Small Area Estimation under a Spatially Non-Linear Model

Hukum Chandra[1], Nicola Salvati[2] and Ray Chambers[3]

Abstract

We describe a methodology for small area estimation of counts that assumes an area-level version of a nonparametric generalized linear mixed model with a mean structure defined using spatial splines. The proposed method represents an alternative to other small area estimation methods based on area level spatial models that are designed for both spatially stationary and spatially non-stationary populations. We develop an estimator for the mean squared error of the proposed small area predictor as well as an approach for testing for the presence of spatial structure in the data and evaluate both the proposed small area predictor and its mean squared error estimator via simulations studies. Our empirical results show that when data are spatially non-stationary the proposed small area predictor outperforms other area level estimators in common use and that the proposed mean squared error estimator tracks the actual mean squared error reasonably well, with confidence intervals based on it achieving close to nominal coverage. An application to poverty estimation using household consumer expenditure survey data from 2011-12 collected by the national sample survey office of India is presented.

Key words: Small area estimation, Nonparametric models, Spatial relationship, Count data, Poverty indicator.

[1] Corresponding Author: Indian Agricultural Statistics Research Institute, Library Avenue, PUSA, New Delhi-110012, India, Phone: +91-11-25843398, Fax: +91-11-25841564, Email: hchandra12@gmail.com
[2] Dipartimento di Economia e Management, University of Pisa, Italy, E-mail: nicola.salvati@unipi.it
[3] Centre for Statistical and Survey Methodology, University of Wollongong, Wollongong, NSW, 2522, Australia. Email: ray@uow.edu.au

1. Introduction

The demand for small area statistics has increased rapidly over the past few years (e.g. measurement of social exclusion and social wellbeing at disaggregate level, see Tzavidis et al. 2008). As a consequence, many small area estimation (SAE) methods based on linear mixed models have been proposed in the literature. In many cases, however, the response variable is not continuously distributed but is binary valued or a count. Such response variables cannot be modelled using standard linear models. When the variable of interest is binary or a count and small area estimates are required for these data, use of standard estimation methods based on linear mixed models becomes problematic. For example, poverty indicators and many other indicators related to socio-economic status and food insecurity usually behave in a non-Gaussian manner at small area levels, and so estimation in these cases is typically based on a generalized linear mixed model (GLMM); see Manteiga *et al.* (2007) and Ruppert e*t al.* (2003, chapter 10). The most commonly used GLMMs are the logistic-normal mixed model (i.e. GLMMs with logistic link function, also referred as the logistic linear mixed model) and the general Poisson-normal mixed model (i.e. GLMMs with log link function, also referred as the log linear mixed model). Unit level predictions generated by a GLMM are generally used to define the empirical predictor for small areas for such data. In many applications this is not possible, for example poverty mapping where data confidentiality restricts access to unit level survey data with small area identifiers, or where the agency carrying out the small area analysis does not have the resources to analyse unit level data, as in many developing countries. In such situations, an area level version of the GLMM can be used for SAE. In particular, when only area level data are available, an area level version of the GLMM is fitted to obtain the plug-in empirical predictor for the small areas, see for example, Chandra e*t al.* (2011) and Johnson *et al.* (2010). Other recent work on this topic include Boubeta *et al.* (2016, 2017) who use area-level Poisson mixed models for estimating small area counting indicators. Other authors have developed SAE under a GLMM using a Bayesian approach. See Torabi and Shokoohi (2015), Rao and Molina (2015), Moura *et al.* (2006), Datta *et al.* (1999), and references therein. Mercer *et al.* (2014), Liu *et al.* (2014) and Franco and Bell (2013) consider the use of survey weights in a Bayesian hierarchical model framework when estimating small area proportions. In this context, we note that the hierarchical Bayes (HB) approach to SAE offers considerable promise because of it can accommodate complex small area models and provide "exact" inferences. Unfortunately,

however, the choice of noninformative priors that can provide frequentist validity for the Bayesian approach may not be easy in practice, especially when complex sampling designs are involved. Also, caution needs to be exercised in the routine use of popular HB model-checking methods, see Rao (2011). In this paper we focus on the situation where only aggregated level data are available and SAE is carried out under area level small area models. Our development is based on a frequentist approach to SAE. This approach is often easier to explain to practitioners, can be less time consuming and inferential expressions can typically be written out explicitly. For this reason, national statistical offices as well as other Government agencies involved in production of statistics often prefer a frequentist approach to estimation and prediction over a Bayesian approach.

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. It is therefore reasonable to assume that the effects of neighbouring areas, defined via a contiguity criterion, are correlated. Chandra and Salvati (2018) and Saei and Chambers (2003) describe an extension of the area level version of GLMM that allows for spatially correlated random effects using a SAR model (SGLMM) and define a plug-in 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. Chandra *et al.* (2017) introduce a spatially nonstationary extension of the area level version of GLMM, using an adaptation of the geographical weighted regression (GWR) concept to extend the GLMM to incorporate spatial nonstationarity (NSGLMM), which they then apply to the SAE problem to define a plug-in empirical predictor (NSEP) for small areas. Non-stationary spatial effects can be also modelled using a spatially non-linear extension of the GLMM. In the GLMM, the relationship between the link function and the covariates is often assumed to be linear. However, when the functional form of the relationship between the link function and the covariates is unknown or has a complicated functional form, an approach based on the use of a non-linear regression model can offer significant advantages compared with one based on a linear model. Torabi and Shokoohi (2015) describe a data cloning approach to fitting a GLMM that uses this idea, but which is based on a unit level GLMM. When geographically referenced area-level responses play a central role in the analysis and need to be converted to maps, we can use bivariate smoothing to

fit a spatially heterogeneous GLMM. In particular, we use P-splines that rely on a set of bivariate basis functions to handle the spatial structures in the data, while at the same time including small area random effects in the model. We denote this nonparametric P-spline-based extension of the usual GLMM by SNLGLMM. See Ugarte *et al.* (2009), Opsomer *et al*. (2008) and Ruppert *et al.* (2003). We then describe a non-linear version of the plug-in empirical predictor for small areas (SNLEP) under an area level version of SNLGLMM. We also develop mean squared error estimation for SNLEP using the approach discussed in Chandra e*t al.* (2011), Johnson *et al.* (2010), Opsomer *et al*. (2008) and Saei and Chambers (2003).

Note that an alternative to computing the plug-in empirical predictor (EP) is to compute the empirical best predictor (EBP, Jiang, 2003). Unfortunately, computing the EBP is generally not straightforward since it does not have a closed form and usually has to be computed via numerical approximation. As a consequence, national statistical agencies tend to favour computation of an analytic approximation such as the EP. It is our understanding that an approximation closely related to the EP is also used in Lopez-Vizcaino *et al*. (2013, 2015) and Molina *et al*. (2007).

The rest of this article is organized as follows. Section 2 introduces the area level version of GLMM to define the plug-in empirical predictor for small areas and reviews the SGLMM and NSGLMM and its corresponding plug-in empirical predictors (SEP and NSEP). In Section 3 we describe the spatially non-linear extension of an area level version of GLMM (i.e. the SNLGLMM) and subsequently use this model to carry out SAE. We focus on the GLMM with logistic link function (i.e. logistic-normal mixed model) for binary data and the GLMM with log link function (i.e. general Poisson-normal mixed model) for count data. The development reported in this paper can be easily generalized to other variants of GLMMs. Section 4 then discusses mean squared error estimation for the proposed small area predictor, and develops a corresponding analytic estimator. Empirical results are provided in Section 5 and application of the proposed small area method to poverty mapping is described in Section 6. Finally, Section 7 summarizes our main conclusions and identifies areas where further research is necessary.

2. SAE under a generalized linear mixed model

Consider a finite population *U* of size *N*, and assume that a sample *s* of size *n* is drawn from this population according to a given sampling design, with the subscripts *s* and *r* used to denote quantities related to the sampled and non-sampled parts of the population. We assume that population is made up of *m* small domains or small areas (or simply domains or areas) $U_i (i = 1,...,m)$, where we use a subscript of *i* to index those quantities associated with area i. In particular, n_i and N_i are used to represent the sample and population sizes in area *i*, respectively. We also assume that the underlying unit level variable of interest *y* is discrete, and in particular is either a binary value or is a non-negative integer, and the aim is to estimate the corresponding small area population proportions or population totals (i.e. counts). Let the total of *y* in area *i* be denoted y_i, and let y_{si} and y_{ri} denote the corresponding sample and non-sample counts for area *i* respectively. We shall assume that area level auxiliary information from secondary data sources, e.g., Census and Administrative records, is available. Let $\mathbf{x}_i$ be the *p*-vector of these covariates for area *i* from these sources. The area level version of the GLMM is then defined as $\Pr(y_i \mid \mathbf{x}_i) \propto \pi_i$, where

$$g(\pi_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i \tag{1}$$

where $g(\cdot)$ is a known function, called the link function, $\pi_i = g^{-1}(\eta_i)$, $\boldsymbol{\beta}$ is the *p*-vector of regression coefficients, often referred to as the fixed effect parameter of the GLMM, and u_i is an area-specific random effect that accounts for between area differences beyond that explained by differences in the auxiliary variables included in the fixed part of the model. We assume that these area effects are independently and identically distributed as normal with mean zero and variance σ_u^2. The model (1) can be used to relate the area level direct survey estimates to area level covariates. This type of model is often referred to as 'area-level' model in SAE, see for example, Rao (2003) and Fay and Herriot (1979). Note that estimation of the fixed effect parameter $\boldsymbol{\beta}$ and the area specific random effects u_i as well as the variance component parameter σ_u^2 uses data from all areas. Collecting the area level models (1), we can write the model (1) as

$$g(\boldsymbol{\pi}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{2}$$

where $\boldsymbol{\pi} = (\pi_1, ..., \pi_m)^T$, $\mathbf{X}=(\mathbf{x}_1^T, ..., \mathbf{x}_m^T)^T$ is a $m \times p$ matrix and $\mathbf{u} = (u_1, ..., u_m)^T$ is a vector of $m \times 1$ of area random effects which is normally distributed with mean zero and variance $\boldsymbol{\Sigma}_u = \sigma_u^2 \mathbf{I}_m$. Here, $\mathbf{I}_m$ is an identity matrix of order m.

When the variable of interest *y* is binary, and unit level values in area *i* are independently and identically distributed, the sample counts y_{si} in area *i*, has a Binomial distribution with parameters n_i and π_i, denoted by $y_{si} \sim \text{Binomial}(n_i, \pi_i)$, where π_i is now the probability of occurrence of an event or probability of prevalence in area *i*, often referred to as the probability of a 'success'. Similarly, the non-sample count y_{ri} in area *i* is such that $y_{ri} \sim \text{Binomial}(N_i - n_i, \pi_i)$. That is, the counts y_{si} and y_{ri} are independent Binomial variables with π_i then corresponding to a common success probability. In this case, the link function $g(\cdot)$ is usually taken to be the logit of the probability π_i. The model (1) linking π_i with the covariates $\mathbf{x}_i$ is then the GLMM with logistic link function given by

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

with $\pi_i = \exp(\eta_i)\{1+\exp(\eta_i)\}^{-1} = \text{expit}(\eta_i) = \text{expit}(\mathbf{x}_i^T\boldsymbol{\beta} + u_i)$ and $u_i \sim N(0, \sigma_u^2)$. Here, $y_{is} | u_i \sim \text{Binomial}\left(n_i, \text{expit}(\mathbf{x}_i^T\boldsymbol{\beta} + u_i)\right)$ and $y_{ir} | u_i \sim \text{Binomial}\left(N_i - n_i, \text{expit}(\mathbf{x}_i^T\boldsymbol{\beta} + u_i)\right)$. The expected values of y_{si} and y_{ri} given u_i are then

$$\mu_{si} = E\left(y_{si} | u_i\right) = n_i \,\text{expit}(\mathbf{x}_i^T\boldsymbol{\beta} + u_i) \text{ and } \mu_{ri} = E\left(y_{ri} | u_i\right) = \left(N_i - n_i\right)\text{expit}(\mathbf{x}_i^T\boldsymbol{\beta} + u_i).$$

Similarly, when the variable of interest *y* is a count, a similar argument leads to the conclusion that *y* follows a general Poisson-normal mixed model, in which case the function $g(\cdot)$ is usually taken to be the log link function and model (1) takes the form

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

with $\pi_i = \exp(\eta_i) = \exp(\mathbf{x}_i^T\boldsymbol{\beta} + u_i)$ and $u_i \sim N(0, \sigma_u^2)$. Here the assumption is that the sample count y_{si} and the non-sample count y_{ri} are independent Poisson variables, with

$$\mu_{si} = E\left(y_{si}|u_i\right) = n_i\left[\exp(\mathbf{x}_i^T\boldsymbol{\beta}+u_i)\right] \text{ and } \mu_{ri} = E\left(y_{ri}|u_i\right) = \left(N_i - n_i\right)\left[\exp(\mathbf{x}_i^T\boldsymbol{\beta}+u_i)\right].$$

The population count in area *i* can be expressed as $y_i = y_{si} + y_{ri}$, where the first term y_{si}, the sample count, is known whereas the second term y_{ri}, the non-sample count, is unknown. A plug-in empirical predictor (EP) of the population count in area *i* is obtained by replacing y_{ri} by its predicted value $\hat{\mu}_{ri} = \hat{E}\left(y_{ri}|u_i\right)$ under the model (1) as

$$\hat{y}_i^{EP} = y_{si} + \hat{\mu}_{ri} = y_{si} + \left(N_i - n_i\right)\hat{\pi}_i^{EP}, \tag{3}$$

with $\hat{\pi}_i^{EP} = \text{expit}(\mathbf{x}_i^T\hat{\boldsymbol{\beta}} + \mathbf{q}_i^T\hat{\mathbf{u}})$ for binary data and $\hat{\pi}_i^{EP} = \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}} + \mathbf{q}_i^T\hat{\mathbf{u}})$ for count data, where $\mathbf{q}_i^T = \left(0,..,1,.,0\right)$ is $1\times m$ vector with 1 in the *i*-th position and $\hat{\mathbf{u}} = (\hat{u}_1,...,\hat{u}_m)^T$. In SAE problems, the sample size n_i is often negligible relative to the population size N_i, then $\hat{y}_i^{EP} = N_i\hat{\pi}_i^{EP}$. An estimate of the proportion or rate in area *i* is given by $\hat{\pi}_i^{EP}$.

