ corresponds to the situation where the model is spatially homogeneous. Given this specification, there are just 2 parameters (λ and ϕ) that need to be estimated. Replacing these unknown parameters by their

estimated values $\hat{\phi}$ and $\hat{c}$, and denoting subsequent plug-in estimators by a 'hat', we define the nonstationary empirical predictor (NSEP) of the population count in area i as

$$\hat{y}_i^{NSEP} = y_{si} + (N_i - n_i) \hat{\pi}_i^{NSEP}(d_i), \quad (8)$$

with $\hat{\pi}_i^{NSEP}(d_i) = \text{expit}[\hat{\eta}_i(d_i)]$ and $\hat{\pi}_i^{NSEP}(d_i) = \exp[\hat{\eta}_i(d_i)]$ for binary and count-valued data respectively, where $\hat{\eta}_i(d_i) = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \mathbf{z}_i^T \hat{\boldsymbol{\gamma}}(d_i) + \hat{u}_i$. Here $\mathbf{z}_i^T$ is the i -th row of $\mathbf{Z}$. A nonstationary empirical predictor of the proportion (or rate) in area i is $\hat{p}_i^{EP} = N_i^{-1} \hat{y}_i^{EP}$.

Extending (8) to synthetic prediction for a non-sample area is straightforward. We set the estimated area effect to zero in this case, and evaluate (8) at the location d_i of the non-sampled area. The result is a nonstationary synthetic predictor (NSSYN) of the total count for the area of the form $\hat{y}_i^{NSSYN} = N_i \hat{\pi}_i^{NSSYN}(d_i)$, with $\hat{\pi}_i^{NSSYN}(d_i) = \text{expit}[\hat{\eta}_i(d_i)]$ and $\hat{\pi}_i^{NSSYN}(d_i) = \exp[\hat{\eta}_i(d_i)]$ for binary and count-valued data respectively, and where now $\hat{\eta}_i(d_i) = \mathbf{x}_{out,i}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{out,i}^T \hat{\boldsymbol{\gamma}}(d_i)$. Note that in contrast to the SYN predictor, the NSSYN predictor uses the location data for the non-sampled areas to 'borrow strength' from neighbouring sampled areas, and so has the potential to improve conventional synthetic prediction for non-sampled areas. In particular, we expect that if in fact the population data exhibit spatial nonstationarity, then NSSYN will be less biased than SYN.

4.1. Parameter estimation for the NSGLMM

Let $f_1(\mathbf{y}_s(d) | \mathbf{u}, \boldsymbol{\gamma}(d))$ be the probability density function of $\mathbf{y}_s(d)$ conditional on the vector of random effects $\{\mathbf{u}, \boldsymbol{\gamma}(d)\}$ and let $f_2(\mathbf{u})$ and $f_3(\boldsymbol{\gamma}(d))$ be the probability density functions of $\mathbf{u}$ and $\boldsymbol{\gamma}(d)$ respectively. Here $\mathbf{y}_s(d) = (y_{s1}(d_1), \dots, y_{sm}(d_m))^T$ is a $m \times 1$ vector of sample counts where $y_{si}(d_i)$ has Binomial (or Poisson) distribution whereas $f_2(\mathbf{u})$ and $f_3(\boldsymbol{\gamma}(d))$ have normal distributions with zero means and covariance matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$ respectively. The 'log-likelihood' for $\mathbf{y}_s(d)$ conditional on the realised value of $\{\mathbf{u}, \boldsymbol{\gamma}(d)\}$ and the 'log-likelihood' for $\mathbf{u}$ and $\boldsymbol{\gamma}(d)$ are then

$$\begin{aligned}
l_1 &= \ln \{f_1(\mathbf{y}_s(d) | \mathbf{u}, \boldsymbol{\gamma}(d))\} \\
&= \begin{cases} \text{Constant} + \sum_{i=1}^m \{y_{si}(d_i)\eta_i(d_i) - n_i \ln [1 + \exp(\eta_i(d_i))]\} & \text{for Binomial data} \\ \text{Constant} + \sum_{i=1}^m \{y_{si}(d_i)\eta_i(d_i) - n_i \exp(\eta_i(d_i))\} & \text{for Poisson data} \end{cases} \\
l_2 &= \text{Constant} - \frac{1}{2} \left\{ \ln |\boldsymbol{\Sigma}_u| + \mathbf{u}^T \boldsymbol{\Sigma}_u^{-1} \mathbf{u} \right\} \\
l_3 &= \text{Constant} - \frac{1}{2} \left\{ \ln |\boldsymbol{\Sigma}_\gamma| + \boldsymbol{\gamma}(d)^T \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\gamma}(d) \right\}.
\end{aligned}$$

For fixed $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$, the values of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}(d)$ and $\mathbf{u}$ that jointly maximise $l = l_1 + l_2 + l_3$ are the maximum penalised quasi-likelihood (MPQL) estimates of these parameters. See Saei and McGilchrist (1998). The ‘log-likelihood’ l is not a likelihood in the conventional sense because it is based on the non-observable $\mathbf{u}$ and $\boldsymbol{\gamma}(d)$. Substituting these estimates into (7) yields the MPQL estimate of $\boldsymbol{\pi}(d)$. In practice the variance components that define the matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$ are unknown and must be estimated from the sample data. It is well known that the MPQL estimates of these variance components are biased, and that this bias increases with the relative contributions of the associated random effects to overall variability. Consequently, this approach is not recommended. Alternative estimates based on a hybrid of MPQL for fixed and random effects and maximum likelihood (ML) or restricted maximum likelihood (REML) for variance components can be defined. These can reduce, but not eliminate, the aforementioned bias. Since prediction of small area quantities, rather than parameter estimation, is our focus, we continue with this hybrid approach, returning to the issue in Section 6 where we provide simulation results that provide a perspective on this bias problem.

Under the hybrid approach, parameter estimates for the NSGLMM are obtained by a two-stage iterative process. At the first stage, MPQL estimates of $\boldsymbol{\beta}$, $\mathbf{u}$ and $\boldsymbol{\gamma}(d)$ are calculated based on specified values of $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$, and at the second stage $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$ are estimated via ML or REML given these MPQL estimates. This hybrid approach, where MPQL estimation is combined with ML or REML estimation in generalized linear mixed models

is due to McGilchrist (1994). Below we outline an algorithm that extends this approach to estimation of the parameters of the GLMM (5).

MPQL Stage

Given Σ_u (or ϕ) and Σ_γ (or λ), the MPQL estimates of β , $\mathbf{u}$ and $\gamma(d)$ are computed via the following iterative process:

1. Initialise the iteration. Set $k=0$ and initial values β_0 , $\gamma_0(d)$ and $\mathbf{u}_0$.
2. Update these values via

$$\begin{bmatrix} \beta_{k+1} \\ \gamma_{k+1}(d) \\ \mathbf{u}_{k+1} \end{bmatrix} = \begin{bmatrix} \beta_k \\ \gamma_k(d) \\ \mathbf{u}_k \end{bmatrix} + \mathbf{V}_k^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix} \frac{\partial l_1}{\partial \boldsymbol{\eta}_k} - \mathbf{V}_k^{-1} \begin{bmatrix} \mathbf{0} \\ \Sigma_\gamma^{-1} \gamma_k(d) \\ \Sigma_u^{-1} \mathbf{u}_k \end{bmatrix}$$

where

$$\mathbf{V}_k = - \begin{bmatrix} \mathbf{X}^T \\ \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix} \left(\frac{\partial^2 l_1}{\partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_k^T} \right) \begin{bmatrix} \mathbf{X} & \mathbf{Z} & \mathbf{I}_m \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_\gamma^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_u^{-1} \end{bmatrix}$$

and $\frac{\partial l_1}{\partial \boldsymbol{\eta}_k}$, $\frac{\partial^2 l_1}{\partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_k^T}$ are the first and second derivatives of l_1 with respect to $\boldsymbol{\eta}$ and

evaluated at $\boldsymbol{\eta}_k$. These are evaluated for given $\boldsymbol{\eta}_k$ via model (7).

3. Return to step 2.
4. At convergence, update $\boldsymbol{\eta}(d) = \mathbf{X}\beta + \mathbf{Z}\gamma(d) + \mathbf{u}$.

We suggest setting $\phi=0.5$ and $\lambda=1$ to start the initial MPQL stage. These values are generally a good compromise for fast convergence of the algorithm.