In many practical situations small areas are unplanned domains, and many have zero sample sizes. These small areas are referred to as non-sampled areas. Traditional survey estimation approaches do not provide a solution to the small area estimation problem in this case. In contrast, model-based SAE methods can be used to derive estimates for such areas. The conventional approach for estimating area proportions or counts in this case is synthetic estimation (Chandra *et al*., 2011), based on a suitable GLMM fitted to the data from the sampled areas. Let $\mathbf{x}_{i,out}$ denote the vector of covariates associated with non-sampled area *i*. Under (1), the synthetic type predictor for the unknown population value y_i of non-sampled area *i* is then

$$\hat{y}_i^{SYN} = N_i\hat{\pi}_i^{SYN}, \tag{4}$$

with $\hat{\pi}_i^{SYN} = \text{expit}(\mathbf{x}_{i,out}^T\hat{\boldsymbol{\beta}})$ for binary data and $\hat{\pi}_i^{SYN} = \exp(\mathbf{x}_{i,out}^T\hat{\boldsymbol{\beta}})$ for count data. Similarly, small area proportions or rates are estimated by $\hat{\pi}_i^{SYN}$.

A major difficulty with using the GLMM (1) for SAE is estimation of the unknown model parameters $\boldsymbol{\beta}$ and $\mathbf{u}$, since the likelihood function for GLMM often involves high dimensional integrals (computed by integrating a product of discrete and normal densities, which has no analytical solution), which are difficult to evaluate numerically. Maximum likelihood estimation

(MLE) is usually not a problem if one assumes a generalized linear model (GLM) for the sample counts. However, when a random effect is introduced into the model, for example the GLMM (1), we can only write down a penalised likelihood. The corresponding solution is then no longer maximum likelihood but is known as restricted (or residual) maximum likelihood. The second differential of the (penalised) likelihood is calculated and Newton-Raphson iteration performed to obtain converged estimates of the parameter values and the area random effect term. The standard errors of these parameters estimates are obtained from the inverse of the second differential. In particular, an iterative procedure that combines the Penalized Quasi-Likelihood (PQL) estimation of $\boldsymbol{\beta}$ and $\mathbf{u}$ with restricted maximum likelihood estimation (REML) of σ_u^2 can then be used to estimate these unknown parameters. See Chandra *et al.* (2011), Manteiga *et al.* (2007) and Saei and Chambers (2003) for a detailed description of this approach. An alternative data cloning approach to fitting a GLMM is described in Torabi and Shokoohi (2015).

2.1. SAE under Non-stationary GLMM

Following Chandra and Salvati (2018) and Saei and Chambers (2003), a spatially stationary simultaneous autoregression (SAR) specification for the GLMM (2) can be written as

$$g(\boldsymbol{\pi}) = \boldsymbol{\eta}^{sp} = \mathbf{X}\boldsymbol{\beta}^{sp} + \boldsymbol{\varsigma}$$

where the vector of random area effects $\boldsymbol{\varsigma} = (\varsigma_i)$ satisfies $\boldsymbol{\varsigma} = \rho\boldsymbol{\Omega}_S\boldsymbol{\varsigma} + \mathbf{u}$, where ρ is a spatial autoregressive coefficient, $\boldsymbol{\Omega}_S$ is a proximity matrix of order m and $\mathbf{u} \sim N(0, \sigma_u^2\mathbf{I}_m)$. Since then $\boldsymbol{\varsigma} = (\mathbf{I}_m - \rho\boldsymbol{\Omega}_S)^{-1}\mathbf{u}$, we see that $E(\boldsymbol{\varsigma}) = \mathbf{0}$ and $\text{var}(\boldsymbol{\varsigma}) = \sigma_u^2[(\mathbf{I}_m - \rho\boldsymbol{\Omega}_S)(\mathbf{I}_m - \rho\boldsymbol{\Omega}_S^T)]^{-1}$. 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 $\boldsymbol{\Omega}_S = (\omega_{jk}; j,k = 1,...,m)$, where $\omega_{jk} = 1$ if area j shares an edge with area k and $\omega_{jk} = 0$ otherwise. In row-standardised form this becomes $\omega_{jk}^* = q_j^{-1}$, if j and k are contiguous and 0 otherwise, where q_j is the total number of areas that share an edge with area j

(including area *j* itself). A two stage iterative procedure combining the maximum penalised quasi-likelihood (MPQL) and REML is then used to estimate the parameters of this model. The plug-in empirical predictor of the small area count defined under this SAR model (SEP) can be written as

$$\hat{y}_i^{SEP} = y_{si} + (N_i - n_i)\,\hat{\pi}_i^{SEP}, \tag{5}$$

with $\hat{\pi}_i^{SEP} = \text{expit}(\mathbf{x}_i^T\hat{\boldsymbol{\beta}}^{sp} + \hat{\zeta}_i)$ and $\hat{\pi}_i^{SEP} = \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}}^{sp} + \hat{\zeta}_i)$ for binary and count-valued data respectively.

The vector of fixed effect parameters $\boldsymbol{\beta}$ in (2) and the corresponding vector $\boldsymbol{\beta}^{sp}$ in (5) are spatially invariant. Chandra *et al.* (2017) define a spatially nonstationary extension of (2) that is appropriate when the parameters associated with the model covariates vary spatially. Let $\mathbf{d} = (d_i) = (d_1, ..., d_m)^T$ denote an *m*-vector of the coordinates of *m* spatial locations (longitude and latitude, typically of the centroids of the small areas), 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

$$g(\boldsymbol{\pi}_i(d_i)) = \eta_i(d_i) = \mathbf{x}_i^T\boldsymbol{\beta}(d_i) + u_i = \mathbf{x}_i^T\boldsymbol{\beta} + \mathbf{x}_i^T\boldsymbol{\delta}(d_i) + u_i, \tag{6}$$

where the nonstationarity is characterised by an area-specific vector of fixed effects $\boldsymbol{\beta}(d_i) = \boldsymbol{\beta} + \boldsymbol{\delta}(d_i)$; u_i is the area-specific random effect, assumed to follow a Gaussian distribution with zero mean and variance ϕ; and $\boldsymbol{\delta}(\mathbf{d}) = \left(\delta_1(\mathbf{d}),\ldots,\delta_p(\mathbf{d})\right)^T$ is a spatially correlated vector-valued random process with $E(\boldsymbol{\delta}(d_i)) = \mathbf{0}$ and such that

$$\text{cov}\left(\delta_k(d_i), \delta_l(d_j)\right) = c_{kl}\left(1 + L(d_i, d_j)\right)^{-1}. \tag{7}$$

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 $\boldsymbol{\Sigma}_\gamma = \boldsymbol{\Omega} \otimes (\mathbf{c}\mathbf{c}^T)$ to be a covariance matrix, with $\boldsymbol{\Omega} = \left[\left(1 + L(l_i, l_j)\right)^{-1}\right]$ defining the matrix of distances between the sample areas, and where $\otimes$ denotes Kronecker product. In general, the only constraint on the vector **c** is that $\boldsymbol{\Sigma}_\delta = \boldsymbol{\Omega} \otimes (\mathbf{c}\mathbf{c}^T)$ is symmetric and non-negative definite. In practice ϕ and **c** are

unknown and have to be estimated from the data. Chandra et al. (2017) restrict their analysis to the simple specification $\mathbf{c} = \sqrt{\lambda}\mathbf{1}_p$ so that $\text{cov}\left(\delta_k(d_i), \delta_l(d_j)\right) = \lambda\left(1 + L(d_i, d_j)\right)^{-1}$, where $\lambda \geq 0$ and $\mathbf{1}_p$ denotes the unit vector of order *p*. Given this specification, there are just 2 parameters (λ and ϕ) that need to be estimated. See Chandra *et al.* (2017) for more detail about the estimation procedure. Replacing these unknown parameters by their estimated values $\hat{\phi}$ and $\hat{\mathbf{c}}$, and denoting subsequent plug-in estimators by a 'hat', the nonstationary empirical predictor (NSEP) of the population count in area *i* as

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

with $\hat{\pi}_i^{NSEP}(d_i) = \text{expit}\left[\hat{\eta}_i(d_i)\right]$ and $\hat{\pi}_i^{NSEP}(d_i) = \exp\left[\hat{\eta}_i(d_i)\right]$ for binary and count-valued data respectively, where $\hat{\eta}_i(d_i) = \mathbf{x}_i^T\hat{\boldsymbol{\beta}} + \mathbf{s}_i^T\hat{\boldsymbol{\gamma}}(d_i) + \hat{u}_i$. Here $\mathbf{s}_i^T$ is the *i*-th row of $\mathbf{S} = \{diag(\mathbf{x}_1), ..., diag(\mathbf{x}_k)\}$. For large values of N_i, a plug-in nonstationary empirical predictor of the proportion (or rate) in area *i* is $\hat{\pi}_i^{NSEP}$. A nonstationary synthetic predictor (NSSYN) of the total or count for area *i* is of the form $\hat{y}_i^{NSSYN} = N_i\,\hat{\pi}_i^{NSSYN}(d_i)$, with $\hat{\pi}_i^{NSSYN}(d_i) = \text{expit}\left[\hat{\eta}_i(d_i)\right]$ and $\hat{\pi}_i^{NSSYN}(d_i) = \exp\left[\hat{\eta}_i(d_i)\right]$ for binary and count-valued data respectively, and where now $\hat{\eta}_i(d_i) = \mathbf{x}_{out,i}^T\hat{\boldsymbol{\beta}} + \mathbf{s}_{out,i}^T\hat{\boldsymbol{\gamma}}(d_i)$.

3. Spatially non-linear generalized mixed model

We now introduce a spatially non-linear extension of an area level GLMM and describe algorithms to fit this model. We refer this model as spatially non-linear generalized linear mixed model (SNLGLMM). A test for the presence of a nonlinear spatial relationship is also suggested. We start by developing the nonparametric extension of the GLMM, and then suggest a spatial extension of this model. Typically, the fixed effect part of a GLMM is assumed to be linear. However, in reality the functional form of this relationship may be unknown or it may have a complicated functional form. Without loss of generality we restrict our development to the case of a single covariate *x* and use nonparametric regression modelling based on a P-spline approximation. The spatially non-linear GLMM (SNLGLMM) is then of the form

$$g(\pi_i) = \eta_i = f(x_i) + u_i, \tag{9}$$

where $u_i \sim N(0, \sigma_u^2)$ is the area specific random effect and $\pi_i = E(y_i \mid u_i, x_i) = h\{f(x_i) + u_i\}$. In particular, the spatially non-linear logistic-normal mixed model and the spatially non-linear Poisson-normal mixed model for binary and count data, respectively, are defined as $logit(\pi_i) = \eta_i = f(x_i) + u_i$ with $\pi_i = \text{expit}\{f(x_i) + u_i\}$ and $log(\pi_i) = \eta_i = f(x_i) + u_i$ with $\pi_i = \exp\{f(x_i) + u_i\}$. The function $f(x_i)$ in (9) is unknown, but can be approximated sufficiently well by the P-spline approximation

$$f(x_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \beta_0 + \beta_1 x_i + ... + \beta_p x_i^p + \sum_{l=1}^{L} \gamma_l (x_i - \kappa_l)_+^p. \qquad (10)$$

Here *p* is the degree of the spline, $(t)_+^p = t^p$ if $t > 0$ and is 0 otherwise, κ_l for $l = 1, ..., L$ is a set of fixed constants called knots, $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)^T$ is the coefficient vector of the parametric portion of the model and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_L)^T$ is the vector of spline coefficients, *L* is the number of spline knots, and $\gamma_l \sim N(0, \sigma_\gamma^2); l = 1, ..., L$. Provided that the knot locations are sufficiently spread out over the range of *x* and *L* is sufficiently large, the class of functions defined by (10) can approximate most smooth functions. Ruppert *et al*. (2003, chapter 5) suggest the use of a knot for every four observations, up to a maximum of about 40 knots for a univariate application. This is usually done by placing these knots at equally spaced quantiles of the distribution of the covariate.

Note that the P-spline approximation consists of a linear combination of appropriately chosen basis functions. For simplicity, the approximating function $f(x, \boldsymbol{\beta}, \boldsymbol{\gamma})$ in (10) uses truncated polynomial spline basis functions $\left\{1, x, \ldots, x^p, (x - \boldsymbol{\kappa}_1)_+^p, \ldots, (x - \boldsymbol{\kappa}_L)_+^p\right\}$. Other basis functions, e.g. B-splines (Eilers and Marx, 1996) or radial functions, can also be used. Using a large number of knots in expression (10) can lead to an unstable fit. In order to overcome this problem, a penalty is usually put on the magnitude of the spline parameters $\boldsymbol{\gamma}$. See Du *et al.* (2011) and Wahba and Wang (1990) for a discussion about estimation of the penalty term.

When geographically referenced responses play a central role in the analysis and need to be converted to maps, we can use bivariate smoothing, $f(x_{1i},x_{2i}) = f(x_{1i},x_{2i},\boldsymbol{\beta},\boldsymbol{\gamma})$ where x_{1i} and x_{2i} are spatial coordinates. This is usually the case with environmental, agricultural, public health and poverty mapping applications. Consequently we assume the following model (for further details see Opsomer *et al.,* 2008)

$$f(x_{1i},x_{2i},\boldsymbol{\beta},\boldsymbol{\gamma}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \mathbf{z}_i\boldsymbol{\gamma}, \tag{11}$$

where $\mathbf{z}_i$ is the *i*-th row of the following $n \times L$ matrix

$$\mathbf{Z} = \left[C\left(\mathbf{x}_i - \boldsymbol{\kappa}_l\right)\right]_{\substack{1\le i\le n\\ 1\le l\le L}} \left[C\left(\boldsymbol{\kappa}_l - \boldsymbol{\kappa}_{l'}\right)\right]^{-1/2}_{1\le l\le L}, \tag{12}$$

where $C(\mathbf{t}) = \|\mathbf{t}\|^2 \log\|\mathbf{t}\|$, $\mathbf{x}_i = \left(x_{1i}, x_{2i}\right)$ and $\boldsymbol{\kappa}_l (l = 1,\ldots,L)$ are knots. Note that the $C(\mathbf{t})$ function is defined so that when there is a knot at every observation (that is, the full rank case) this model for bivariate smoothing leads to a thin plate spline (Green and Silverman, 1994). The second matrix on the right-hand side of (12) applies a linear transformation to the radial basis functions defining the first matrix, and was recommended by Ruppert *et al.* (2003) as a way of making the radial spline behave approximately like a thin plate spline. More details on the **Z** matrix can be found in Ruppert *et al.* (2003, chapter 13) and Kammann and Wand (2003). This type of bivariate smoothing will be used in the application in Section 6 in order to take into account the spatial information in the data.

As suggested by Ruppert *et al*. (2003), fitting the approximation (10) can be simplified by treating the vector γ as a random-effect vector in a mixed model specification, which allows estimation of β and γ by maximum likelihood methods. Following Opsomer *et al*. (2008), Wand (2003) and Ruppert *et al*. (2003, chapter 4), the spline approximation to (9), and to (11), can be written as

$$g(\boldsymbol{\pi}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}. \tag{13}$$

If other covariates are available, they can be included in the model as parametric terms, by being added to the matrix **X**. In this model $\boldsymbol{\gamma}$ is assumed to be a Gaussian random vector of dimension *L*. In particular, it is assumed that $\boldsymbol{\gamma} \sim N_L(\mathbf{0}, \boldsymbol{\Sigma}_\gamma = \sigma_\gamma^2 \mathbf{I}_L)$, where $\mathbf{I}_t$ denotes the identity matrix of

dimension *t*. Here **u** is the *m*-vector of random area effects. As usual, it is assumed that the area effects in **u** are distributed independently of the spline effects in $\boldsymbol{\gamma}$ with $\mathbf{u} \sim N_m(\mathbf{0}, \boldsymbol{\Sigma}_u = \sigma_u^2 \mathbf{I}_m)$.

Under (13), a plug-in spatially non-linear empirical predictor (SNLEP) for the total count y_i in area *i* is given by

$$\hat{y}_i^{SNLEP} = y_{si} + \left(N_i - n_i\right)\hat{\pi}_i^{SNLEP}, \qquad (14)$$

where $\hat{\pi}_i^{SNLEP} = \text{expit}(\mathbf{x}_i^T\hat{\boldsymbol{\beta}} + \mathbf{z}_i^T\hat{\boldsymbol{\gamma}} + \mathbf{q}_i^T\hat{\mathbf{u}})$ for binary data and $\hat{\pi}_i^{SNLEP} = \exp(\mathbf{x}_i^T\hat{\boldsymbol{\beta}} + \mathbf{z}_i^T\hat{\boldsymbol{\gamma}} + \mathbf{q}_i^T\hat{\mathbf{u}})$ for count data, and $\mathbf{x}_i^T$, $\mathbf{z}_i^T$ and $\mathbf{q}_i^T$ denote respectively the rows of **X**, **Z** and $\mathbf{I}_m$ that correspond to area *i*. When n_i is negligible compared to N_i, the SNLEP (14) is $\hat{y}_i^{SNLEP} = N_i\hat{\pi}_i^{SNLEP}$. An estimate of the proportion or rate in area *i* is $\hat{\pi}_i^{SNLEP}$. For synthetic nonparametric prediction exactly the same approach can be taken with the spline-based small area model (13). When geo-referenced population location data are available, and spline smoothing is over these locations (e.g. using radial basis functions), the nonparametric model (13) is effectively accounting for spatial correlation in the population values of *y* over and above that 'explained' by the random area effects. In this case, model (13) has the potential to improve conventional synthetic estimation for out of sample areas. Under (13), the nonparametric synthetic (SNLSYN) predictor for area *i* is defined as $\hat{y}_i^{SNLSYN} = N_i\hat{\pi}_i^{SNLSYN}$. That is, SNLSYN is SNLEP with $\hat{u}_i = 0$. Here, $\hat{\pi}_i^{SNLSYN} = \text{expit}(\mathbf{x}_{i,out}^T\hat{\boldsymbol{\beta}} + \mathbf{z}_{i,out}^T\hat{\boldsymbol{\gamma}})$ for binary data and $\hat{\pi}_i^{SNLSYN} = \exp(\mathbf{x}_{i,out}^T\hat{\boldsymbol{\beta}} + \mathbf{z}_{i,out}^T\hat{\boldsymbol{\gamma}})$ for count data, and $\mathbf{z}_{i,out}^T$ denote the row of **Z** associated with non-sampled area *i*.

3.1 Parameter estimation

Let $\mathbf{y}_s = \left(y_{s1}, ..., y_{sm}\right)^T$ denotes the $m \times 1$ vector of sample counts, with $f_1\left(\mathbf{y}_s \mid \mathbf{u}, \boldsymbol{\gamma}\right)$ denoting the probability density function of $\mathbf{y}_s$ conditional on $\{\mathbf{u}, \boldsymbol{\gamma}\}$ and let $f_2(\mathbf{u})$ and $f_3(\boldsymbol{\gamma})$ be the probability density functions of **u** and $\boldsymbol{\gamma}$ respectively. Here $y_{si}; i = 1, .., m$ has a Binomial (or Poisson) distribution whereas $f_2(\mathbf{u})$ and $f_3(\boldsymbol{\gamma})$ have normal distributions with zero means and variances $\boldsymbol{\Sigma}_u = \sigma_u^2\mathbf{I}_m$ and $\boldsymbol{\Sigma}_\gamma = \sigma_\gamma^2\mathbf{I}_L$ respectively. The log-likelihood function defined by the

vector $\mathbf{y}_s$ conditional on fixed $\{\mathbf{u},\boldsymbol{\gamma}\}$ and the logarithm of the probability density functions of $\mathbf{u}$ and $\boldsymbol{\gamma}$ are

$$l_1 = \ln\{f_1(\mathbf{y}_s \mid \mathbf{u},\boldsymbol{\gamma})\} = \begin{cases} \text{Constant} + \sum_{i=1}^{m}\{y_{si}\eta_{si} - n_i \ln[1+\exp(\eta_{si})]\} & \text{for Binomial data} \\ \text{Constant} + \sum_{i=1}^{m}\{y_{si}\eta_{si} - n_i \exp(\eta_{si})\} & \text{for Poisson data} \end{cases},$$

$$l_2 = \text{Constant} - \frac{1}{2}\{\ln|\boldsymbol{\Sigma}_u| + \mathbf{u}^T\boldsymbol{\Sigma}_u^{-1}\mathbf{u}\},$$

$$l_3 = \text{Constant} - \frac{1}{2}\{\ln|\boldsymbol{\Sigma}_\gamma| + \boldsymbol{\gamma}^T\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\gamma}\}.$$

For fixed $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$, the $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\mathbf{u}$ values that jointly maximize $l = l_1 + l_2 + l_3$ (sum of three components based on joint distribution of $\mathbf{y}_s$, $\mathbf{u}$ and $\boldsymbol{\gamma}$) are called the maximum penalised quasi-likelihood (MPQL) estimates. 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 $\boldsymbol{\gamma}$. Substituting these estimates into (14) ($\boldsymbol{\eta} = logit(\boldsymbol{\pi})$ or $\boldsymbol{\eta} = log(\boldsymbol{\pi})$ for Binomial or Poisson data respectively) yields the MPQL estimate of $\boldsymbol{\pi}$. In practice the variance components parameters defining the matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$ are unknown and have to 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 maximum likelihood (ML) and restricted maximum likelihood (REML) can be defined. In particular, the bias in the REML estimates is typically small. 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 5 where we provide simulation results that provide a perspective on this bias problem.

Under the hybrid approach, parameter estimates for the SNLGLMM are obtained by a two-stage iterative process. In the first step (a) MPQL estimates of $\boldsymbol{\beta}$, $\mathbf{u}$ and $\boldsymbol{\gamma}$ are obtained based on $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$ (assumed known) and in the second step (b) the estimates of $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$ are obtained by either a ML or REML procedure given these MPQL estimates. This hybrid algorithm combining

MPQL with ML and REML for fitting generalized linear mixed models was developed in McGilchrist (1994). Below we outline the extension of McGilchrist (1994) to estimation of variance components for the small area estimation problem.

The estimation process is repeated with the updated estimates of $\boldsymbol{\beta}$, $\mathbf{u}$, $\boldsymbol{\gamma}$, $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\gamma$ being used in step (a) until the required convergence is achieved. The SNLEP estimate of $\boldsymbol{\eta}$ is obtained by substituting the converged values $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$ and $\hat{\boldsymbol{\gamma}}$ in the right hand side of (14). The inverse link function is then used to derive $\boldsymbol{\pi}$ from $\boldsymbol{\eta}$. This leads to the SNLEP estimate (14). Similar, the 'log-likelihood' for the Poisson observation vector $\mathbf{y}_s$ conditional on fixed $\{\mathbf{u},\boldsymbol{\gamma}\}$ and the 'log-likelihood' for $\mathbf{u}$ and $\boldsymbol{\gamma}$ are defined. This algorithm for parameter estimation is therefore implemented as follows:

MPQL Stage

Given $\boldsymbol{\Sigma}_u$ (or σ_u^2) and $\boldsymbol{\Sigma}_\gamma$ (or σ_γ^2), an iterative procedure can be used to obtain the MPQL estimates of $\boldsymbol{\beta}$, $\mathbf{u}$ and $\boldsymbol{\gamma}$. This is:

1. Initialize the iteration. Set $k=0$ and initial values $\boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_0$ and $\mathbf{u}_0$.

2. Update these values via

$$\begin{bmatrix}\boldsymbol{\beta}_{k+1}\\ \boldsymbol{\gamma}_{k+1}\\ \mathbf{u}_{k+1}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\beta}_{k}\\ \boldsymbol{\gamma}_{k}\\ \mathbf{u}_{k}\end{bmatrix}+\mathbf{V}_k^{-1}\begin{bmatrix}\mathbf{X}^T\\ \mathbf{Z}^T\\ \mathbf{I}_m\end{bmatrix}\frac{\partial l_1}{\partial \boldsymbol{\beta}_k}-\mathbf{V}_k^{-1}\begin{bmatrix}\mathbf{0}\\ \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\boldsymbol{\gamma}_k\\ \boldsymbol{\Sigma}_u^{-1}\mathbf{u}_k\end{bmatrix},$$

where

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

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

3. Return to step 2.

4. At convergence, update $\boldsymbol{\eta}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\boldsymbol{\gamma}+\mathbf{u}$.

ML/REML Stage

Given the estimates of $\boldsymbol{\beta}$, $\mathbf{u}$ and $\boldsymbol{\gamma}$ from steps 1 - 3 of the MPQL stage, we use ML or REML to estimate the variance components $\boldsymbol{\Sigma}_u$ (or σ_u^2) and $\boldsymbol{\Sigma}_\gamma$ (or σ_γ^2). Following Schall (1991), define an adjusted variable

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

with $\mathbf{e}^* = \mathbf{e}\left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}}\right)$ as follows $\mathbf{y}^* \cong \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{I}_m\mathbf{u} + \mathbf{e}^*$. Here, $\mathbf{e}^* = (e_i^*)$ is called an adjusted error term. Its variance is $\mathrm{Var}(\mathbf{e}^*) = \tau\boldsymbol{\Sigma}_e^*$ where, τ is dispersion parameter and

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

Note that for a GLMM based on the Binomial or Poisson models, the variance is a known function of the mean and so $\tau = 1$. We assume that the random variable $\mathbf{y}^*$ has an approximately normal marginal distribution with mean $\mathbf{X}\boldsymbol{\beta}$ and $Var(\mathbf{y}^*) = \mathbf{V} = \boldsymbol{\Sigma}_u + \mathbf{Z}\boldsymbol{\Sigma}_\gamma\mathbf{Z}^T + \Sigma_e^*$. Maximising the log-likelihood generated by $\mathbf{y}^*$ with respect to σ_u^2 and σ_γ^2 we obtain the solutions

$$\boldsymbol{\sigma}_u^2 = m^{-1}\left[tr(\mathbf{T}) + \mathbf{u}^T\mathbf{u} \right]$$

and

$$\boldsymbol{\sigma}_\gamma^2 = (L)^{-1}\left[tr(\mathbf{T}) + \boldsymbol{\gamma}^T\boldsymbol{\gamma} \right]$$

where

$$\mathbf{T} = \left\{ \begin{bmatrix} \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix} \boldsymbol{\Sigma}_e^{*-1} \begin{bmatrix} \mathbf{Z}^T \\ \mathbf{I}_m \end{bmatrix}^T + \begin{bmatrix} \boldsymbol{\sigma}_\gamma^2\mathbf{I}_L & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}_u^2\mathbf{I}_m \end{bmatrix} \right\}^{-1}.$$

REML estimation of σ_u^2 and σ_γ^2 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, which is achieved when the squared difference between the estimated model parameters obtained from two successive iterations is less than a very small value. R code that implements this algorithm is available from the authors. Note that the starting point for $\boldsymbol{\sigma}_{\boldsymbol{\gamma}}^2$ in the preceding algorithm can be defined based on information about the strength of spatial non-linearity in the data. Similarly, the starting point for σ_u^2 can be set to the estimated value of the variance component defined by the area effects under the model (2). Our experience is that good convergence performance is usually achieved by choosing the starting points of the hybrid MPQL/REML procedure to be the values 0.5 for σ_u^2 and 1.0 for $\boldsymbol{\sigma}_{\boldsymbol{\gamma}}^2$. We also suggest that the values of $\boldsymbol{\beta}, \mathbf{u}$, $\boldsymbol{\gamma}$ all be set equal to 0 as a starting point. These values represent a good compromise for fast convergence of the algorithm. We have tested different starting points in our simulation experiments, and the hybrid MPQL/REML algorithm with these starting points usually converges, on average, after fifteen iterations.