ML/REML Stage

Given the estimates of β , $\mathbf{u}$ and $\gamma(d)$ from steps 1 - 3 of the MPQL stage, we use ML or REML to estimate the variance components Σ_u (or ϕ) and Σ_γ (or λ). Following Schall (1991), define an adjusted variable

$$\mathbf{y}^* = g(\mathbf{y}) \cong g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu}) g'(\boldsymbol{\mu}) = \boldsymbol{\eta} + \mathbf{e} \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}} \right) = \boldsymbol{\eta} + \mathbf{e}^*$$

where $\mathbf{e}^* = \mathbf{e} \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}} \right)$. Under the GLLM (5), $\mathbf{y}^* \cong \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}(d) + \mathbf{u} + \mathbf{e}^*$ where $\mathbf{e}^* = (e_i^*)$ with $\text{var}(\mathbf{e}^*) = \tau \boldsymbol{\Sigma}_e^*$. Here τ is a dispersion parameter and

$$\boldsymbol{\Sigma}_e^{*-1} = \text{diag} \left[V(\mu_i)^{-1} \left\{ \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right\}^2 \right].$$

For either the Binomial or the Poisson models, this variance is a known function of the mean and so $\tau = 1$. Finally, we assume that $\mathbf{y}^*$ is approximately Gaussian with $E(\mathbf{y}^*) = \mathbf{X}\boldsymbol{\beta}$ and $\text{var}(\mathbf{y}^*) = \mathbf{V} = \boldsymbol{\Sigma}_u + \mathbf{Z}\boldsymbol{\Sigma}_\gamma \mathbf{Z}^T + \boldsymbol{\Sigma}_e^*$. Maximising the log-likelihood generated by $\mathbf{y}^*$ with respect to $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$ we obtain the solutions

$$\boldsymbol{\phi} = m^{-1} [\text{tr}(\mathbf{T}) + \mathbf{u}^T \mathbf{u}]$$

and

$$\boldsymbol{\lambda} = (mp)^{-1} \left[\text{tr} \left(\mathbf{T} (\boldsymbol{\Omega} \otimes \mathbf{1}_p \mathbf{1}_p^T)^{-1} \right) + \boldsymbol{\gamma}(d)^T (\boldsymbol{\Omega} \otimes \mathbf{1}_p \mathbf{1}_p^T)^{-1} \boldsymbol{\gamma}(d) \right]$$

where

$$\mathbf{T} = \left\{ \begin{bmatrix} \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix} \boldsymbol{\Sigma}_e^{*-1} \begin{bmatrix} \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix}^T + \begin{bmatrix} \lambda (\boldsymbol{\Omega} \otimes \mathbf{1}_p \mathbf{1}_p^T) & \mathbf{0} \\ \mathbf{0} & \phi \mathbf{I}_m \end{bmatrix}^{-1} \right\}^{-1}.$$

REML estimation of $\boldsymbol{\phi}$ and $\boldsymbol{\lambda}$ can be carried out by maximising the restricted log-likelihood generated of $\mathbf{y}^*$, where we replace $\mathbf{T}$ by

$$\mathbf{T}_{22} = \mathbf{T} + \mathbf{T} \begin{bmatrix} \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix}^T \boldsymbol{\Sigma}_e^{*-1} \mathbf{X} \mathbf{T}_{11} \mathbf{X}^T \boldsymbol{\Sigma}_e^{*-1} \begin{bmatrix} \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix} \mathbf{T}$$

with

$$\mathbf{T}_{11} = \left\{ \mathbf{X}^T \boldsymbol{\Sigma}_e^{*-1} \mathbf{X} - \mathbf{X}^T \boldsymbol{\Sigma}_e^{*-1} \begin{bmatrix} \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix} \mathbf{T} \begin{bmatrix} \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix}^T \boldsymbol{\Sigma}_e^{*-1} \mathbf{X} \right\}^{-1}.$$

Updated estimates of the variance components are then used as inputs to the MPQL stage. Both stages are finally iterated to convergence.

4.2. A diagnostic for spatial nonstationarity in the NSGLMM

Using the ideas set out in Opsomer *et al.* (2008) and Chandra *et al.* (2015), we describe a bootstrap procedure to test the spatial nonstationarity hypothesis in the context of the simple one parameter NSGLMM (5) considered in this paper, i.e. the hypothesis $H_0 : \lambda = 0$ versus the one-sided alternative $H_1 : \lambda > 0$. Two models are fitted, first without spatial random effects, and second with these effects. The test then involves comparing the restricted log-likelihoods under each hypothesis. Put $\mathbf{y}_0^* \cong \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e}^*$, $\mathbf{y}_1^* \cong \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}(d) + \mathbf{u} + \mathbf{e}^*$ and $\mathbf{P}_k = \mathbf{V}_k^{-1} - \mathbf{V}_k^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}_k^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}_k^{-1}$, where $\mathbf{V}_0 = \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_e^*$ and $\mathbf{V}_1 = \boldsymbol{\Sigma}_u + \mathbf{Z}\boldsymbol{\Sigma}_\gamma\mathbf{Z}^T + \boldsymbol{\Sigma}_e^*$. The test statistic is $\ell = 2(\ell_1 - \ell_0)$, where (modulo an additive constant)

$$\ell_k = -\frac{1}{2} \log |\mathbf{P}_k^T \mathbf{V}_k \mathbf{P}_k| - \frac{1}{2} (\mathbf{P}_k^T \mathbf{y}_k^*)^T (\mathbf{P}_k^T \mathbf{V}_k \mathbf{P}_k)^{-1} (\mathbf{P}_k^T \mathbf{y}_k^*)$$

is the restricted log-likelihood under H_k ($k = 0, 1$). The level of significance of ℓ is calculated via a parametric bootstrap. That is, substituting estimates $\hat{\phi}$ and $\hat{\beta}$ obtained under the null, we generate bootstrap realisations $\boldsymbol{\pi}^{*(b)}$ of $\boldsymbol{\pi}$ as described in Section 5.2. For each bootstrap replication, the null and the alternative models are then fitted and the bootstrap value $\ell^{*(b)}$ of ℓ is calculated. The level of significance of the calculated value of ℓ is estimated by comparing it with the bootstrap distribution of $\ell^{*(b)}$.

5. Mean squared error estimation

Analytic estimation of the MSE of the small area empirical predictor SEP (4) follows along the same lines as reported in Saei and Chambers (2003), Johnson *et al.* (2010), Chandra *et al.* (2011) and references therein. A corresponding analytic approach to estimating the MSE of the NSEP (8) under the NSGLMM (7) is developed in Section 5.1 below. In Section 5.2 we propose a parametric bootstrap procedure for estimating this MSE.

5.1. An analytic MSE estimator for the NSEP

Write $\mathbf{W} = (\mathbf{Z}, \mathbf{I}_m)$, $\mathbf{v} = (\boldsymbol{\gamma}^T(d), \mathbf{u}^T)^T$ and $\boldsymbol{\Sigma}_v = \text{var}(\mathbf{v}) = \text{diag}\{\boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_\gamma\}$. Let $\boldsymbol{\delta} = (\phi, \lambda)^T$ denote the variance components underpinning $\boldsymbol{\Sigma}_v$. Let $\mathbf{y} = (y_1, \dots, y_m)^T$, $\mathbf{y}_s = (y_{s_1}, \dots, y_{s_m})^T$ and $\mathbf{y}_r = (y_{r_1}, \dots, y_{r_m})^T$ denote the vectors of population, sample and non-sample counts, respectively, so $\mathbf{y} = \mathbf{y}_s + \mathbf{y}_r$. From (7), $\boldsymbol{\eta}(d) = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{v}$ and so $\hat{\boldsymbol{\eta}}(d) = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{W}\hat{\mathbf{v}}$. The NSEP (8) is $\hat{\mathbf{y}}^{NSEP} = \mathbf{y}_s + \hat{\boldsymbol{\mu}}_r$, where $\hat{\boldsymbol{\mu}}_r = \text{diag}\{N_i - n_i; i = 1, \dots, m\} \hat{\boldsymbol{\pi}}^{NSEP}(d)$ and $\hat{\boldsymbol{\pi}}^{NSEP}(d) = (\hat{\pi}_1^{NSEP}(d_1), \dots, \hat{\pi}_m^{NSEP}(d_m))^T$. Similarly $\boldsymbol{\mu}_r = \text{diag}\{N_i - n_i; i = 1, \dots, m\} \boldsymbol{\pi}(d)$ with $\boldsymbol{\pi}(d) = (\pi_1(d_1), \dots, \pi_m(d_m))^T$. Here $\hat{\pi}_i^{NSEP}(d_i) = \text{expit}[\mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \mathbf{z}_i^T \hat{\boldsymbol{\gamma}}(d_i) + \hat{u}_i]$ for binary data and $\hat{\pi}_i^{NSEP}(d_i) = \exp[\mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \mathbf{z}_i^T \hat{\boldsymbol{\gamma}}(d_i) + \hat{u}_i]$ for count-valued data. In the case of the empirical predictor $\hat{\mathbf{y}}^B = \hat{\boldsymbol{\mu}}$ of Boubeta *et al.* (2016), $\hat{\boldsymbol{\mu}} = \text{diag}\{N_i; i = 1, \dots, m\} \hat{\boldsymbol{\pi}}^{EP}$ with $\hat{\boldsymbol{\pi}}^{EP} = (\hat{\pi}_1^{EP}, \dots, \hat{\pi}_m^{EP})^T$ so $\boldsymbol{\mu} = \text{diag}\{N_i; i = 1, \dots, m\} \boldsymbol{\pi}$ with $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)^T$.

Define $\hat{\boldsymbol{\mu}}_s = \text{diag}\{n_i; i = 1, \dots, m\} \hat{\boldsymbol{\pi}}^{EP}$ and $\boldsymbol{\mu}_s = \text{diag}\{n_i; i = 1, \dots, m\} \boldsymbol{\pi}$. The prediction error of the NSEP is $\hat{\mathbf{y}}^{NSEP} - \boldsymbol{\mu} = \mathbf{y}_s + \hat{\boldsymbol{\mu}}_r - \boldsymbol{\mu} = \mathbf{y}_s + (\hat{\boldsymbol{\mu}}_r + \hat{\boldsymbol{\mu}}_s) - \hat{\boldsymbol{\mu}}_s - \boldsymbol{\mu} = (\mathbf{y}_s - \hat{\boldsymbol{\mu}}_s) + (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$. It follows that the prediction MSE of $\hat{\mathbf{y}}^{NSEP}$ is

$$MSE(\hat{\mathbf{y}}^{NSEP}) = E(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T + E(\mathbf{y}_s - \hat{\boldsymbol{\mu}}_s)(\mathbf{y}_s - \hat{\boldsymbol{\mu}}_s)^T + 2E(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\mathbf{y}_s - \hat{\boldsymbol{\mu}}_s)^T.$$

The second term and third terms in the above expansion for the MSE are associated with variability in the prediction of sample counts, and are of a lower order of magnitude than the first term. We therefore focus on the approximation $MSE(\hat{\mathbf{y}}^{NSEP}) \approx E(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T$.

Using first order Taylor linearization, we can write $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \approx (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^T \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \Big|_{\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}} \right)$. So

$$MSE(\hat{\mathbf{y}}^{NSEP}) \approx E(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T = \mathbf{A}^T \left\{ E[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^T] \right\} \mathbf{A}, \text{ with}$$

$$\mathbf{A} = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \Big|_{\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}} \right) = \begin{cases} \frac{\exp(\boldsymbol{\eta}(d))}{[1 + \exp(\boldsymbol{\eta}(d))]^2} = \boldsymbol{\mu}(1 - \boldsymbol{\mu}) & \text{for Binomial data} \\ \exp(\boldsymbol{\eta}(d)) = \boldsymbol{\mu} & \text{for Poisson data} \end{cases}.$$

Define $\hat{\boldsymbol{\theta}} = \mathbf{A}\hat{\boldsymbol{\eta}} = \mathbf{A}\{\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\gamma}}(d) + \hat{\mathbf{u}}\} = \mathbf{A}\{\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{W}\hat{\mathbf{v}}\} = \mathbf{P}\hat{\boldsymbol{\beta}} + \mathbf{Q}\hat{\mathbf{v}}$, with $\mathbf{P} = \mathbf{A}\mathbf{X}, \mathbf{Q} = \mathbf{A}\mathbf{W}$.

Then $\hat{\boldsymbol{\theta}}$ is a linear combination of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}}$ with

$$MSE(\hat{\mathbf{y}}^{NSEP}) \approx MSE(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \quad (9)$$

The results of Henderson (1963) and Prasad and Rao (1990) can be adapted to obtain an approximately unbiased estimator of the MSE of the NSEP. See Rao (2003, chapter 6, page 96-105) and Rao and Molina (2015, chapter 5, page 100-107). See also Saei and Chambers (2003) and Manteiga *et al.* (2007). An approximation to the MSE of the NSEP (8) is

$$MSE(\hat{\mathbf{y}}^{NSEP}) \approx M_1(\boldsymbol{\delta}) + M_2(\boldsymbol{\delta}) + M_3(\boldsymbol{\delta}). \quad (10)$$

The first two components $M_1(\boldsymbol{\delta})$ and $M_2(\boldsymbol{\delta})$ of (10) constitute the leading terms of this MSE approximation. These correspond to the MSE of the best linear unbiased predictor when $\boldsymbol{\delta}$ is assumed known. The third component $M_3(\boldsymbol{\delta})$ is the variability due to estimation of $\boldsymbol{\delta}$ (Rao, 2003, page 96-105), it can be proved that (10) is then a second order approximation to the MSE of the NSEP, see Rao and Molina (2015, chapter 5, page 105), Das *et al.* (2004), Opsomer *et al.* (2008) and references therein. The first component of the approximate MSE (10) is

$$M_1(\boldsymbol{\delta}) = \mathbf{Q}\boldsymbol{\Phi}\mathbf{Q}^T \quad (11)$$

with $\boldsymbol{\Phi} = \{\mathbf{W}^T\mathbf{B}\mathbf{W} + \boldsymbol{\Sigma}_v^{-1}\}^{-1}$ and where

$$\mathbf{B} = \begin{cases} \text{diag}\{n_i\pi_i(d_i)(1 - \pi_i(d_i)); i = 1, \dots, m\} & \text{for binary data;} \\ \text{diag}\{n_i\pi_i(d_i); i = 1, \dots, m\} & \text{for count-valued data} \end{cases}$$

Next, following McGilchrist (1994) and Saei and Chambers (2003), we define

$$\mathbf{V} = \begin{bmatrix} \mathbf{X}^T \\ \mathbf{W}^T \end{bmatrix} \mathbf{B} \begin{bmatrix} \mathbf{X} & \mathbf{W} \end{bmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{\Sigma}_v^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{B}\mathbf{X} & \mathbf{X}^T\mathbf{B}\mathbf{W} \\ \mathbf{W}^T\mathbf{B}\mathbf{X} & \boldsymbol{\Phi}^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$$

and

$$\mathbf{V}^{-1} = \mathbf{I}_{(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})}^{-1} = \text{Var}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})^T = \begin{pmatrix} \mathbf{X}^T\mathbf{B}\mathbf{X} & \mathbf{X}^T\mathbf{B}\mathbf{W} \\ \mathbf{W}^T\mathbf{B}\mathbf{X} & \boldsymbol{\Phi}^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix},$$

where the partitioning of the matrix $\mathbf{V}$ and its inverse $\mathbf{V}^{-1}$ is defined by the dimensions of $\boldsymbol{\beta}$ and $\mathbf{v}$. It follows

$$F_{11} = [\mathbf{X}^T \mathbf{B} \mathbf{X} - \mathbf{X}^T \mathbf{B} (\mathbf{W} \Phi \mathbf{W}^T) \mathbf{B} \mathbf{X}]^{-1} = [\mathbf{X}^T (\mathbf{B} - \mathbf{B} \mathbf{W} \Phi \mathbf{W}^T \mathbf{B}) \mathbf{X}]^{-1},$$

and

$$F_{22} = \Phi + \Phi \{ (\mathbf{W}^T \mathbf{B} \mathbf{X}) F_{11} (\mathbf{X}^T \mathbf{B} \mathbf{W}) \} \Phi.$$

Put $\mathbf{S} = \mathbf{A} [\mathbf{X} - \mathbf{W} \Phi \mathbf{W}^T \mathbf{B} \mathbf{X}]$. Then

$$M_2(\delta) = \mathbf{S} F_{11} \mathbf{S}^T. \quad (12)$$

Finally, we develop the expression for the third term $M_3(\delta)$, that is, the variability due to estimation of the unknown parameters of the variance components δ defining the covariance matrix Σ_v . Put $\Delta = \mathbf{A} \mathbf{W} \Phi = \mathbf{Q} \Phi$ and define $\mathbf{Q}_{(k)}$ to be the k^{th} row of the matrix $\mathbf{Q}$, with derivative $\nabla_{(k)} = \frac{\partial \Delta_{(k)}}{\partial \delta} = \frac{\partial (\mathbf{Q}_{(k)} \Phi)}{\partial \delta} = \frac{\partial (\mathbf{Q}_{(k)} \Phi)}{\partial (\lambda, \phi)^T} = \mathbf{Q}_{(k)} \Phi \Sigma_v^{-2} \Phi$. See Saei and

Chambers (2003). Then

$$M_3(\delta) = tr \left[\left(\nabla_{(k)} \Sigma_v^+ \nabla_{(l)}^T \right) \text{var}(\hat{\delta}) \right], \quad (13)$$

with $\Sigma^+ = \mathbf{W}^T \mathbf{B} \mathbf{W} + \mathbf{W}^T \mathbf{B} \mathbf{W} \Sigma_v \mathbf{W}^T \mathbf{B} \mathbf{W} = \mathbf{W}^T (\mathbf{B} + \mathbf{B} \mathbf{W} \Sigma_v \mathbf{W}^T \mathbf{B}) \mathbf{W}$. Here, $\text{var}(\hat{\delta})$ is the asymptotic covariance matrix of the estimates $\hat{\delta}$ of the variance components, which can be evaluated as the inverse of the appropriate Fisher information matrix. It is common practice to estimate the MSE of the predictors by replacing the unknown parameters including components of the variance by their respective estimators. However, this can lead to underestimation of the true MSE, see Rao (2003). Here, under NSGLMM (7), following the results of Prasad and Rao (1990) and Das *et al.* (2004), we note that

$$E(M_1(\hat{\delta}) - M_3(\hat{\delta})) = M_1(\delta) + o(m^{-1}),$$

$$E(M_2(\hat{\delta})) = M_2(\delta) + o(m^{-1}) \text{ and}$$

$$E(M_3(\hat{\delta})) = M_3(\delta) + o(m^{-1}).$$

That is, the bias of $M_1(\hat{\delta})$ is of the order of $M_3(\hat{\delta})$, see Rao and Molina (2015, chapter 5, page 105-107). An approximately unbiased estimator of the MSE (10) is therefore

$$mse(\hat{\mathbf{y}}^{NSEP}) = M_1(\hat{\delta}) + M_2(\hat{\delta}) + 2M_3(\hat{\delta}). \quad (14)$$

5.2. A parametric bootstrap MSE estimator

An alternative procedure for estimating the MSE of the NSEP (8) is based on the parametric bootstrap (Hall and Maiti, 2006). Note that the MSE estimator defined by this procedure is consistent provided the model parameter estimators are consistent. The steps of this parametric bootstrap procedure are as follows.