3.2 A diagnostic for spatial nonlinearity

In the spirit of Chandra *et al.* (2017) and Opsomer *et al.* (2008), we develop a bootstrap procedure to test the spatial nonlinearity hypothesis, that is, the hypothesis $H_{0\sigma_\gamma^2}: \sigma_\gamma^2 = 0$ versus the one-sided alternative $H_{1\sigma_\gamma^2}: \sigma_\gamma^2 > 0$. In the proposed procedure, two models are fitted, the first without random effects that characterizes the spatial relationship in the data (denoted by $H_{0\sigma_\gamma^2}: \sigma_\gamma^2 = 0$), and second with these random effects included (denoted by $H_{1\sigma_\gamma^2}: \sigma_\gamma^2 > 0$). This involves first calculating the value $l_{\sigma_\gamma^2} = 2(l_1 - l_0)$, where ℓ_0 denotes the restricted log-likelihood under the null $H_{0\sigma_\gamma^2}$ and ℓ_1 denotes the corresponding value under the alternative $H_{1\sigma_\gamma^2}$. The level of significance of $l_{\sigma_\gamma^2}$ is then calculated via a parametric bootstrap as follows:

1) Given the sample counts $\mathbf{y}_s$, calculate the parameter estimates using model (2). Let $\hat{\sigma}_u^2$ and $\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}$, denote the resulting estimates.

2) Generate a vector $\mathbf{t}_1^*$ whose elements are m independent realisations of a $N(0,1)$ variable. Construct the bootstrap vector $\mathbf{u}^* = \hat{\sigma}_u \mathbf{t}_1^*$.

3) Calculate the small area bootstrap population parameters

$\boldsymbol{\pi}^* = (\pi_1^*, \ldots, \pi_m^*)^T = \text{expit}\left\{\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{u}^*\right\}$ or $\boldsymbol{\pi}^* = \exp\left\{\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{u}^*\right\}$

depending upon whether the data are binary or count valued respectively.

4) Generate a bootstrap sample of independent bootstrap Binomial realisations (n_i, π_i^*) or Poisson (π_i^*) and fit both the null and the alternative models. Calculate the bootstrap value $l^*_{\sigma_\gamma^2}$ of the $l_{\sigma_\gamma^2}$.

Repeat steps 2 - 4 B times. In the *b*-th bootstrap replication, let $l^{*(b)}_{\sigma_\gamma^2}$ be the value of the difference between the restricted log-likelihood under the null $H_{0\sigma_\gamma^2}$ and the corresponding value under the alternative $H_{1\sigma_\gamma^2}$. The significance of the calculated value of $l_{\sigma_\gamma^2}$ is then evaluated by comparing it with the bootstrap distribution of $l^{*(b)}_{\sigma_\gamma^2}$. In Section 5 we provide simulation results that provide a perspective on the type I error and power of this test statistic.

4. Mean Squared Error estimation

The mean squared error (MSE) estimation of the empirical predictor (3) follows from Johnson *et al.* (2010) and Chandra e*t al.* (2011). In this Section we develop an approximation to the MSE of the SNLEP (14). Let us denote by $\mathbf{G} = (\mathbf{Z}, \mathbf{I}_m)$, $\mathbf{v} = (\boldsymbol{\gamma}^T, \mathbf{u}^T)^T$ and $\boldsymbol{\Sigma}_v = \text{var}(\boldsymbol{v}) = \begin{pmatrix} \boldsymbol{\Sigma}_\gamma & 0 \\ 0 & \boldsymbol{\Sigma}_u \end{pmatrix}$ as covariates matrix of order $(m+L)\times m$, vector of random effects of order $(m+L)\times 1$ and covariance matrix of order $(m+L)\times(m+L)$ respectively. Here, m and L are number of areas and number of spline knots respectively. The variance component parameters in covariance matrix $\boldsymbol{\Sigma}_v$ are $\boldsymbol{\delta} = (\sigma_u^2, \sigma_\gamma^2)^T$. From (13), we write $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{G}\hat{\mathbf{v}}$ as predictor of $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{G}\mathbf{v}$. Let $\mathbf{y} = (y_1, \ldots, y_m)^T$, $\mathbf{y}_s = (y_{s_1}, \ldots, y_{s_m})^T$ and $\mathbf{y}_r = (y_{r_1}, \ldots, y_{r_m})^T$ denote the vectors of population, sample and non-sample counts, respectively. We further decompose the total count as $\mathbf{y} = \mathbf{c}_s^T\mathbf{y}_s + \mathbf{c}_r^T\mathbf{y}_r$ and write the SNLEP as $\hat{\mathbf{y}}^{SNLEP} = \mathbf{c}_s^T\mathbf{y}_s + \mathbf{c}_r^T\hat{\boldsymbol{\mu}}$. Here $\hat{\boldsymbol{\mu}} = diag\left\{(N_i - n_i); i = 1, \ldots, m\right\}\hat{\boldsymbol{\pi}}$ with

$\hat{\boldsymbol{\pi}} = (\hat{\pi}_1^{SNLEP},...,\hat{\pi}_m^{SNLEP})^T$ and $\boldsymbol{\mu} = diag\{(N_i - n_i); i = 1,...,m\}\boldsymbol{\pi}$ with $\boldsymbol{\pi} = (\hat{\pi}_1,...,\hat{\pi}_m)^T$. For binary data, $\boldsymbol{\pi} = \text{expit}[\boldsymbol{\eta}]$ and $\hat{\boldsymbol{\pi}} = \text{expit}[\hat{\boldsymbol{\eta}}]$ and for count data, $\boldsymbol{\pi} = \exp[\boldsymbol{\eta}]$ and $\hat{\boldsymbol{\pi}} = \exp[\hat{\boldsymbol{\eta}}]$. Under the area level version of model (13), we have $\mathbf{c}_s = \mathbf{c}_r = \mathbf{c} = \mathbf{I}_m$. The prediction error of SNLEP is therefore $\hat{\mathbf{y}}^{SNLEP} - \mathbf{y} = \mathbf{c}^T\hat{\boldsymbol{\mu}} - \mathbf{c}^T\mathbf{y}_r = \mathbf{c}^T\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \mathbf{y}_r)\right]$. The MSE of SNLEP is then approximated as

$$MSE(\hat{\mathbf{y}}^{SNLEP} - \mathbf{y}) = E\left[(\hat{\mathbf{y}}^{SNLEP} - \mathbf{y})(\hat{\mathbf{y}}^{SNLEP} - \mathbf{y})^T\right] \approx \mathbf{c}^T E\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\right]\mathbf{c}. \quad (15)$$

The approximation (15) follows from the fact that the cross product term is zero and the component $E\left[(\boldsymbol{\mu} - \mathbf{y}_r)(\boldsymbol{\mu} - \mathbf{y}_r)^T\right]$, which measures the spread of observations about the mean while estimating the actual values rather than the expected value, is not required under the area level model. Using first order Taylor linearization, we can approximate $\hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \approx (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^T \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}|_{\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}}$. Then, from expression (15), we have

$$\mathbf{c}^T E\left[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T\right]\mathbf{c} = \mathbf{C}^T\left\{E\left[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^T\right]\right\}\mathbf{C}, \text{ with } \mathbf{H} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}|_{\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}} \text{ and } \mathbf{C} = \mathbf{cH}.$$

We now define $\hat{\tau} = \mathbf{C}\hat{\boldsymbol{\eta}} = \mathbf{cH}\hat{\boldsymbol{\eta}} = \mathbf{C}\left\{\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\gamma}} + \hat{\mathbf{u}}\right\} = \mathbf{C}\left\{\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{G}\hat{\mathbf{v}}\right\}$. This leads to

$$MSE(\hat{\mathbf{y}}^{SNLEP} - \mathbf{y}) = MSE(\hat{\tau} - \tau) \approx \mathbf{C}^T\left\{E\left[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^T\right]\right\}\mathbf{C}.$$

That is, $\hat{\tau}$ can be expressed as a linear combination of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{v}}$. Consequently we can use the results of Prasad and Rao (1990) to define an estimator of an approximation to the MSE of the SNLEP that reflects the true variability associated with this estimator. This approximate MSE of the SNLEP is

$$MSE(\hat{\mathbf{y}}^{SNLEP}) \approx M_1(\boldsymbol{\delta}) + M_2(\boldsymbol{\delta}) + M_3(\boldsymbol{\delta}). \quad (16)$$

The components $M_1(\boldsymbol{\delta})$ and $M_2(\boldsymbol{\delta})$ constitute the largest part of the MSE (15). These are the MSE of the best linear unbiased predictor-type estimator when the variance components in $\boldsymbol{\delta}$ are assumed to be known (Rao, 2003). The third component $M_3(\boldsymbol{\delta})$ is the variability due to estimation of the variances of the random effects from the data, see Rao and Molina (2015, chapter 5, page 100-101), Saei and Chambers (2003), Manteiga *et al*. (2007) and references therein. Following Das *et al.* (2004) and Opsomer *et al.* (2008) we define estimators of the

different components of the approximate MSE (16). Let us write $\mathbf{B} = -\frac{\partial^2 l_1}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}|_{\hat{\boldsymbol{\eta}}=\boldsymbol{\eta}}$, $\boldsymbol{\Psi} = \mathbf{G}^T\mathbf{B}\mathbf{G} + \boldsymbol{\Sigma}_v^{-1}$, $\boldsymbol{\Omega} = \mathbf{G}^T\left(\mathbf{B} + \mathbf{B}\mathbf{G}\boldsymbol{\Sigma}_v\mathbf{G}^T\mathbf{B}\right)\mathbf{G}$. In particular,

$$\mathbf{B} = \begin{cases} diag\left\{n_i\hat{\pi}_i^{SNLEP}\left(1-\hat{\pi}_i^{SNLEP}\right); i=1,...,m\right\} & \text{for Binomial data} \\ diag\left\{n_i\hat{\pi}_i^{SNLEP}; i=1,...,m\right\} & \text{for Poisson data} \end{cases},$$

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

$\mathbf{C} = \mathbf{c}\boldsymbol{\mu}(1-\boldsymbol{\mu}) = diag\left\{\left(N_i - n_i\right)\hat{\pi}_i^{SNLEP}\left(1-\hat{\pi}_i^{SNLEP}\right); i=1,...,m\right\}$ and

$\mathbf{C} = \mathbf{c}\boldsymbol{\mu} = diag\left\{\left(N_i - n_i\right)\hat{\pi}_i^{SNLEP}; i=1,...,m\right\}$ are for binary and count data respectively. Following McGilchrist (1994), we define

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

and

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

and subsequently denote the partitioning of the matrix $\mathbf{V}$ and its inverse $\mathbf{V}^{-1}$ defined by dimension of $\boldsymbol{\beta}$ and $\mathbf{v}$. Here, sub matrices of $\mathbf{V}^{-1}$ are given by

$$F_{11} = \left[\mathbf{X}^T\left(\mathbf{B} - \mathbf{B}\mathbf{G}\boldsymbol{\Psi}^{-1}\mathbf{G}^T\mathbf{B}\right)\mathbf{X}\right]^{-1}, \text{ and } F_{22} = \boldsymbol{\Psi}^{-1} + \boldsymbol{\Psi}^{-1}\left\{\left(\mathbf{G}^T\mathbf{B}\mathbf{X}\right)F_{11}\left(\mathbf{X}^T\mathbf{B}\mathbf{G}\right)\right\}\boldsymbol{\Psi}^{-1}.$$

Let us define $\Delta = \mathbf{C}\mathbf{G}\boldsymbol{\Psi}^{-1} = \mathbf{G}^*\boldsymbol{\Psi}^{-1}$ with $\mathbf{G}^* = \mathbf{C}\mathbf{G}$ and let $\mathbf{G}^*_{(k)}$ be the *k*-th row of the matrix $\mathbf{G}^*$, with its derivative given by $\frac{\partial \Delta_{(k)}}{\partial \boldsymbol{\delta}} = \frac{\partial\left(\mathbf{G}^*_{(k)}\boldsymbol{\Psi}^{-1}\right)}{\partial \boldsymbol{\delta}} = \nabla_{(k)} = \mathbf{G}^*_{(k)}\boldsymbol{\Psi}^{-1}\left(\Sigma_v^{-1}\right)\left(\Sigma_v^{-1}\right)\boldsymbol{\Psi}^{-1}$. Using these expressions together with the results and approximations given in Opsomer *et al.* (2008), Das *et al.* (2004) and Prasad and Rao (1990), the components of MSE (16) are finally defined as

$$M_1(\boldsymbol{\delta}) = \mathbf{C}\left(\mathbf{G}\boldsymbol{\Psi}^{-1}\mathbf{G}^T\right)\mathbf{C}^T,$$

$$M_2(\boldsymbol{\delta}) = \mathbf{C}\left[\mathbf{X} - \mathbf{G}\boldsymbol{\Psi}^{-1}\mathbf{G}^T\mathbf{B}\mathbf{X}\right]F_{11}\left[\mathbf{X}^T - \mathbf{X}^T\mathbf{B}\mathbf{G}\boldsymbol{\Psi}^{-1}\mathbf{G}^T\right]\mathbf{C}^T,$$

$$M_3(\boldsymbol{\delta}) = tr\left[\left(\nabla_{(k)}\boldsymbol{\Omega}\nabla_{(l)}^T\right)Var(\hat{\boldsymbol{\delta}})\right].$$

Here $Var(\hat{\boldsymbol{\delta}})$ is the asymptotic covariance matrix of the estimates of variance components $\hat{\boldsymbol{\delta}}$, which can be evaluated as the inverse of the appropriate Fisher information matrix for $\hat{\boldsymbol{\delta}}$. Note that this also depends upon whether we are using maximum likelihood or restricted maximum likelihood estimates for $\hat{\boldsymbol{\delta}}$. Following the results from Opsomer *et al.* (2008) and Das *et al.* (2004), under model (13) we then have

$$E\left(M_1(\hat{\boldsymbol{\delta}}) - M_3(\hat{\boldsymbol{\delta}})\right) \approx M_1(\boldsymbol{\delta}) + o(m^{-1}),$$

$$E\left(M_2(\hat{\boldsymbol{\delta}})\right) \approx M_2(\boldsymbol{\delta}) + o(m^{-1}) \text{ and}$$

$$E\left(M_3(\hat{\boldsymbol{\delta}})\right) \approx M_3(\boldsymbol{\delta}) + o(m^{-1}).$$

That is, the bias of $M_1(\hat{\boldsymbol{\delta}})$ is of the same order as $M_3(\hat{\boldsymbol{\delta}})$, see Rao and Molina (2015, chapter 5, page 106). Therefore, an approximately model unbiased estimator of the MSE approximation (16) is

$$mse(\hat{\mathbf{y}}^{SNLEP}) \approx M_1(\hat{\boldsymbol{\delta}}) + M_2(\hat{\boldsymbol{\delta}}) + 2M_3(\hat{\boldsymbol{\delta}}), \qquad (17)$$

where $M_k(\hat{\boldsymbol{\delta}}); k = 1,2,3$ are obtained from $M_k(\boldsymbol{\delta})$ replacing $\boldsymbol{\delta}$ by its estimate $\hat{\boldsymbol{\delta}}$. Note that the bias of the MSE estimator (17) is of order $o(m^{-1})$ since both $M_2(\hat{\boldsymbol{\delta}})$ and $M_3(\hat{\boldsymbol{\delta}})$ have bias of order $o(m^{-1})$.