- 1) Given the vector of sample counts $\mathbf{y}_s$, calculate the parameter estimates using the method described in Section 4.1. Let $(\hat{\phi}, \hat{\lambda})$ and $\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\gamma}}(d)$ denote these estimates.
- 2) Given the estimates obtained in step 1, generate a vector $\mathbf{t}_1^*$ of length pm corresponding to a realisation from the $N(\mathbf{0}, \boldsymbol{\Omega} \otimes (\mathbf{1}_p \mathbf{1}_p^T))$ distribution. Construct the bootstrap vector $\boldsymbol{\gamma}^*(d) = \hat{\lambda}^{1/2} \mathbf{t}_1^*$.
- 3) Generate a vector $\mathbf{t}_2^*$ whose elements are m independent realisations of a $N(0,1)$ variable, independently of the generation of $\mathbf{t}_1^*$. Construct the bootstrap vector $\mathbf{u}^* = \hat{\phi}^{1/2} \mathbf{t}_2^*$.
- 4) Calculate the small area bootstrap population parameters $\boldsymbol{\pi}^*(d) = (\pi_1^*, \dots, \pi_m^*)^T = \text{expit}\{\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\boldsymbol{\gamma}^*(d) + \mathbf{u}^*\}$ or $\boldsymbol{\pi}^*(d) = \exp\{\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\boldsymbol{\gamma}^*(d) + \mathbf{u}^*\}$ depending upon whether the underlying data are binary or count-valued respectively.
- 5) Generate a bootstrap sample of independent Binomial realisations $\text{Bin}(n_i, \pi_i^*)$ or $\text{Poisson}(n_i \pi_i^*)$. Together with the values of $\mathbf{X}$ and d , calculate the bootstrap estimates $(\hat{\phi}^*, \hat{\lambda}^*)$ and $\hat{\boldsymbol{\beta}}^*, \hat{\mathbf{u}}^*, \hat{\boldsymbol{\gamma}}^*(d)$ using the method described in Section 4.1.
- 6) Use these estimates, and (8), to calculate the bootstrap value $\hat{y}_i^{*NSEP}$ of the NSEP.
- 7) Repeat steps 2 - 6 B times. In the b^{th} bootstrap replication, let $\hat{y}_i^{*NSEP(b)}$ be the bootstrap NSEP value in area i . The bootstrap estimator of the MSE of $\hat{y}_i^{NSEP}$ is then

$$mse_{boot}(\hat{y}_i^{NSEP}) = \mathbf{B}^{-1} \sum_{b=1}^B \left\{ \hat{y}_i^{*NSEP(b)} - \hat{y}_i^{NSEP} \right\}^2. \quad (15)$$

6. Simulation studies

In this section we use model-based simulation to investigate the performance of the NSEP (8), as well as its two MSE estimators (14) and (15). Model-based simulations are a common way of illustrating the sensitivity of an estimation procedure to variation in assumptions about the structure of the population of interest. In addition to allowing for spatial nonstationarity via simulated NSGLMM data, our simulations also allow for data generated via an area-level version of a spatial generalised linear mixed model (SGLMM) that assumes dependence between areas is due to SAR, i.e. simultaneous autoregression (Pratesi and Salvati, 2008; Saei and Chambers, 2003). In this alternative scenario we assume a spatially stationary SAR specification for the GLMM (3), replacing it by

$$g(\boldsymbol{\pi}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha}$$

where the vector of random area effects $\boldsymbol{\alpha} = (\alpha_i)$ satisfies $\boldsymbol{\alpha} = \rho\boldsymbol{\Omega}_s\boldsymbol{\alpha} + \mathbf{u}$, where ρ is a spatial autoregressive coefficient, $\boldsymbol{\Omega}_s$ is a proximity matrix of order m and $\mathbf{u} \sim N(0, \boldsymbol{\Sigma}_u = \sigma_u^2 \mathbf{I}_m)$. Since then $\boldsymbol{\alpha} = (\mathbf{I}_m - \rho\boldsymbol{\Omega}_s)^{-1}\mathbf{u}$, we see that $E(\boldsymbol{\alpha}) = \mathbf{0}$ and $\text{var}(\boldsymbol{\alpha}) = \sigma_u^2 [(\mathbf{I}_m - \rho\boldsymbol{\Omega}_s)(\mathbf{I}_m - \rho(\boldsymbol{\Omega}_s)^T)]^{-1}$. In this case the matrix $\boldsymbol{\Omega}_s$ specifies which random effects from neighbouring areas are related, whereas ρ defines the strength of this spatial relationship. It is standard practice to define $\boldsymbol{\Omega}_s$ as a contiguity matrix, i.e. the elements of $\boldsymbol{\Omega}_s$ take non-zero values only for those pairs of areas that share a common border. For ease of interpretation, this matrix is generally defined in row-standardized form, in which case ρ is called the spatial autocorrelation parameter (Pratesi and Salvati, 2008). Formally, we write $\boldsymbol{\Omega}_s = [\omega_{jk}]$, where $\omega_{jk} = 1$ if area j shares an edge with area k and $\omega_{jk} = 0$ otherwise. In row-standardised form this becomes $\omega_{jk}^* = c_j^{-1}$, if j and k are

contiguous and 0 otherwise, where c_j is the total number of areas that share an edge with area j (including area j itself). The same two stage iterative procedure combining MPQL and REML that was described in Section 4.1 is then used to estimate the parameters of this model. We refer to the empirical predictor (4) of the small area count defined under this SAR model as the spatial empirical predictor, or SEP. The different estimators evaluated in the simulation studies are set out in Table 1.

[Table 1 about here]

Let t index the T independent simulations in a study, and let $\hat{y}_{it}$ denote an estimate of the count y_{it} for area i at simulation t , with associated MSE estimate mse_{it} . The main performance indicators used in our simulations are the Absolute Relative Bias (ARB) and the Relative Root MSE (RRMSE), both expressed in percentage terms, and defined as follows:

$$ARB_i = \left| \left(T^{-1} \sum_{t=1}^T y_{it} \right)^{-1} \left\{ T^{-1} \sum_{t=1}^T (\hat{y}_{it} - y_{it}) \right\} \right| \times 100,$$

$$RRMSE_i = \left(T^{-1} \sum_{t=1}^T y_{it} \right)^{-1} \left\{ \sqrt{T^{-1} \sum_{t=1}^T (\hat{y}_{it} - y_{it})^2} \right\} \times 100.$$

We also computed these performance indicators for the estimated MSE values associated with each estimated count. In this case we replaced $\hat{y}_{it}$ in the preceding ARB and RRMSE formulae by mse_{it} and replaced the actual count y_{it} by the empirical simulation-based MSE $T^{-1} \sum_{t=1}^T (\hat{y}_{it} - y_{it})^2$.

We simulated data for three sets of small areas corresponding to $m = 49, 100$ and 196 areas and with area-specific sample sizes of either 10 or 20 for all areas. We simulated three scenarios of population counts for these small areas. In the first two we generated these counts under the GLMM (2) and the NSGLMM (5) respectively. In the third, we generated them under the GLMM (2) with a SAR-type area effect. In all three cases the population counts were simulated using a Binomial specification (logistic link function) as well as a Poisson specification (log link function).

The first scenario corresponded to a stationary non-spatial GLMM. For the Binomial case, we generated data for area i under the model: $\eta_i = 1 + x_i + u_i$, where $x_i \sim U[-1, 1]$, with random area effects u_i generated as m independent realizations from a $N(0, \sigma_u^2 = 0.0625)$ distribution. In this case the population count for area i was generated as the sum of N_i independent binary values y_{ij} , each such that $\Pr(y_{ij} = 1) = \mu_i = \exp(\eta_i) \{1 + \exp(\eta_i)\}^{-1}$. Here N_i is the population size of area i . The corresponding sample count was then the sum of the first n_i of these binary values. For Poisson case, we generated data from the model: $\eta_i = -1 + 0.6x_i + u_i$ with x_i and u_i as previously specified, and with the N_i individual counts within the area generated as independent realisations of a $Pois(\mu_i)$ random variable, where $\mu_i = \exp(\eta_i)$. Again, the individual sample counts, that were aggregated to the overall sample count for the area, were defined by the first n_i of these realised counts.

The second scenario corresponded to a spatially non-stationary GLMM. In this case the m areas were assumed to be located on a $\sqrt{m} \times \sqrt{m}$ grid with intersections uniformly spread between -1 and 1 and with a distance $2/(\sqrt{m}-1)$ between any two neighbouring intersections along both the horizontal and vertical axes. For example, when $m = 100$ the area locations d_i are the set of pairs defined by every pair of values from the set $\{-1, -0.77, -0.55, -0.33, -0.11, 0.11, 0.33, 0.55, 0.77, 1\}$. For the Binomial case we generated population counts using $\eta_i(d_i) = 1 + x_i + \gamma_0(d_i) + x_i\gamma_1(d_i) + u_i$, where both x_i and u_i were defined as in scenario one, and where the vector $\boldsymbol{\gamma}(d_i) = (\gamma_0(d_i), \gamma_1(d_i))^T$ was a random draw from $N(\mathbf{0}, \boldsymbol{\Omega} \otimes \lambda \mathbf{I}_2)$, with $\boldsymbol{\Omega}$ equal to the matrix of Euclidean distances between the grid points, and with values of λ set to be either 0.5, 1 or 2. Binomial population and sample counts for each area were then simulated using the same procedure for binomial data described for scenario one. Poisson counts for this scenario were generated in the same way as Poisson counts in scenario one, with the only difference being that in this case we set $\eta_i(d_i) = -1 + 0.6x_i + \gamma_0(d_i) + x_i\gamma_1(d_i) + u_i$.

In the third scenario, counts were generated under a stationary spatial GLMM with SAR-type area effects. That is, effects for neighbouring areas had a SAR-type correlation. For the Binomial case, we generated data under $\eta_i = 1 + x_i + \alpha_i$ where $x_i \sim U[-1, 1]$ as usual but where the vector $\alpha = (\alpha_i)$ of area effects was generated via a random draw from $N(\mathbf{0}, \sigma_u^2 [(\mathbf{I}_m - \rho \mathbf{\Omega}_s)(\mathbf{I}_m - \rho (\mathbf{\Omega}_s)^T)]^{-1})$, with values of ρ set to 0.25, 0.5 or 0.75 and with the proximity matrix $\mathbf{\Omega}_s$ defined using the actual proximities of the $m = 71$ Districts making up the State of Uttar Pradesh in India. Note that this implies that the elements of $\mathbf{\Omega}_s$ were non-zero only for those pairs of districts that are adjacent. Binomial counts for each area were then obtained as in scenario one. For Poisson case, we generated data under $\eta_i = -1 + 0.6x_i + \alpha_i$ with random area effects $\alpha = (\alpha_i)$ generated as in the Binomial case. Poisson counts were then obtained as in scenario one.

We generated $T = 1000$ independent data sets under each of these models specified in the scenarios above and estimated the corresponding small area total counts (or proportions) using the predictors defined in Table 1. In Table 2 we compare the empirical performances of these predictors for both the stationary (scenario one) and the nonstationary (scenario two) cases for Binomial data, while in Table 3 we show similar results for Poisson data. In Table 4 and 5 we show corresponding results for scenario three, i.e. where data are generated under a stationary spatial process. In all cases the values of the performance measures ARB and RRMSE are reported as averages over the areas. The corresponding area-averaged ARB and RRMSE results for the analytical MSE estimator (14) and the bootstrap MSE estimator (15) of the NSEP are reported in Table 6 (Binomial data) and in Table 7 (Poisson data) for both scenario one and scenario two. Note that Tables 6 and 7 only contain results for selected cases since performances for remaining cases were very similar.

Two points emerge from an inspection of Table 2. The first is that all the three small area predictors (EP, NSEP and SEP) are comparable when the data are stationary, with the EP predictor showing marginally smaller values of RRMSE, as one would expect. Second,

when the underlying data are nonstationary it is clear that NSEP performs better than both the EP and the SEP. In particular, in terms of RRMSE the NSEP shows consistently better performance than both the EP and the SEP in the nonstationary case. As one might expect, the RRMSE of the NSEP decreases with increasing number of areas (i.e. $m = 49$ to $m = 196$), increases with increasing intensity of spatial nonstationarity (i.e. $\lambda = 0.5$ to $\lambda = 2.0$) and decreases with increasing area specific sample sizes (i.e. $n_i = 10$ to $n_i = 20$). Furthermore, we see that the SEP dominates the EP for small area specific sample sizes (i.e. $n_i = 10$) but both are comparable for larger area specific sample size (i.e. $n_i = 20$). There is not much difference in performance between the three small area predictors when we look at their average ARB values, however.

[Table 2 about here]

Figure 1 shows the distribution of area-specific values of actual Relative Bias, ARB and RRMSE for EP, NSEP and SEP (all expressed in percentage terms) obtained in simulations where $m = 100$ and $n_i = 20$, with a Binomial response. Here the actual Relative Bias is defined in exactly the same way as ARB, but no absolute value is applied. The plots confirm the results presented in Table 2: the NSEP dominates the other predictors in terms of efficiency as the intensity of spatial nonstationarity increases. The plots show also that, in terms of Relative Bias and ARB, the performance of NSEP is less variable between small areas than those of EP and SEP.

[Figure 1 about here]

The results in Table 3 are for Poisson counts. When the underlying data generation process is spatially stationary, these indicate a similar pattern to that observed with Binomial counts in Table 2. All three small area predictors are again comparable, with the EP predictor showing marginally smaller values of RRMSE. For spatially nonstationary Poisson data, the NSEP again performs better than the EP and the SEP in terms of RRMSE, but the gains are much less marked compared with the Binomial case. Furthermore, we now see that there are no clear trends in RRMSE for all three predictors as the intensity of spatial nonstationarity increases or as m increases, though RRMSE generally decreases with increasing n_i . Furthermore, in contrast with the Binomial case, the EP outperforms the SEP in terms of RRMSE when the underlying Poisson counts are spatially nonstationary.

Again, in terms of ARB we see no real differences in performance of three small area predictors.

[Table 3 about here]

Figure 2 shows the boxplots of area-specific values of actual Relative Bias, ARB and RRMSE for EP, NSEP and SEP (all expressed in percentage terms) obtained in simulations where $m = 100$ and $n_i = 20$, with a Poisson response. Again, the NSEP outperforms EP and SEP in terms of the distribution of RRMSE between areas, whereas there are no noticeable differences in the distribution of the Relative Bias and ARB between the three small area predictors.

[Figure 2 about here]

Generally, the results set out in Tables 2 and 3 support the conclusion that the use of geo-coded information improves estimates for small areas, with the proposed NSEP emerging as the best performing of the methods that we investigated in the nonstationary data case.

A potential issue with the NSEP is its sensitivity to the unit of measurement used to define the matrix Ω of distances between areas. We therefore also investigated the effect of changing the length unit (for example from kilometres to meters) by running a further simulation experiment where we replicated scenarios 1 and 2 but with area locations now defined on two-dimensional grid consisting of $\sqrt{m} \times \sqrt{m}$ points uniformly spread between -1000 and 1000 instead of between -1 and 1. Although the results of ARB and RRMSE for the EP and NSEP reported in Tables 2 and 3 were basically unchanged in this new simulation, we did observe a significant increase in ARRMSE for SEP. As a matter of prudence, however, we suggest that the variables defining location coordinates be generally normalized to lie between zero and one before NSGLMM parameter estimation.

In scenario three the target was to evaluate the performance of NSEP under a spatially stationary process. Our prior expectation was that the three estimators would behave differently under different levels of spatial correlation, but the SEP should dominate unless the level of spatial correlation is low. To a large extent our expectations were confirmed with the SEP more accurate and efficient than the EP and NSEP in both the Binomial and

the Poisson case. However, again as expected, the gain from using the SEP reduced with decreasing spatial correlation.

[Tables 4 and 5 about here]

Tables 6 and 7 present results that allow us to compare the performances of the two MSE estimators of the NSEP, i.e., the analytic MSE estimator (14) and the bootstrap MSE estimator (15) for Binomial data (Table 6) and Poisson data (Table 7). Here we see that the analytic MSE estimator (14) is less biased than the bootstrap MSE estimator (15). The relative bias of the analytic estimator (14) reduces when the number of areas or values of λ (the ‘intensity’ of spatial clustering) increases. However, the bootstrap estimator (15) is slightly more stable.

[Tables 6 and 7 about here]

Figure 3 shows boxplots of area-specific values of actual Relative Bias, ARB and RRMSE for the analytic and bootstrap MSE estimators for the NSEP (all expressed in percentage terms) obtained in simulations where $m = 100$ and $n_i = 10$, with a Binomial response. As expected, the bootstrap MSE estimator (15) is less variable and more biased than the analytic MSE estimator (14).

[Figure 3 about here]

Finally, we present simulation results that provide an insight into the behaviour of the hybrid MPQL/REML parameter estimation algorithm described in Section 4.1. In particular, Tables 8 and 9 show simulation results for parameter estimation using this algorithm under NSGLMM with Binomial and Poisson responses, with $m = 100$ and $n_i = 10, 20$ under scenario 1 and scenario 2 with $\lambda = 1.5$. The MPQL/REML procedure results in small downward biases in the estimation of the regression model parameters β_0 and β_1 , as well as the variance component σ_u , for the Binomial case under both scenarios, with a more substantial negative bias of around 12% when estimating the spatial intensity parameter λ . As expected, both Relative Bias and RRMSE decrease with increasing n_i . For Poisson data the results are similar for β_0 and σ_u but now show a downward bias of around 15% for β_1 under the non-stationary scenario, while estimation of λ continues to

show a small downward bias (9-10%). Overall, we conclude that although parameter estimation via the MPQL/REML algorithm works reasonably well in the stationary case, there is room for improvement in the nonstationary case.

[Tables 8 and 9 about here]

7. Application

Here we use SAE based on the spatially nonstationary GLMM (5) to obtain estimates of the proportions of poor households at District level in the State of Uttar Pradesh in India. We use survey data from the Household Consumer Expenditure Survey 2011-12 of NSSO 68th round and the Indian Population Census 2011, and assume a binomial specification for the observed District level sample proportions. Model specification for this application was discussed in Section 3, and resulted in the identification of three PCA-based covariates, labelled G11, G21 and G31 there.

Figure 4 shows contour maps of the estimated District-specific intercepts and slopes from a GWR fit (Fotheringham *et al.*, 2002) to the sample proportions for the different Districts. These maps support the case for spatial nonstationarity in the NSSO data. In particular, we see that the coefficients for G11, G21 and G31 in the GWR fit vary considerably, ranging from -1.3 to 2.9 for G11, from -1.2 to 2.8 for G21 and from -0.9 to 4.3 for G31. Moreover the contour map of the intercept coefficients also shows considerable spatial variation, ranging from -5.3 (South-East) to 0.2 (Centre-West).