5. Simulation studies

This Section presents the results from simulation studies that compare the empirical performance of the proposed SNLEP estimator (14) under the SNLGLMM (13) with the SEP under the SAR model (5), the NSEP (8) under the NSGLMM (6) and the EP (3) under the GLMM (1). The performance of the analytical MSE estimator (17) for the SNLEP is also evaluated. The performance criteria used are the percentage Relative Bias (*RB*) and the percentage Relative Root MSE (*RRMSE*), defined as:

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

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

where y_{it} is the true value of the parameter for small area *i* in iteration *t*, $\hat{y}_{it}$ is its predicted value and *T* is the number of simulated samples. In addition, we computed the square root of the actual (i.e. Monte Carlo) MSE (TRMSE) and the square root of the average of the estimated values of this MSE (ERMSE), i.e.

$$TRMSE_i = \sqrt{TMSE_i} \text{ with } TMSE_i = T^{-1}\sum_{t=1}^{T}\left(\hat{y}_{it} - y_{it}\right)^2 \text{ and } ERMSE_i = \sqrt{T^{-1}\sum_{t=1}^{T} mse_{it}}.$$

As above, the subscript *i* indexes the small areas and the subscript *t* indexes the *T* Monte Carlo simulations, with mse_{it} denoting the simulation *t* value of the MSE estimator in area *i*, and $TMSE_i$ denotes the actual MSE in area *i*. We also calculated performance indicators for the MSE estimates (17). The first was based on the fact that in many applications of small area estimation, MSE estimators are used to calculate confidence intervals for the small area quantities of interest. Consequently it was interesting to evaluate the coverage properties of such intervals. In particular, we focused on the 'two sigma' (i.e. nominal 95 per cent) intervals based on a normality assumption for the prediction error, and calculated the per cent coverage rate for area *i* as

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

We also calculated the percentage Relative Bias and the percentage Relative RMSE (denoted by RE) of the MSE estimator (17), using the same definitions as above, but replacing the predicted value there by the MSE estimate and the true value of the parameter by the true, i.e. Monte Carlo, MSE.

The model based simulations were of two scenarios, corresponding to the logistic-normal mixed model (binary unit level data, aggregated to area level) and the Poisson-normal mixed model (area level counts). We set the number of small areas *m* = 100 in both and considered two values for the area specific sample sizes $n_i = 10$ and 50 with $N_i = 100$ and 5000, respectively. We used an area level version GLMM to generate data. For the binary case, response values were generated from $y_i \sim Binomial(n_i, \pi_i)$ and $logit(\pi_i) = \eta_i = f(x_i) + u_i$ with

$\pi_i = \exp(\eta_i)\{1+\exp(\eta_i)\}^{-1}$. For the case of count data, response values were generated from $y_i \sim Poiss(n_i\pi_i)$ where $\pi_i = \exp(\eta_i)$ and $\log(\pi_i) = \eta_i = f(x_i) + u_i$. Spatial locations were simulated as the values of two independently distributed uniform [0, 1] covariates x_1 and x_2, and the random area effects u_i were generated as *m* independent realizations from a $N\left(0, \sigma_u^2 = 0.0625\right)$ distribution. We considered two different choices of the response function $f(x_1, x_2)$:

(i) Plane: $f(x_1, x_2) = 0.5x_1 + 0.2x_2$,

(ii) Mountain:

$$f(x_1, x_2) = \frac{40\exp\left[8\left\{\left(x_1-0.5\right)^2+\left(x_2-0.5\right)^2\right\}\right]}{\exp\left[8\left\{\left(x_1-0.2\right)^2+\left(x_2-0.7\right)^2\right\}\right]+\exp\left[8\left\{\left(x_1-0.7\right)^2+\left(x_2-0.2\right)^2\right\}\right]}.$$

The first case corresponds to a situation in which the GLLM underlying the EP (3) is a good representation of the true underlying model while the other predictors may be too complex. The second choice corresponds to a more complicated relationship between *y* and *x*: the model surface in this case has a ridge along the 45° line and we expect that the predictors based on spatial models will work better than the EP. A total of $T = 1000$ data sets were independently generated under each of these models and the predicted small area counts for the different predictors developed in the previous Sections were calculated. Note that in these simulations the SNLGLMM uses the radial basis functions described in Section 3 with $L = 25$ knots, following Pratesi *et al.* (2009).

Table 1 shows the average values of percentage relative bias (RB) and the average values of percentage relative root MSE (RRMSE) recorded by the SAE methods investigated in our simulations. In particular, Table 1 sets out the average relative biases and the average relative RMSEs of the four small area predictors (EP, SNLEP, NSEP and SEP) across the two different types of response function $f(x_1, x_2)$ (i.e. plane and mountain) used in SNLGLMM for data generation and two types of data (binary and count), allowing one to compare the four predictors across different data types. In Table 2 we show the performance of the MSE estimator (17)

corresponding to the SNLEP predictor. In all cases (both Tables 1 and 2) the values of the performance measures are reported as averages over the areas.

The results in Table 1 are essentially as one would expect. Here we clearly see that the performances of EP and SNLEP are on a par when data are generated using the plane function for both the binary and count situations. However, two of the alternative predictors, the SEP and the NSEP, perform worse than the other predictors especially in terms of relative RMSE. In contrast, the simplicity of the EP predictor comes at a price when data are generated using the mountain response function. In this case, the SNLEP performs best, with low bias and is more efficient than the EP, NSEP and SEP. As expected, these results improve as the sample size increases.

Table 1. Percentage relative bias (*RB*) and the percentage relative root MSE (*RRMSE*) of the EP, SNLEP, NSEP and SEP with $n_i = 10$ and $n_i = 50$ under the plane and the mountain response functions. Values are averages over the small areas.

$f(x)$	Predictors	$n_i = 10$		$n_i = 50$	
		RB	RRMSE	RB	RRMSE
		Binary data			
Plane	EP	-0.073	12.89	-0.029	9.51
	SNLEP	-0.074	13.95	-0.029	9.80
	NSEP	-0.075	16.14	-0.030	10.21
	SEP	-0.075	17.04	-0.029	10.08
Mountain	EP	6.320	26.65	1.729	15.96
	SNLEP	1.587	20.61	0.527	13.22
	NSEP	2.386	23.42	0.798	14.35
	SEP	2.284	27.64	1.506	15.91
		Count data			
Plane	EP	0.189	20.08	0.048	12.25
	SNLEP	0.183	20.86	0.046	12.39
	NSEP	0.187	20.24	0.047	12.28
	SEP	0.184	21.44	0.047	12.29
Mountain	EP	6.704	33.19	1.987	17.11
	SNLEP	2.165	25.94	0.733	15.41
	NSEP	3.433	29.16	1.127	16.08
	SEP	6.011	33.35	3.983	18.62

Figure 1 shows the distribution of area-specific values of Relative Bias and RRMSE for all four predictors (all expressed in percentage terms) obtained in simulations where m = 100 and the area sample size $n_i = 50$, with a Binomial response. These plots confirm the results presented in Table 1: the SNLEP dominates the other predictors in terms of efficiency when the spatial relationship is non-linear.

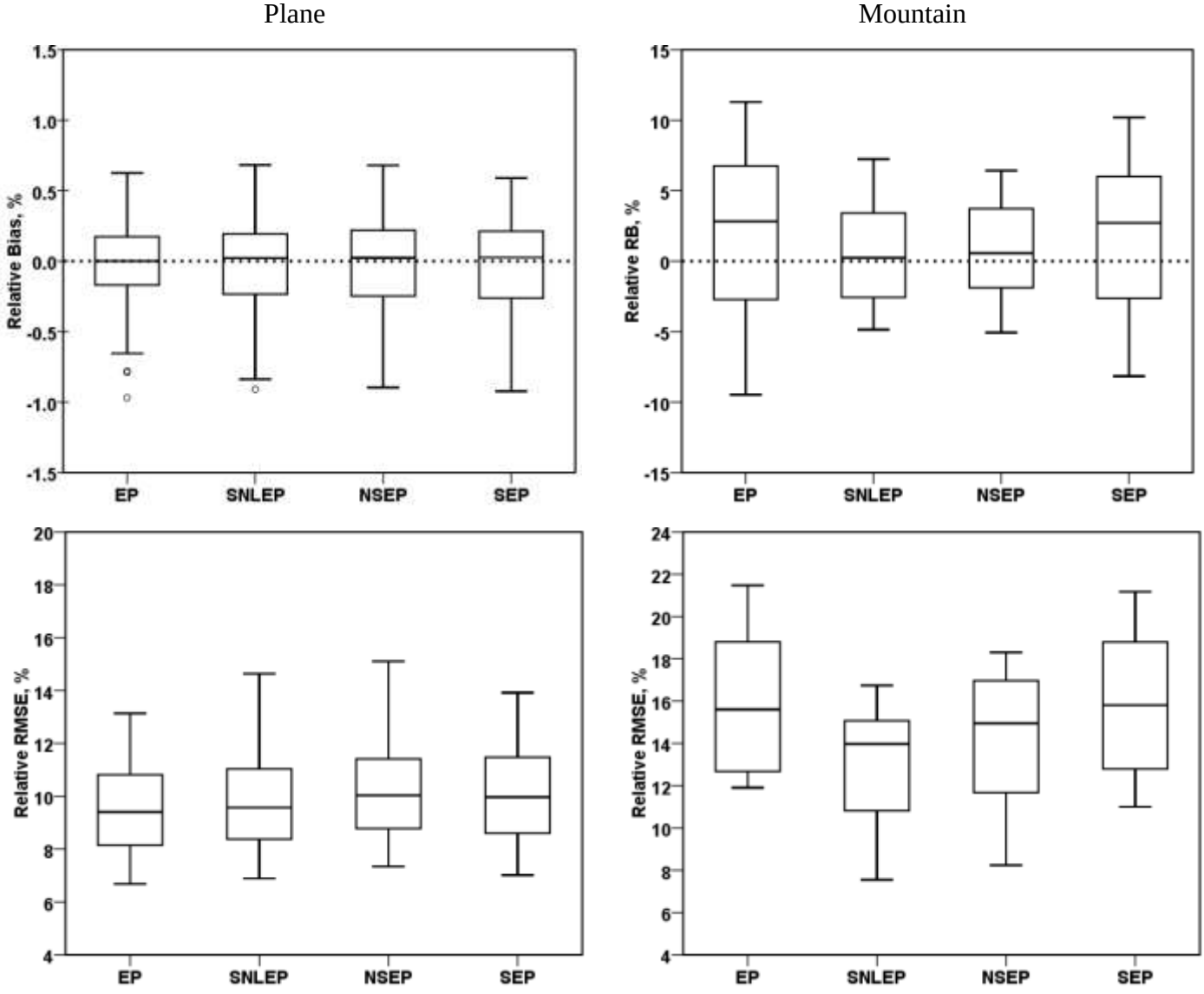

Figure 1. Boxplots of area-specific values of actual relative bias (top) and relative RMSE (bottom) for EP, SNLEP, NSEP and SEP (all expressed in percentage terms) obtained in simulations where m = 100, n_i = 50, and with a Binomial response.

Figure 2 shows the same set of boxplots as in Figure 1, but this time for a Poisson response. Again, the SNLEP outperforms the other predictors in terms of efficiency when there is a strong spatial signal in the data, as when the mountain response function underpins these values.

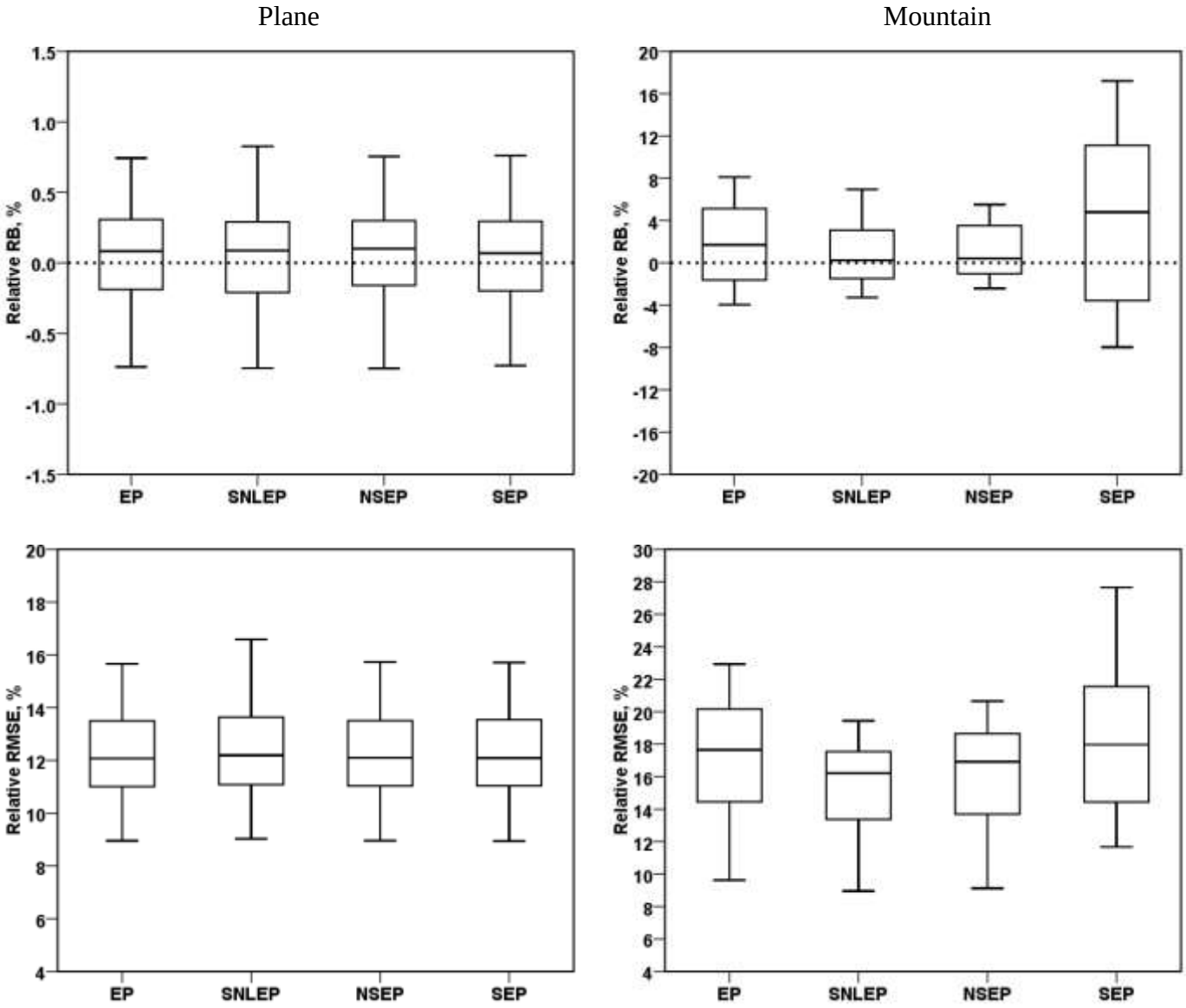

Figure 2. Boxplots of area-specific values of actual relative bias (top) and relative RMSE (bottom) for EP, SNLEP, NSEP and SEP (all expressed in percentage terms) obtained in simulations where $m = 100$, $n_i = 50$, and with a Poisson response.