[Figure 4 about here]

Additional GWR-based analyses were carried out in order to determine the scale of the spatial nonstationarity in the local parameter coefficients obtained in the GWR fit. This was done by increasing the fixed kernel bandwidth of the GWR "window" from a 30km to a 125 km radius in 5 km increments. The variation in the regression coefficient for each variable at each location was subsequently used to calculate an approximate index of stationarity (Brunsdon *et al.*, 1998), defined by dividing the interquartile range of the GWR coefficient by twice the standard error of the same variable from the global logistic regression (Windle *et al.*, 2010). Values greater than one for this ratio suggest that the relationship between the independent and dependent variables in the GWR model might be

spatially non-stationary. These index values are plotted against the kernel bandwidth distances in Figure 5, allowing one to examine the spatial scale at which the regression relationships eventually became stationary.

[Figure 5 about here]

Inspection of Figure 5 shows that the G11 and G21 components of the model remain non-stationary at least up to a bandwidth of 125 km, whereas the G31 component becomes stationary beyond approximately 100 km.

The diagnostic procedure described in Section 4.2 for testing for the presence of spatial nonstationarity, that is, the hypothesis $H_0 : \lambda = 0$ versus the one-sided alternative $H_1 : \lambda > 0$, was also applied to the NSSO data. The test statistic value that was generated was highly significant (p-value = 0.00), indicating strong evidence for nonstationarity, with the estimated value of the variance component λ characterising the intensity of this nonstationarity equal to 0.244.

We next applied the NSEP method to produce estimates of the proportions of poor households (Head Count Ratio, or HCR) by District across Uttar Pradesh. Table 10 shows the estimates of the regression coefficients for the global GLMM as well as descriptive statistics for the District-specific estimated parameter coefficients produced under the NSGLMM.

[Table 10 about here]

The G11 and G21 variables had both negative and positive parameter values, although most values are negative. In contrast, all parameter values for the G31 variable are positive. As one would expect, the NSGLMM is an improvement over the GLMM for predicting HCR, with the NSGLMM log-likelihood (2375.5) significantly larger than the corresponding log-likelihood generated by the GLMM (-2970.2). The contour maps shown in Figure 6 support the nonstationarity hypothesis. Here we see that the fitted regression coefficients in the NSGLMM applied to the district level data change substantially over the study space.

[Figure 6 about here]

Figure 7 shows the fitted distributions of the predicted random effects in the NSGLMM applied to the survey data. These include histograms as well as Normal q - q plots of District

residuals as well as the spatial random effects. These indicate that the Gaussian assumption fits reasonably well for this application.

[Figure 7 about here]

Brown *et al.* (2001) discuss diagnostics for SAE, rather than diagnostics for model fit. They note that small area estimates should be (a) consistent with unbiased direct survey estimates, i.e. they should provide an approximation to the direct survey estimates that is consistent with these values being "close" to the expected values of the direct estimates and (b) more precise than direct survey estimates, as evidenced by lower MSE estimates. We therefore consider three commonly used diagnostics developed by these authors for this purpose: the bias diagnostic, the goodness of fit (GOF) statistic and the percent coefficient of variation (CV) diagnostic.

The basic idea underpinning the bias diagnostic is that since direct estimates are unbiased, their regression on the true values should be linear and correspond to the identity line. If model-based small area estimates are close to these true values the regression of the direct estimates on these model-based estimates should be similar. We therefore plot direct estimates (y-axis) vs. model-based small area estimates (x-axis) and we looked for divergence of the fitted least squares regression line from the line of equality. The scatter plot of the District-level direct estimates against the corresponding NSEP estimates is shown in Figure 8, with fitted least squares regression line (dashed line) and line of equality (solid line) superimposed.

[Figure 8 about here]

Inspection of Figure 8 reveals that the District estimates generated by NSEP are less extreme than the direct survey estimates, demonstrating the typical SAE outcome of shrinkage towards the average. However, it is also clear that the NSEP estimates deviate somewhat from the line of equality at the extremes of their distribution. This is not unexpected, since the NSEP estimates are random variables and so the regression of the direct estimates on the NSEP estimates is biased for a test of common expected values. Such a test is provided by the GOF statistic, which is equivalent to a Wald test for whether the differences $D_i = \hat{y}_i^{direct} - \hat{y}_i^{Nssyn}$ have a zero mean, and is computed as

$$W = \sum_i \left\{ \frac{D_i^2}{\text{var}(\hat{y}_i^{direct}) + \text{mse}(\hat{y}_i^{NSSYN})} \right\}$$

where $\hat{y}_i^{NSSYN}$ is the synthetic version NSSYN of the NSEP, defined following (8). See Brown *et al.* (2001). Under the assumption that $\hat{y}_i^{direct}$ and $\hat{y}_i^{NSSYN}$ are independently distributed, which is not unreasonable for large sample sizes, the value of W can be compared with an appropriate critical value from a chi square distribution with degrees of freedom D equal to the number of Districts. For our analysis, $D = 71$, with a critical value of 91.7 at a 5% level of significance. In contrast, when we calculate our MSE estimate for $\hat{y}_i^{NSSYN}$ using the bootstrap method described in Section 5.2, we obtain $W = 35.9$, and so conclude that the NSEP estimates are consistent with the direct estimates. Note that the GOF statistic can be calculated substituting the NSEP estimator for the NSSYN estimator, in which case $W = 29.7$. However, this is not recommended since the inclusion of an estimated area effect in the NSEP implies a potential correlation with the direct estimator.

The final SAE diagnostic that we compute compares the extent to which the model-based small area estimates (EP, SEP and NSEP) improve in precision compared to the direct survey estimates. Recollect that, as noted earlier, the proximity matrix used to construct the SEP under SGLMM for this application is a simple row-standardized contiguity matrix, with elements that take non-zero values only for those pairs of Districts (or areas) that are adjacent. We assess precision via the percentage CV of an estimate $\hat{y}_i$, defined as $CV(\hat{y}_i) = (SE(\hat{y}_i) / \hat{y}_i) \times 100$. Noting that substitution of model-based analytic estimators of RRMSE for estimated standard errors in this diagnostic is problematic since these will all depend on different model assumptions, we use a common NSGLMM-based parametric bootstrap (see Section 5.2) for this purpose. Summary percentage CV values for the different SAE methods are shown in Table 11, while Figure 9 displays the District level values of percentage CV for the direct, EP, SEP and NSEP estimation methods. These show that the direct survey estimators for the proportion of poor households within each District are unstable, with CVs varying from 8.67% to 232.44 %. Furthermore, the CVs of the direct estimators are greater than 20% (40%) in 36 (16) out of the 71 Districts (see Figure 6). All

three small area estimators (EP, SEP and NSEP) are better than the direct estimator with respect to CV, with the NSEP the best performing method. In particular, the NSEP records smaller CVs than the EP in 53 of the 71 Districts of Uttar Pradesh.

[Table 11 about here]

[Figure 9 about here]

Figure 10 is a map showing the proportions of poor households by District for Uttar Pradesh, estimated by the NSEP method. This clearly shows an east-west divide in the distribution of household poverty. For example, in the western part of Uttar Pradesh there are many districts with low poverty incidence (Saharanpur, Hathras, Meerut, Baghpat, Muzaffarnagar, Bulandshahar, etc). Similarly, in the eastern part and in the Bundelkhand region (north-east) we see many districts (Azamgarh, Sitapur, Chitrakoot, Bahraich, Siddharthnagar, Banda, Fatehpur, Basti and Kaushambi, etc) with high poverty incidence. This is an example of a "poverty map" showing reliable estimates of poverty incidence across a region of interest. This type of map is a useful aid for policy planners and administrators charged with taking effective financial and administrative decisions that can impact differentially across the region.

[Figure 10 about here]

8. Conclusions

This paper describes how SAE can be carried out under an extension of the area level version of the generalised linear mixed model that accounts for the presence of spatial nonstationarity in the data. Extensive simulations then provide evidence that for Binomial and Poisson counts, the proposed plug-in nonstationary empirical predictor (NSEP) is more efficient than the usual plug-in empirical predictor (EP) where there evidence of spatial nonstationarity in the data. The approach extends easily to the non-sampled case, i.e. where the area of interest has no sample, and therefore no direct estimate. Synthetic estimation is the standard approach here, and we develop a synthetic version NSSYN of the NSEP. We also investigate area level SAE for counts that can be modelled via generalised linear mixed models with spatial random area effects. In particular, we investigate the case where these random effects can be characterised via a simultaneous autoregression or SAR model (Pratesi and Salvati, 2008; Chandra and Salvati, 2017; Saei and Chambers, 2003). We refer

to the plug-in empirical predictor for this case as the SEP and use simulation to compare its performance with the EP and the NSEP. Our results, and particularly those of the NSEP, are promising, indicating that extensions, particularly to nonstationary multi-category counts, are worth investigating.

We also developed two MSE estimation methods for the NSEP, an analytical method based on a linearisation argument and a parametric bootstrap method. In our simulations, the analytical MSE estimator showed smaller bias but the bootstrap MSE estimator was more stable. Finally, we used the NSEP methodology to produce a poverty map showing how household poverty incidence varies by District across the State of Uttar Pradesh in India. We conclude by observing that the estimates and spatial distribution of poverty incidence generated from this research should be useful for meeting the data requirements for policy research and strategic planning by different international organizations and by Departments and Ministries in the Government of India.

Acknowledgment

The authors would like to acknowledge the valuable comments and suggestions of the Editor, Associate Editor and an anonymous referee. These led to a considerable improvement in the paper. The work of Hukum Chandra was done under ICAR-National Fellow Scheme at ICAR-IASRI, New Delhi, India. The work of Salvati was carried out with the support of project PRIN-SURWEY [http://www. sp.unipg.it/surwey/](http://www.sp.unipg.it/surwey/) (Grant 2012F42NS8, Italy).

References

- Anselin, L., 1992. Spatial Econometrics. Methods and Models. Kluwer Academic, Boston.
- Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* 43, 1-59.
- Brown, G., Chambers, R., Heady, P., Heasman, D, 2001. Evaluation of small area estimation methods - an application to the unemployment estimates from the UK LFS. In. *Proc. Statistics Canada Symp. Achieving Data Quality in a Statistical Agency: A Methodological Perspective*. Statistics Canada.

- Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis* 28, 281-298.
- Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1998. Geographically weighted regression - modelling spatial nonstationarity. *The Statistician* 47, 431-443.
- Boubeta, M., Lombardía, M.J., Morales D., 2016. Empirical best prediction under area-level Poisson mixed models. *TEST*, 25, 548-569.
- Cantoni, E., Ronchetti, E., 2001. Robust inference for generalized linear models. *J. Amer. Statist. Assoc.* 96, 1022--1030.
- Chambers, R., Chandra, H., Salvati, N., Tzavidis, N., 2014. Outlier robust small area estimation. *J. Roy. Statist. Soc. Ser. B* 76 (1), 47-69.
- Chambers, R., Salvati, N., Tzavidis, N., 2016. Semiparametric small area estimation for binary outcomes with application to unemployment estimation for local authorities in the UK. *J. Roy. Statist. Soc. Ser. A* 179 (2), 453-479.
- Chandra, H., Salvati, N., 2017. Small Area Estimation for Count Data Under a Spatial Dependent Aggregated Level Random Effects Model. *Comm. Stat. - Theory and Meth.* under revision.
- Chandra, H., Salvati, N., Chambers, R., 2015. A Spatially Nonstationary Fay-Herriot Model for Small Area Estimation. *J. Surv. Stat. Methodol.* 3, 109-135.
- Chandra, H., Salvati, N., Sud, U.C., 2011. Disaggregate-level Estimates of indebtedness in the state of Uttar Pradesh in India-an application of small area estimation technique. *J. Appl. Statist.* 38(11), 2413-2432.
- Clayton, D., Kaldor, J., 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43, 671-681.
- Cressie, N., 1993. *Statistics for Spatial Data*. Wiley, New York.
- Das, K., Jiang, J., Rao, J.N.K., 2004. Mean squared error of empirical predictor. *Ann. Statist.* 32, 818-840.
- Fay, R. E., Herriot, R. A., 1979. Estimation of income from small places: an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.* 74, 269-277.
- Fotheringham, A.S., Brunsdon, C., Charlton, M.E., 2002. *Geographically Weighted Regression*. John Wiley & Sons.

- Hall, P., Maiti, T., 2006. On parametric bootstrap methods for small-area prediction. *J. Roy. Statist. Soc. Ser. B* 68, 221-238.
- Henderson, C. R., 1963. Selection index and expected genetic advance, in *Statistical Genetics and Plant Breeding*, publication 982, Washington DC: National Academy of Science, National research council, 141-163.
- Jiang, J., 2003. Empirical method of moments and its applications. *J. Stat. Plan. Inference* 115, 69–84.
- Johnson, F.A., Chandra, H., Brown, J., Padmadas, S., 2010. Estimating district-level births attended by skilled attendants in Ghana using demographic health survey and census data: an application of small area estimation technique. *J. Off. Stat.* 26 (2), 341–359.
- Manteiga, G.W., Lombardía, M.J., Molina, I., Morales, D., Santamaría, L., 2007. Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Comput. Stat. & Data Anal.* 51, 2720-2733.
- Marhuenda, Y., Molina, I., Morales, D., 2013. Small area estimation with spatio-temporal Fay–Herriot models. *Comput. Stat. & Data Anal.* 58, 308–325.
- McGilchrist, C.E., 1994. Estimation in generalized mixed models. *J. Roy. Statist. Soc. Ser. B* 56, 61-69.
- Molina, I., Salvati, N., Pratesi, M., 2009. Bootstrap for estimating the MSE of the spatial EBLUP. *Comput. Stat.* 24, 441–458.
- Opsomer, J.D., Claeskens, G., Ranalli, M.G., Kauermann, G., Breidt, F.J., 2008. Nonparametric small area estimation using penalized spline regression. *J. Roy. Statist. Soc. Ser. B.* 70, 265-286.
- Porter, A.T., Holan, S.H., Wikle, C.K., Cressie, N., 2014. Spatial Fay–Herriot models for small area estimation with functional covariates. *Spat. Stat.* 10, 27-42.
- Prasad, N.G.N., Rao, J.N.K., 1990. The estimation of the mean squared error of small area estimators. *J. Amer. Statist. Assoc.* 85, 163-171.
- Pratesi, M., Salvati, N., 2008. Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Stat. Methods Appl.* 17, 114-131.
- Pratesi, M., Salvati, N., 2009. Small area estimation in the presence of correlated random area effects. *Journal of Official Statistics*, 25, 37-53.
- Rao, J.N.K., 2003. *Small Area Estimation*. John Wiley and Sons.

- Rao, J.N.K., Molina, I., 2015. Small Area Estimation, 2nd Edition. John Wiley and Sons.
- Saei, A., Chambers, R., 2003. Small area estimation under linear and generalized linear mixed models with time and area effects, Meth. W.P. No. M03/15, Southampton Statistical Sciences Research Institute, University of Southampton, UK.
- Saei, A., McGilchrist, C., 1998. Longitudinal threshold models with random components. The Statistician Ser. D, 47, 365-375.
- Schall, R., 1991. Estimation in generalized linear models with random effects. Biometrika, 78, 719-727.
- Singh, B., Shukla, G., Kundu, D. 2005. Spatio-temporal models in small area estimation. Survey Methodology, 31, 183-195.
- Windle, M.J.S., Rose, G.A., Devillers, R., Fortin, M., 2010. Exploring spatial nonstationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. ICES Journal of Marine Science, 67, 145-154.

Table 1. Description of the small area predictors considered in the simulation studies.

Predictor	Description
EP	Predictor (4) under model (2)
NSEP	Predictor (8) under model (5)
SEP	Spatial EP under SGLMM

Table 2. Values of percentage absolute relative bias (ARB, %) and percentage relative RMSE (RRMSE, %) from model-based simulations for Binomial data. All values are averages over the m areas.

		ARB,%				RRMSE,%			
# Areas	Predictor	Stationary	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$	Stationary	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$
$n_i = 10$									
$m=49$	EP	0.193	0.373	0.404	0.463	7.42	14.96	18.72	20.56
	NSEP	0.183	0.285	0.247	0.297	7.81	12.29	13.66	14.79
	SEP	0.175	0.336	0.318	0.371	7.82	13.96	15.71	16.30
$m=100$	EP	0.186	0.402	0.531	0.549	6.92	14.80	18.77	20.50
	NSEP	0.191	0.320	0.380	0.382	7.10	11.39	12.80	14.04
	SEP	0.187	0.397	0.407	0.382	7.26	13.56	15.15	15.71
$m=196$	EP	0.172	0.370	0.502	0.609	6.68	14.56	18.74	20.09
	NSEP	0.173	0.272	0.300	0.313	6.77	10.60	11.90	13.10
	SEP	0.179	0.350	0.373	0.382	7.03	13.07	14.53	15.40
$n_i = 20$									
$m=49$	EP	0.166	0.309	0.281	0.289	6.59	11.70	12.12	12.09
	NSEP	0.154	0.225	0.203	0.243	6.78	10.07	10.87	11.42
	SEP	0.158	0.290	0.258	0.294	6.78	11.72	12.64	12.49
$m=100$	EP	0.164	0.309	0.340	0.289	6.33	11.60	11.82	12.02
	NSEP	0.165	0.263	0.299	0.286	6.42	9.51	10.38	11.02
	SEP	0.163	0.367	0.357	0.300	6.53	11.93	12.35	12.04
$m=196$	EP	0.156	0.324	0.331	0.309	6.21	11.42	11.60	11.87
	NSEP	0.160	0.222	0.249	0.250	6.26	8.98	9.80	10.48
	SEP	0.160	0.328	0.300	0.300	6.41	11.59	11.63	11.71

Table 3. Values of percentage absolute relative bias (ARB, %) and percentage relative RMSE (ARRMSE, %) from model-based simulations for Poisson data. All values are averages over the m areas.

		ARB,%				RRMSE,%			
# Areas	Predictor	Stationary	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$	Stationary	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$
$n_i = 10$									
$m=49$	EP	0.556	0.929	0.873	1.054	23.36	33.43	31.33	29.61
	NSEP	0.564	0.863	0.876	0.955	23.83	31.59	30.33	29.09
	SEP	0.572	1.169	0.991	1.097	23.89	35.07	32.60	30.19
$m=100$	EP	0.572	0.913	0.853	0.825	22.59	32.77	31.20	30.56
	NSEP	0.567	0.885	0.828	0.783	22.82	30.31	29.67	29.61
	SEP	0.589	1.074	0.966	0.846	23.04	36.08	33.13	31.41
$m=196$	EP	0.521	0.819	0.806	0.892	22.38	32.65	31.63	31.65
	NSEP	0.527	0.781	0.785	0.877	22.46	29.41	29.46	30.22
	SEP	0.534	0.948	0.899	0.917	22.83	35.81	33.29	32.17
$n_i = 20$									
$m=49$	EP	0.541	0.623	0.529	0.620	20.34	24.82	22.81	21.68
	NSEP	0.550	0.561	0.502	0.583	20.57	24.11	22.40	21.46
	SEP	0.538	0.784	0.618	0.666	20.68	25.98	23.45	22.01
$m=100$	EP	0.537	0.661	0.609	0.662	19.94	24.38	22.77	22.29
	NSEP	0.537	0.665	0.624	0.684	20.07	23.27	22.11	21.88
	SEP	0.552	0.742	0.650	0.662	20.16	26.31	23.76	22.77
$m=196$	EP	0.455	0.650	0.607	0.667	19.78	24.32	23.22	23.22
	NSEP	0.455	0.614	0.574	0.663	19.84	22.85	22.21	22.56
	SEP	0.456	0.703	0.646	0.666	19.95	26.26	24.17	23.57

Table 4. Values of percentage absolute relative bias (ARB, %) and percentage relative RMSE (RRMSE, %) from model-based simulations for Binomial data. All values are averages over the m areas. Data generated under SAR model.