Turning now to the results presented in Table 2, these show the empirical performance of the MSE estimator (17) for the SNLEP (14). We see that this MSE estimator performs well, showing reasonable bias and good empirical coverage for different types of data except for count data with $n_i = 10$ where both MSE and coverage rates are slightly underestimated. As the sample size increases, the bias performance of the MSE estimator improves. Overall, the MSE estimator (17) appeared to provide a good approximation to the actual MSE in our simulations, and generated confidence intervals close to nominal coverage.

Table 2. The actual (i.e. Monte Carlo) values of the RMSE (TRMSE), the average values of the estimated RMSE (ERMSE), the percentage relative bias (*RB*), the percentage relative RMSE of the MSE estimator (*RE*) and the percentage the coverage rate (CR) of the proposed MSE estimator (17) of SNLEP (14). Values are averages over the small areas.

$f(x)$	$n_i = 10$					$n_i = 50$				
	TRMSE	ERMSE	RB	RE	CR	TRMSE	ERMSE	RB	RE	CR
				Binary data						
Plane	0.067	0.067	-0.05	7.42	95	0.047	0.048	1.05	11.02	96
Mountain	0.076	0.072	-6.65	13.60	95	0.048	0.048	0.19	12.84	96
				Count data						
Plane	0.221	0.202	-8.40	32.64	94	0.129	0.128	-0.82	28.45	95
Mountain	0.179	0.165	-7.91	36.70	94	0.103	0.102	-0.81	30.15	95

Finally, we present simulation results that provide an insight into the behaviour of the hybrid MPQL/REML parameter estimation algorithm described in Section 3.1 and the bootstrap test of spatial non-linearity described in Section 3.2. In particular, Table 3 shows simulation results for parameter estimation using this algorithm under the SNLGLMM with Binomial and Poisson responses, with m = 100 and $n_i = 10$. For the Binomial case we generated the data under the model $\eta_i = -1 + x_i + u_i + \mathbf{z}_i\gamma$ where $x_i \sim U[-1,1]$, with random area effects u_i generated as m independent realizations from a $N\left(0, \sigma_u^2 = 0.0625\right)$ distribution and γ generated as $L = 30$ independent realizations from a $N\left(0, \sigma_\gamma^2\right)$, with $\sigma_\gamma^2 = 0$ in the non-spatial case and $\sigma_\gamma^2 = 1$ in spatially non-linear case. Under these scenarios the population count for area i was generated as the sum of N_i independent binary values y_{ij}, each such that

$\Pr\left(y_{ij}=1\right)=\mu_i=\exp(\eta_i)\{1+\exp(\eta_i)\}^{-1}$. 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: $\eta_i=-1+0.6x_i+u_i+\mathbf{z}_i\gamma$ with x_i, u_i and γ 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.

From Table 3 we see that the hybrid MPQL/REML procedure results in small upward biases in the estimation of the regression model parameters β_0 and small downward biases in the estimation of β_1, as well as the variance component σ_γ, for the Binomial case with a negative bias of around 2% when estimating the variance component σ_u. For Poisson data the results are similar for β_1 and σ_u but now show a downward bias of around 15% for β_0 under the spatially non-linear scenario, while estimation of σ_γ shows a small downward bias. Note that values of %Rel Bias and the %RRMSE cannot be computed in case of a non-spatial relationship with $\sigma_\gamma^2=0$ because the denominator of these summary measures is 0. Overall, we conclude that although parameter estimation via the MPQL/REML algorithm works reasonably well in the case of no spatial relationship $(\sigma_\gamma^2=0)$ there is room for improvement in the spatial non-linear case $(\sigma_\gamma^2=1)$. This could be a topic for future research.

The results of the bootstrap-based spatial non-linearity test at significance levels of $\alpha=0.10,0.05,0.01$ under the null hypothesis are presented in Table 4 for binary data with *m*=49 and 100. In all cases when $\sigma_\gamma^2=0$ and the null hypothesis is true the Type I errors (actual rejection probabilities for null in this case) are close to the nominal value α. These probabilities increase as we move away from the null and as the small area sample sizes increase. The results for Poisson data, which are not presented here for reason of space, show a similar pattern.

Table 3. Simulation results for parameter estimation using hybrid MPQL/REML algorithm under the SNLGLMM - m = 100, n_i = 10, 1000 Monte Carlo simulations.

	β_0	β_1	σ_u	σ_γ
		Binary data		
True Value	-1.000	1.000	0.250	0.000
Average Estimated Value	-0.993	0.998	0.243	0.005
%Rel Bias	0.671	-0.197	-2.771	NA
%RRMSE	13.470	15.327	5.353	NA
True Value	-1.000	1.000	0.250	1.000
Average Estimated Value	-0.980	0.980	0.248	0.989
%Rel Bias	1.967	-2.011	-2.181	-4.589
%RRMSE	98.076	20.521	3.611	26.607
		Poisson data		
True Value	-1.000	0.600	0.250	0.000
Average Estimated Value	-0.975	0.596	0.244	0.001
%Rel Bias	2.489	-0.696	-2.190	NA
%RRMSE	10.692	28.546	2.550	NA
True Value	-1.000	0.600	0.250	1.000
Average Estimated Value	-1.153	0.595	0.246	0.930
%Rel Bias	-14.33	-0.819	-1.587	-10.162
%RRMSE	88.492	72.631	6.401	26.039

Table 4. Rejection probabilities for the null hypothesis for the bootstrap test statistic of Section 3.2 under the SNLGLMM: *m=49,* 100, n_i = 10, 50, 1000 Monte Carlo simulations. Binary data.

Area sample size	Nominal type 1 error rate α	Actual rejection probability for null			
		m=49		m=100	
		$\sigma_\gamma^2 = 0$	$\sigma_\gamma^2 = 1$	$\sigma_\gamma^2 = 0$	$\sigma_\gamma^2 = 1$
n_i = 10	0.10	0.112	0.581	0.095	0.593
	0.05	0.065	0.552	0.055	0.571
	0.01	0.015	0.523	0.015	0.555
n_i = 50	0.10	0.092	0.730	0.095	0.745
	0.05	0.044	0.710	0.050	0.735
	0.01	0.012	0.690	0.010	0.700

6. Application to NSSO data

In this Section we show how two of the estimators (EP and SNLEP) described in this paper can be used to obtain estimates of the proportion of poor households at small area levels in the State of Uttar Pradesh in India, using 2011-12 data from the Household Consumer Expenditure (HCE)

Survey that was carried out by National Sample Survey Office (NSSO) of India and also data from the Population Census 2011. In particular, since the underlying variable is binary (poor / not poor) we used a Binomial model for the area level counts. Small areas are defined as the different districts of the State of Uttar Pradesh in India. Poverty is an important and persistent social issue in India. Existing data based on socio-economic surveys produce state and nationally representative poverty estimates but these surveys cannot be used directly to generate reliable district level estimates. This motivated us to apply the SAE methods developed in this paper to poverty estimation in India. However, we also note that the approach we describe is a general methodology for small area prediction of count data, and is not limited to poverty estimation based on socio-economic survey data. The sampling design used in the NSSO data is stratified multi-stage random sampling with districts as strata, villages as first stage units and households as the second stage units. A total of 5916 households were surveyed from the 71 districts of the Uttar Pradesh. The district-wise sample size ranges from 32 to 128 with average of 83. It is evident that district level sample sizes are very small with a very low (0.0002) average sampling fraction. As a consequence it is difficult to derive reliable estimates and their standard errors at district level, and use of model-based SAE is obvious for such cases. The target variable used for the study was an indicator for whether a household is poor or not, defined as monthly per capita consumer expenditure below the 2011-12 poverty line (Rs. 768), as specified by the Planning Commission, Government of India. Many countries use 60% of median household income in order to define a poverty line. However, in India the poverty line is defined as the bare minimum income required to provide basic food requirements, and does not account for other essentials such as health care and education. It is an economic benchmark used by the government to indicate economic disadvantage and to identify individuals and households in need of government assistance and aid. It is published by the Government of India, for every state and for rural and urban sectors separately. The poverty indicator of interest is then the proportion of poor households, often referred to as the head count ratio (HCR), at the district level.

Auxiliary variables (covariates) are drawn from the Population Census 2011. There are around 50 covariates available from this source to consider for modelling. We use Principal Component Analysis (PCA) to derive a composite score for selected groups of variables. In particular, we consider three groups for PCA, namely G1, G2 and G3. The first group (G1) is based on gender-

wise literacy rate and gender-wise worker population. The first principal component for this first group (G11) explained 51% of the variability in the dataset, while adding the second component (G12) explained 85%. The second group (G2) is based on gender-wise main worker, gender-wise main cultivator and gender-wise main agricultural labourers. The first principal component (G21) for this second group explained 49% of the variability in the dataset, while adding the second component (G22) explained 67%. Finally, the third group (G3) is derived from gender-wise marginal cultivator and gender-wise marginal agriculture labourers. The first principal component (G31) for this third group explained 61% of the variability in the dataset, while adding the second component (G22) explained 78%. After fitting a generalized linear model using direct survey estimates of proportion of poor households as the response variable and these six variables (i.e. G11, G12, G21, G22, G31 and G32) as covariates, only three significant covariates namely G11, G21 and G31 were used to generate district level estimates of the HCR. Initially, this model was fitted using the glm() function in R and specifying the family as "binomial" and the district specific sample sizes as the weight, see R Development Core Team (2013). The residual deviance and AIC values for this fitted model were 327.18 and 636.89, respectively. The actual model to estimate the HCR in the small areas of interest can be written as

$$g(\boldsymbol{\pi}) = \boldsymbol{\eta} = f(x_1, x_2) + \beta_3 G11 + \beta_4 G21 + \beta_5 G31 + \mathbf{u},$$

where (x_1, x_2) are the coordinates (latitude and longitude) of the centroid of the districts of the State of Uttar Pradesh and $f(\cdot)$ is an unknown smooth bivariate function. We estimate this function using *P*-splines. The choice of knots in two dimensions is more challenging than in one. Following the advice of Ruppert *et al.* (2003), and carrying out estimation with a different number *L* of knots, we finally chose $L = 30$, since we found that the fit of the spline stabilizes after this number of knots. The diagnostic procedure to test for spatial nonlinearity (i.e. the hypothesis $H_{0\sigma_\gamma^2} : \sigma_\gamma^2 = 0$ versus the one-sided alternative $H_{1\sigma_\gamma^2} : \sigma_\gamma^2 > 0$) in the data was also applied and with $L = 30$ it was significant (*p* value of 0.0001) and so the null hypothesis of lack of spatial structure in the data was rejected.

We estimated the district level HCR using the EP and the SNLEP. For the GLMM (1), the estimate of the between area variance component was $\hat{\sigma}_u^2 = 0.265$, whereas for the SNLGLMM

(13) the corresponding estimates of the model variance components were $\hat{\sigma}_u^2 = 0.025$ and $\hat{\sigma}_\gamma^2 = 0.105$. We next used diagnostic measures to check model assumptions as well as the empirical performance and underlying assumptions of the model-based SAE methods. The GLMM and SNLGLMM are based on distributional assumption, i.e. the random effects in both models are independently and identically distributes values from a normal distribution with mean zero. Standard diagnostics such as histograms and normal probability (*q*-*q*) plots were used for this purpose. Although these results are not reported here, they indicated that these distribution features hold for both the GLMM and the SNLGLMM when fitted to NSSO data. We also observed that the district level residuals are randomly distributed, and that their line of best fit does not significantly differ from the line *y*=0 in all cases. Model diagnostics were therefore satisfactory for both the GLMM and the SNLGLMM.

To validate the reliability of the small area estimates produced by the EP and SNLEP estimation methods we examined bias diagnostics and coefficients of variation (CV). Bias diagnostics are used to investigate if the small area estimates are less extreme when compared to the direct survey estimates, when these are available. In addition, if the direct estimates are unbiased, their regression on the true values should be linear and correspond to the identity line. If the small area estimates are also close to the true values the regression of the direct estimates on the model-based estimates should be similar. We plotted direct estimates on the vertical axis against the small area estimates on the horizontal axis and looked for divergence of the fitted regression line from *y*= *x* by testing for intercept = 0 and slope = 1. See Brown *et al.* (2001) and Chandra *et al*. (2011). These bias scatter plots for EP and SNLEP are displayed in Figure 3.