	ARB,%			RRMSE,%		
$n_i = 10$						
Predictor	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
EP	0.351	0.772	0.648	9.66	15.23	14.03
NSEP	0.279	0.592	0.551	9.64	14.28	13.55
SEP	0.261	0.298	0.363	11.30	12.96	12.68
$n_i = 20$						
EP	0.270	0.361	0.383	8.57	11.42	10.98
NSEP	0.240	0.312	0.373	8.40	11.38	10.97
SEP	0.212	0.230	0.275	8.73	10.13	9.88

Table 5. Values of percentage absolute relative bias (ARB, %) and percentage relative RMSE (RRMSE, %) from model based simulations for Poisson data. All values are averages over the m areas. Data generated under SAR model.

ARB,%				RRMSE,%		
$n_i = 10$						
Predictor	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$	$\rho = 0.25$	$\rho = 0.50$	$\rho = 0.75$
EP	1.504	1.889	2.019	32.71	38.25	38.18
NSEP	1.195	1.761	1.970	32.02	39.07	38.37
SEP	0.954	1.101	1.368	32.94	36.36	35.90
$n_i = 20$						
EP	0.877	1.075	0.383	26.61	28.91	28.66
NSEP	0.776	1.060	0.373	26.43	29.70	28.93
SEP	0.603	0.832	0.275	25.85	27.73	27.47

Table 6. Values of percentage absolute relative bias (ARB,%) and percentage relative RMSE (RRMSE,%) of MSE estimator for NSEP from model based simulations for Binomial data and $n_i = 10$. All values are averages over the m areas.

		ARB,%			RRMSE,%		
# Areas	MSE Estimator	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$
$m=49$	Analytic	5.26	5.85	3.93	0.75	0.60	0.65
	Bootstrap	4.75	5.95	6.08	0.38	0.35	0.32
$m=100$	Analytic	3.32	2.11	1.85	0.54	0.55	0.60
	Bootstrap	5.01	5.60	6.05	0.38	0.35	0.32

Table 7. Values of percentage absolute relative bias (ARB,%) and percentage relative RMSE (RRMSE,%) of MSE estimator for NSEP from model based simulations for Poisson data and $n_i = 10$. All values are averages over the m areas.

		ARB,%			RRMSE,%		
# Areas	MSE Estimator	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$	$\lambda=0.5$	$\lambda=1.0$	$\lambda=2.0$
$m=49$	Analytic	3.15	3.83	5.96	1.19	1.57	1.99
	Bootstrap	12.29	13.98	11.26	1.07	1.13	1.03
$m=100$	Analytic	3.50	3.70	4.39	1.26	1.64	2.02
	Bootstrap	13.67	9.51	7.56	1.13	1.21	1.50

Table 8. Simulation results for parameter estimation using hybrid MPQL/REML algorithm under NSGLMM - Binomial response (1000 Monte Carlo simulations).

	β_o	β_1	σ_u	λ
$m = 100, n_i = 10$				
True Value	1.000	1.000	0.250	0.000
Average Estimated Value	0.992	0.994	0.245	0.000
%Rel Bias	-0.850	-0.688	-2.020	NA
%RRMSE	7.969	13.969	2.158	NA
True Value	1.000	1.000	0.250	1.500
Average Estimated Value	0.950	0.930	0.241	1.314
%Rel Bias	-4.981	-6.932	-3.295	-12.921
%RRMSE	76.381	81.421	7.747	25.869
$m = 100, n_i = 20$				
True Value	1.000	1.000	0.250	0.000
Average Estimated Value	0.991	0.992	0.245	0.000
%Rel Bias	-0.892	-0.787	-2.006	NA
%RRMSE	5.919	10.206	2.408	NA
True Value	1.000	1.000	0.250	1.500
Average Estimated Value	0.967	0.944	0.242	1.372
%Rel Bias	-3.247	-5.590	-2.932	-9.198
%RRMSE	77.127	81.123	6.362	22.118

Table 9. Simulation results for parameter estimation using hybrid MPQL/REML algorithm under NSGLMM - Poisson response (1000 Monte Carlo simulations).

	β_o	β_1	σ_u	λ
$m = 100, n_i = 10$				
True Value	-1.000	0.600	0.250	0.000
Average Estimated Value	-0.978	0.598	0.244	0.001
%Rel Bias	2.144	-0.216	-2.289	NA
%RRMSE	6.362	16.647	3.760	NA
True Value	-1.000	0.600	0.250	1.500
Average Estimated Value	-0.985	0.513	0.244	1.354
%Rel Bias	-1.483	-14.366	-2.246	-9.787
%RRMSE	71.154	131.377	2.440	25.348
$m = 100, n_i = 20$				
True Value	-1.000	0.600	0.250	0.000
Average Estimated Value	-0.983	0.599	0.245	0.001
%Rel Bias	-1.866	-0.299	-2.322	NA
%RRMSE	5.078	13.255	4.562	NA
True Value	-1.000	0.600	0.250	1.500
Average Estimated Value	-1.021	0.508	0.244	1.371
%Rel Bias	-2.861	-15.451	-2.198	-8.062
%RRMSE	71.563	129.295	2.424	21.794

Table 10. Summary statistics for the NSGLMM parameter estimates, including the overall percentages of negative (% –) and positive (% +) values and parameter estimates for the global GLMM.

	Intercept	G11	G21	G31
Descriptive Statistics	NSGLMM			
Minimum	-2.01	-0.33	-0.54	0.00
Q1	-1.67	-0.21	-0.16	0.10
Mean	-1.32	-0.12	-0.09	0.17
Median	-1.26	-0.13	-0.09	0.16
Q3	-1.01	-0.04	-0.02	0.19
Maximum	-0.70	0.14	0.21	0.53
% –	100	90.2	77.5	0
% +	0	9.8	22.5	100
	GLMM			
GLMM Model	-1.35	-0.21	-0.10	0.36

Table 11. Summary of area distributions of percentage coefficients of variation (CV, %) for different SAE methods applied to NSSO data. Note that all CVs were calculated using the Parametric Bootstrap MSE estimation procedure described in Section 5.2.

Values	Direct	SEP	EP	NSEP
Minimum	8.67	9.28	6.72	6.54
Q1	13.94	14.04	12.70	11.42
Mean	32.61	22.50	21.58	20.67
Median	19.97	18.68	17.93	15.41
Q3	33.32	29.10	28.05	26.04
Maximum	232.40	56.47	57.38	70.02

Figure 1. Boxplots of area-specific values of actual Relative Bias, ARB and RRMSE for EP and NSEP (all expressed in percentage terms) obtained in simulations where $m = 100$, $n_i = 20$, and with a Binomial response. Note that actual Relative Bias is defined in exactly the same way as ARB, but no absolute value is applied.

Figure 2. Boxplots of area-specific values of actual Relative Bias, ARB and RRMSE for EP and NSEP (all expressed in percentage terms) obtained in simulations where $m = 100$, $n_i = 20$, and with a Poisson response. Note that actual Relative Bias is defined in exactly the same way as ARB, but no absolute value is applied.

Figure 3. Boxplots of area-specific values of actual Relative Bias, ARB and RRMSE for Analytic and Bootstrap MSE estimators for the NSEP (all expressed in percentage terms) obtained in simulations where $m = 100$, $n_i = 10$, and with a Binomial response. Note that actual Relative Bias is defined in exactly the same way as ARB, but no absolute value is applied.

Figure 4. Maps showing the spatial variation in the District specific intercept and slope estimates that are generated when the GWR model is fitted to the NSSO data.

Figure 5. Index of spatial stationarity for variables used in the GWR models for the HCR in the State of Uttar Pradesh in India. The index is calculated by dividing the interquartile range of a GWR regression coefficient by twice the standard error of the same parameter estimate from the global model. Spatial nonstationarity is suggested by index values greater than 1.

Figure 6. Maps showing the spatial variation of estimated regression coefficients $\beta(d_i) = \beta + \gamma(d_i)$ when NSGLMM model is fitted to the NSSO data.

Figure 7. Histograms (left side) and normal q-q plots (right side) of the estimated District-level random effects and the coefficient-specific estimated random effects associated with spatial nonstationarity for the NSSO data.

Figure 8 Bias diagnostic plot with $y = x$ line (solid) and regression line (dotted) for proportion of poor households in Uttar Pradesh: NSEP estimates versus direct estimates.

Figure 9. Percentage coefficients of variation (CV, %) by District for direct (dotted line, o) , SEP (dash line, *) , EP (thin line, ∇) and NSEP(solid line, $\bullet$) estimators applied to NSSO data.

Figure 10. NSEP estimates showing the spatial distribution of proportions of poor households by District in Uttar Pradesh.