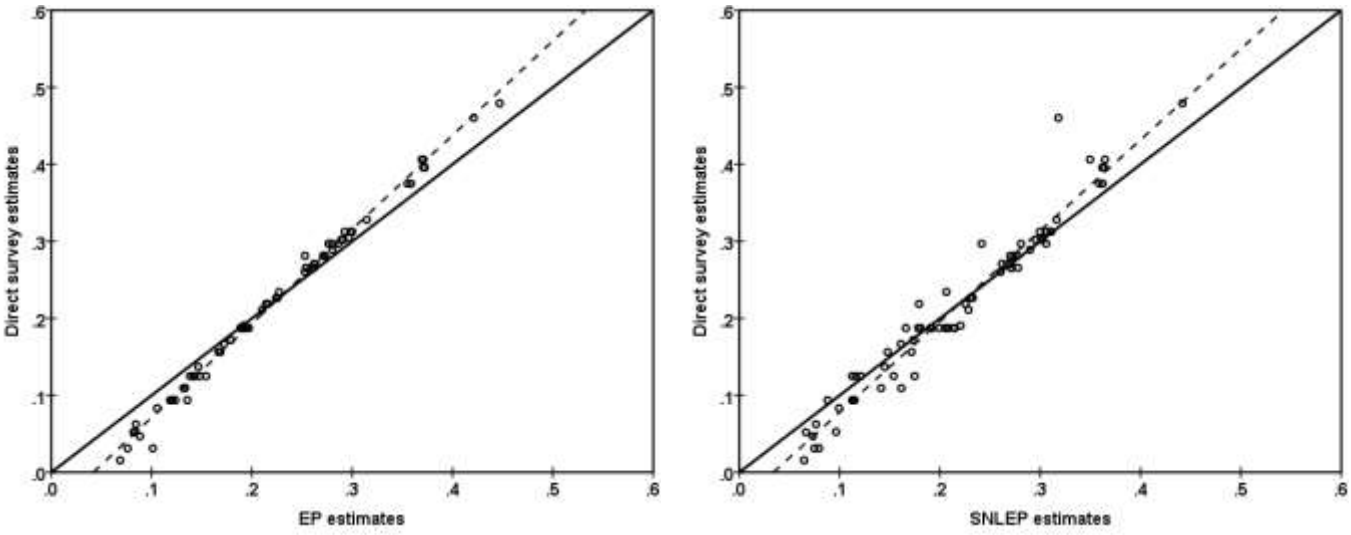

Figure 3. Bias diagnostics plots with $y = x$ line (solid) and regression line (dotted) for EP (left) and SNLEP (right).

The plots in Figure 3 reveal that the small area estimates generated by both the EP and the SNLEP methods are less extreme when compared to the direct survey estimates, demonstrating the typical SAE outcome of shrinking more extreme values towards the average. That is, the estimates of poverty incidence generated by both EP and SNLEP methods lie along the line $y= x$ for most of the districts which indicates that they are approximately design unbiased. We computed the per cent CV to assess the improved precision of the model-based small area estimates (EP and SNLEP) compared to the direct survey estimates. The CV shows the sampling variability as a percentage of the estimate. Estimates with large CVs are considered unreliable (i.e. smaller is better). There are no internationally accepted tables available that allow us to judge what is "too large". Figure 4 shows the district-wise values of percentage CV for direct (DIR), EP and SNLEP estimation methods. Estimates of proportions of poor households (poverty incidence) in Uttar Pradesh by District obtained via the DIR, EP and SNLEP methods along with their percentage CVs are set out in Table 5.

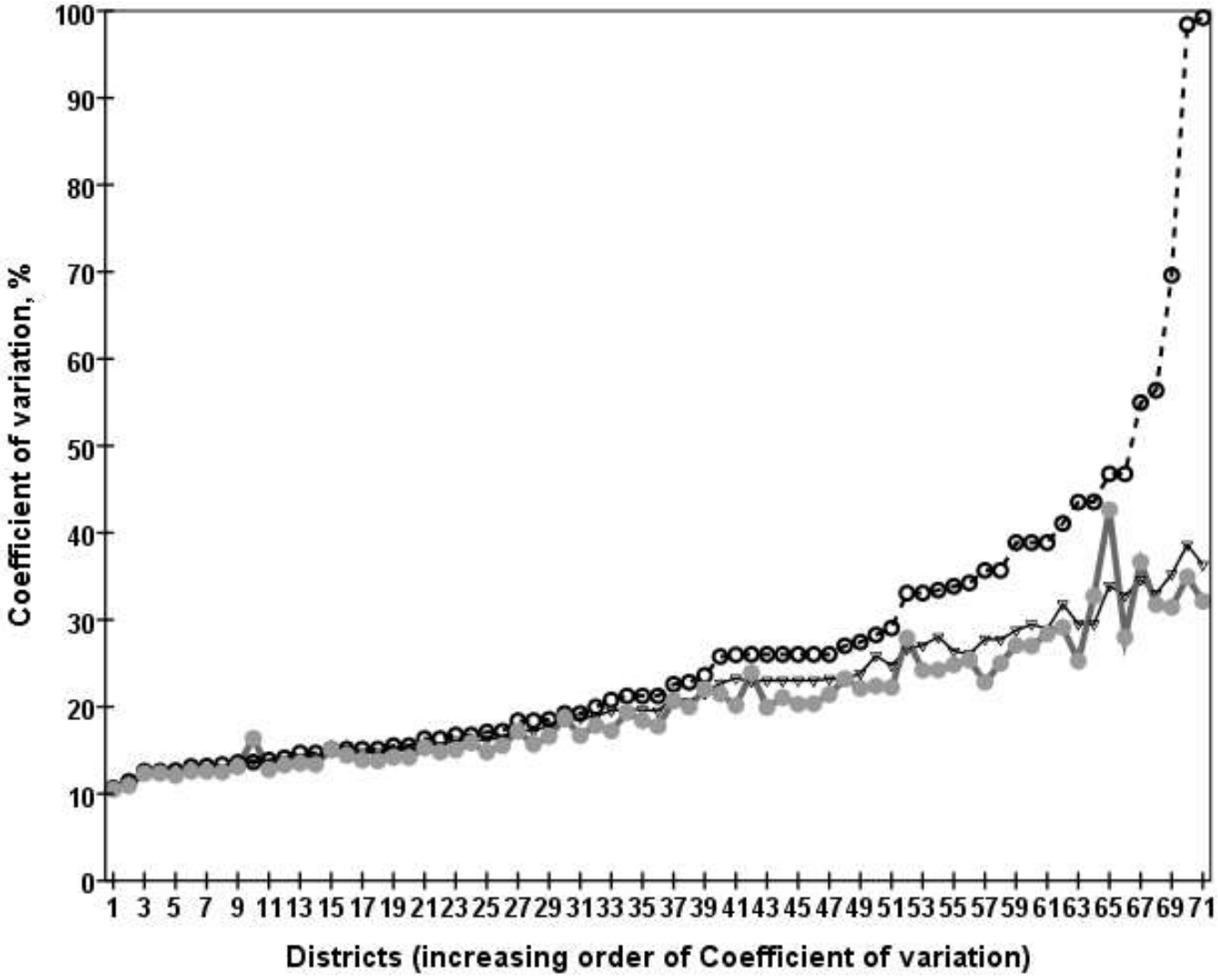

Figure 4. District-wise percentage coefficient of variation (CV, %) for direct (dotted line, o), EP (thin line, ∇) and SNLEP (solid line, •) method applied to NSSO data.

The results presented in Table 5 and displayed in Figure 4 clearly show that direct survey estimates of poverty incidence are unstable, with CVs varying from 10.64 to 99.22 % with an average of 27.03 %. Further, the CVs of these direct estimates are greater than 20% and 40% in 39 and 11 (out of 71) districts respectively. In contrast, the two model-based small area predictors (EP and SNLEP) perform better than the direct estimate. The average CV values for EP and SNLEP are 21.30 % and 20.18 % respectively. Furthermore, the CV of the SNLEP is smaller than that of the EP in 62 out of the 71 districts. We therefore conclude that the SNLEP is the best performing estimation method of the three that we investigated for this application.

Although the relevant results are not presented here, we also note that the SNLEP had narrower confidence intervals than the EP.

Finally, in Figure 5 we present the poverty map produced by the SNLEP estimation method which shows the estimated proportions of poor households in different districts of Uttar Pradesh. In Figure 6 we present the corresponding map of the percentage coefficients of variation of the SNLEP in the different districts. In Figure 5, districts with light colouring are estimated to have a low proportion of poor households. Similarly, in Figure 6 a light colour corresponds to districts with smaller values of percentage coefficient of variation. For example, in the western part of Uttar Pradesh there are many districts with low poverty incidence (light colour), such as Saharanpur, Hathras, Meerut, Baghpat, Muzaffarnagar, Bulandshahar etc (Figure 5). Values of percentage coefficient of variation are large (dark colouring) for these districts (Figure 6). In eastern region and in the Bundelkhand region (north-east part of map) there are number of districts (for example, Azamgarh, Sitapur, Chitrakoot, Bahraich, Siddharthnagar, Banda, Fatehpur, Basti and Kaushambi etc) with high (dark colour) levels poverty incidence and with corresponding smaller values of coefficients of variation. In general, we note that districts with high poverty incidence tend to have higher precision (i.e. lower coefficient of variation) compared to the districts with low poverty incidence. Overall, poverty maps like that displayed in Figure 5 are very useful for policy planners and administrators who require accurate information for effective financial and administrative decisions.

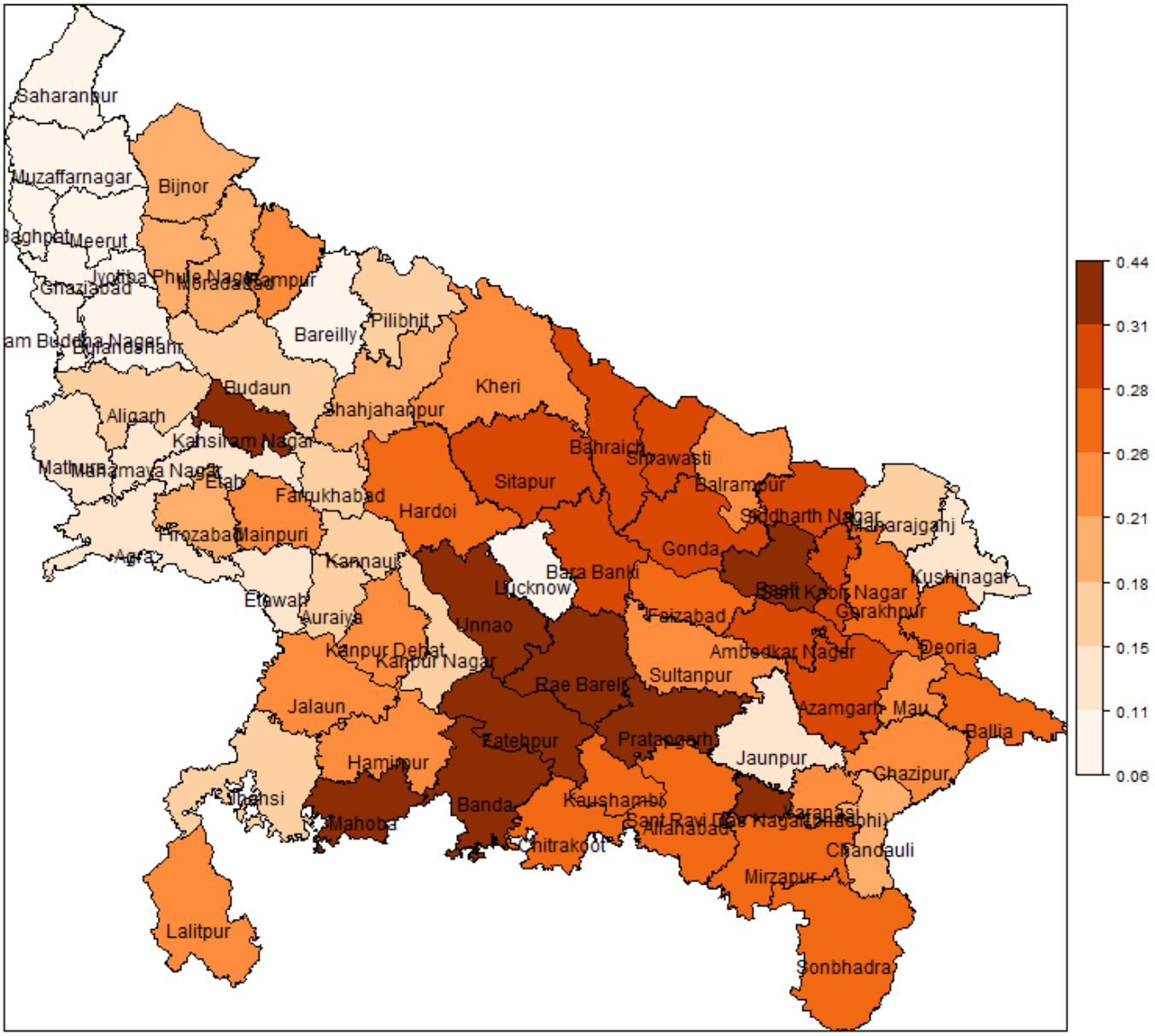

Figure 5. District-wise distribution of poverty incidence in the state of Uttar Pradesh: SNLEP estimates.

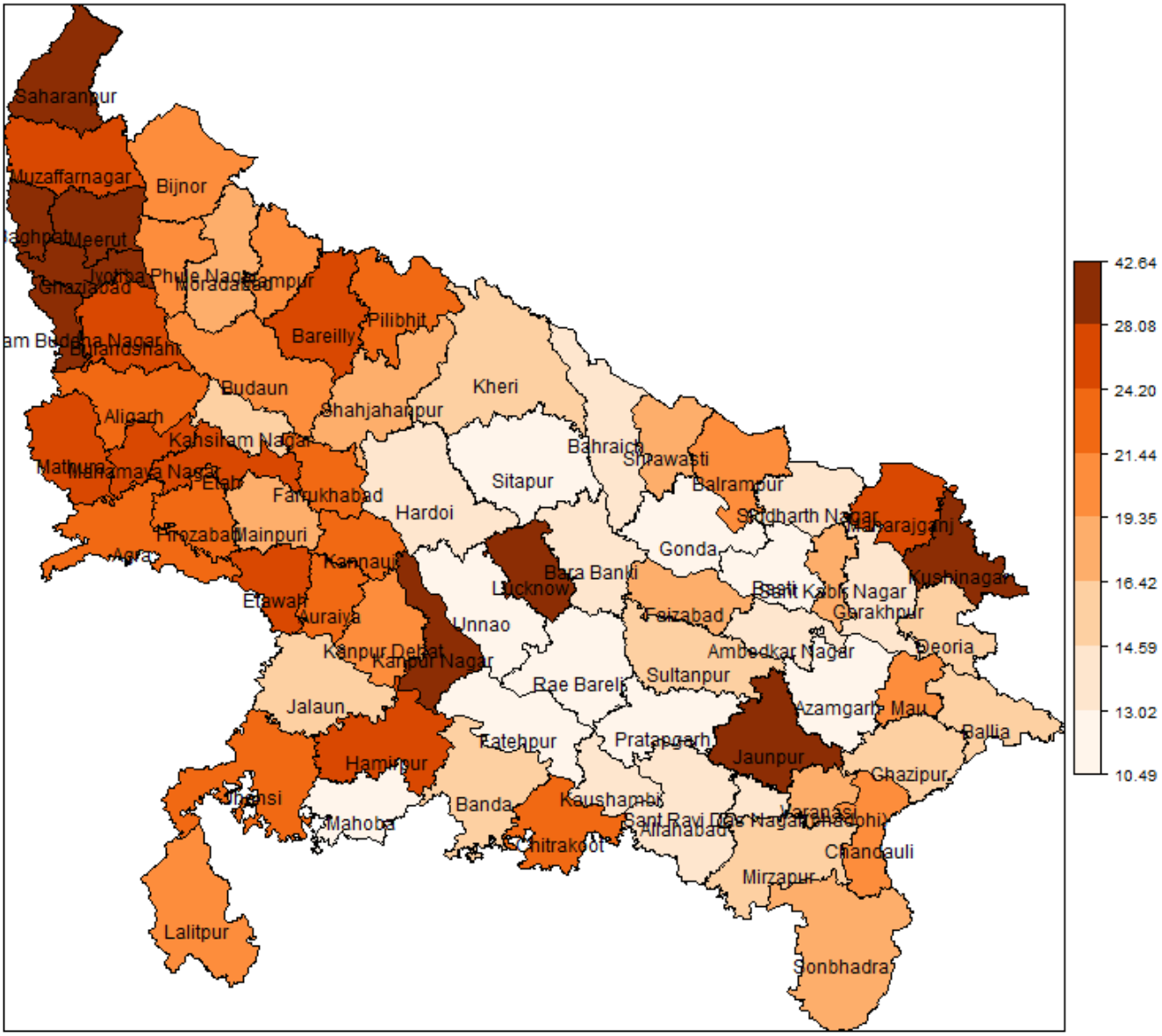

Figure 6. District-wise distribution of percent coefficients of variation in the state of Uttar Pradesh: SNLEP estimates.

Table 5. District wise sample size, direct estimate (DIR), EP and SNLEP, percentage coefficient of variation of DIR, EP and SNLEP for the poverty incidence in 2011-12 obtained from NSSO data.

District	Sample size	Estimated poverty incidence			Coefficient of variation, %		
		DIR	EP	SNLEP	DIR	EP	SNLEP
Saharanpur	96	0.05	0.08	0.07	43.54	29.51	32.73
Muzaffarnagar	128	0.06	0.08	0.08	34.23	26.08	25.38
Bijnor	96	0.19	0.19	0.18	21.25	19.71	19.35
Moradabad	128	0.19	0.19	0.18	18.40	17.39	17.19
Rampur	64	0.23	0.23	0.21	22.59	21.11	20.68
Jyotiba Phule Nagar	64	0.19	0.19	0.17	26.02	22.92	23.91
Meerut	64	0.02	0.07	0.06	99.22	36.32	32.12
Baghpat	32	0.09	0.14	0.09	54.96	34.60	36.65
Ghaziabad	64	0.05	0.09	0.07	56.37	33.01	31.74
Gautam B Nagar	32	0.03	0.10	0.08	98.43	38.62	34.93
Bulandshahr	96	0.08	0.11	0.10	33.85	26.38	24.82
Aligarh	95	0.14	0.15	0.15	25.77	22.71	21.46
Mahamaya Nagar	64	0.03	0.08	0.07	69.60	35.23	31.43
Mathura	64	0.13	0.14	0.11	33.07	26.70	27.88
Agra	96	0.13	0.14	0.12	27.00	23.12	23.21
Firozabad	64	0.22	0.22	0.18	23.62	21.52	22.04
Etah	64	0.09	0.12	0.11	38.86	28.79	27.05
Mainpuri	64	0.30	0.28	0.24	19.24	18.40	18.73
Budaun	96	0.17	0.17	0.16	22.82	20.61	19.98
Bareilly	95	0.05	0.08	0.10	43.53	29.53	25.27
Pilibhit	64	0.13	0.14	0.15	33.07	27.07	24.19
Shahjahanpur	96	0.19	0.19	0.19	21.25	19.65	18.39
Kheri	128	0.28	0.27	0.28	14.13	13.86	13.29
Sitapur	128	0.31	0.30	0.30	13.11	12.96	12.58
Hardoi	128	0.30	0.29	0.28	13.60	13.40	13.10
Unnao	96	0.40	0.37	0.36	12.61	12.59	12.31
Lucknow	64	0.19	0.19	0.21	26.02	23.02	19.93
Rae Bareli	128	0.33	0.31	0.32	12.65	12.55	12.10
Farrukhabad	64	0.17	0.18	0.17	27.44	23.85	22.12
Kannauj	64	0.19	0.19	0.19	26.02	23.03	21.05
Etawah	64	0.09	0.12	0.11	38.86	29.50	26.99
Auraiya	64	0.16	0.17	0.17	29.05	24.75	22.21
Kanpur Dehat	64	0.11	0.13	0.16	35.67	27.76	22.81
Kanpur Nagar	64	0.19	0.19	0.21	26.02	23.02	20.29
Jalaun	64	0.11	0.13	0.14	35.67	27.72	25.00
Jhansi	64	0.09	0.12	0.11	38.86	28.93	28.35
Lalitpur	32	0.13	0.15	0.12	46.77	33.93	42.64
Hamirpur	32	0.22	0.21	0.23	33.41	28.04	24.22
Maharajganj	32	0.13	0.15	0.17	46.77	32.77	27.99
Banda	64	0.41	0.37	0.35	15.11	15.20	15.10
Chitrakoot	32	0.28	0.25	0.27	28.26	25.86	22.39
Fatehpur	96	0.40	0.37	0.36	12.61	12.66	12.34
Pratapgarh	128	0.38	0.36	0.36	11.41	11.36	10.91
Kaushambi	63	0.46	0.42	0.32	13.64	13.80	16.36
Allahabad	128	0.27	0.26	0.27	14.70	14.23	13.49
Bara Banki	96	0.30	0.29	0.30	15.51	15.10	14.18
Faizabad	64	0.27	0.25	0.28	20.78	19.55	17.22
Ambedkar Nagar	96	0.31	0.30	0.31	15.14	14.78	13.87
Sultanpur	128	0.21	0.21	0.23	17.10	16.24	14.79

Bahraich	96	0.30	0.29	0.30	15.51	15.09	14.15
Shrawasti	64	0.31	0.29	0.31	18.54	17.87	16.61
Balrampur	63	0.19	0.19	0.22	25.97	23.26	20.15
Gonda	128	0.29	0.28	0.29	13.86	13.54	12.78
Siddharth Nagar	96	0.31	0.30	0.31	15.14	14.78	13.77
Basti	96	0.48	0.45	0.44	10.64	10.82	10.49
Sant Kabir Nagar	64	0.30	0.28	0.31	19.24	18.70	16.69
Mahoba	96	0.38	0.36	0.36	13.18	13.08	12.58
Gorakhpur	128	0.27	0.26	0.27	14.70	14.26	13.34
Kushinagar	128	0.23	0.22	0.23	16.33	15.62	15.29
Deoria	96	0.27	0.26	0.27	16.75	16.12	15.05
Azamgarh	128	0.30	0.30	0.30	13.35	13.07	12.50
Mau	64	0.19	0.20	0.21	26.02	23.01	20.34
Ballia	96	0.27	0.26	0.26	16.75	16.10	15.86
Jaunpur	128	0.19	0.19	0.21	18.40	17.20	15.69
Ghazipur	128	0.23	0.23	0.23	16.33	15.58	14.78
Chandauli	64	0.19	0.19	0.20	26.02	23.22	21.39
Varanasi	96	0.19	0.19	0.21	21.25	19.52	17.77
Sant Ravi Das Nagar	64	0.41	0.37	0.36	15.11	15.09	14.41
Mirzapur	96	0.26	0.25	0.26	17.20	16.48	15.54
Sonbhadra	64	0.28	0.27	0.27	19.98	18.89	17.87
Kansiram Nagar	32	0.16	0.17	0.15	41.08	31.78	29.13
Mean					**27.03**	**21.30**	**20.18**

7. Concluding remarks

This paper describes a spatially non-linear (or nonparametric) extension of the area level version of the generalized linear mixed model (SNLGLMM) and considers SAE under this model. The corresponding estimator is referred to as the spatially non-linear empirical predictor (SNLEP) for small areas. This estimator can accommodate situations where the functional form of the spatial relationship between the variable of interest and the covariates is unknown. A bootstrap based procedure for testing for spatial nonlinearity in the data is also described. An analytical MSE estimator is also proposed for the SNLEP. Empirical evaluations based on simulation studies indicate that the proposed SNLEP method is more efficient than other model-based SAE methods developed under the area level generalized linear mixed model when there is a spatial relationship between the variable of interest and the covariates. The proposed analytic MSE estimator also performed reasonably well, with good coverage performance for nominal confidence intervals based on it. We also applied the SNLEP to real survey data to estimate the Head Count Rate (HCR) poverty indicator values for the districts of the State of Uttar Pradesh in India and produced a poverty map of these districts based on these HCR estimates. These estimates and their spatial distribution will be useful for various Departments and Ministries in Government of India as well as International organizations for their policy research and strategic

planning. They are also useful for budget allocation and intervention of targeted welfare for below poverty line (BPL) households. Furthermore, the methodology developed in this paper and demonstrated in the application presented in this paper can be used generally for calculating reliable area level estimates of counts and rates. Further extension of this work could be a comparison of the MSE estimator proposed in this paper with alternative bootstrap-based MSE estimators, since the latter may offer a better and more stable approximation to the actual MSE. MSE estimators based on block bootstrapping methods, e.g. the random effects block bootstrap of Chambers and Chandra (2013), are worth considering in this regard.

Acknowledgement

The authors would like to acknowledge the valuable comments and suggestions of the Editor, the Associate Editor and four referees. These led to a considerable improvement in the paper. The work of Hukum Chandra was carried out under an ICAR-National Fellow Project at ICAR-IASRI, New Delhi, India. The work of Nicola Salvati carried out with the support of the project InGRID 2 (Grant Agreement No 312691, EU).

References

Brown, G., Chambers, R., Heady, P. and Heasman, D., 2001. Evaluation of small area estimation methods: an application to unemployment estimates from the UK LFS. Proceedings of Statistics Canada Symposium. Achieving data quality in a statistical agency: A Methodological Perspective.

Boubeta M., Lombardía M.J. and Morales D., 2017. Poisson mixed models for studying the poverty in small areas. Computational Statistics and Data Analysis 107, 32-47.

Boubeta M., Lombardía M.J. and Morales D., 2016. Empirical best prediction under area level Poisson mixed models. TEST 25, 548–569.

Chambers, R. and Chandra, H., 2013. A random effect block bootstrap for clustered data. Journal of Computational and Graphical Statistics 22 (2), 452-470.

Chandra, H., Salvati, N. and Chambers, R., 2015. A spatially nonstationary Fay-Herriot model for small area estimation. Journal of Survey Statistics and Methodology 3, 109-135.

Chandra, H. and Salvati, N., 2018. Small area estimation for count data under a spatial dependent aggregated level random effects model. Communications in Statistics - Theory and Methods, 47 (5), 1234 -1255.

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

Chandra, H., Salvati, N. and Sud, U.C., 2011. Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India-an application of small area estimation technique. Journal of Applied Statistics 38(11), 2413-2432.

Das, K., Jiang, J., Rao, J.N.K., 2004. Mean squared error of empirical predictor. The Annals of Statistics 32, 818-840.

Datta, G.S., Lahiri, P., Maiti, T., Lu, K.L., 1999. Hierarchical Bayes estimation of unemployment rates for the US states. Journal of the American Statistical Association 94, 1074-1082.

Du, P., Jiang, Y., Wang, Y. 2011. Smoothing spline ANOVA frailty model for recurrent event data. Biometrics 67, 1330-1339.

Eilers, P. H. C. and B. D. Marx, 1996. Flexible smoothing with B-splines and penalties. Statistical Science 11 (2), 89–121.

Fay, R. E. and Herriot, R. A., 1979. Estimation of income from small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association 74, 269-277.

Franco, C. and Bell, W.R., 2013. Applying bivariate binomial/logit normal models to small area estimation. Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association, Alexandria, VA, 690-702.

Green, P.J. and Silverman, B.W., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman & Hall Ltd.

Jiang, J., 2003. Empirical best prediction for small-area inference based on generalized linear mixed models. Journal of Statistical Planning and Inference 111, 117–127.

Johnson, F.A., Chandra, H., Brown, J., and 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. Journal of Official Statistics 26 (2), 341–359.

Kammann, E.E. and Wand, M.P., 2003. Geoadditive models. Journal of Royal Statistical Society, C 52, 1-18.

Liu, B., Lahiri, P. and Kalton, G., 2014. Hierarchical Bayes modeling of survey-weighted small area proportions. Survey Methodology 40 (1), 1-13.

Lopez-Vizcaino, E., Lombardìa, M. and Morales, D., 2013. Multinomial-based small area estimation of labour force indicators. Statistical Modelling 13, 153–178.

Manteiga, G.W., Lombardìa, M.J., Molina, I., Morales, D., and Santamarìa,L., 2007. Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. Computational Statistics and Data Analysis 51, 2720-2733.

McGilchrist, C.E., 1994. Estimation in generalized mixed models. Journal of the Royal Statistical Society B 56, 61-69.

Mercer, L., Wakefield, J., Chen, C. and Lumley T., 2014. A comparison of spatial smoothing methods for small area estimation with sampling weights. Spatial Statistics 8, 69–85.

Molina, I., Saei, A. and José Lombardìa, M., 2007. Small area estimates of labour force participation under a multinomial logit mixed model. Journal of the Royal Statistical Society A 170, 975–1000.

Mourra, N., Lesurtel, M. & Flejou, J.F., 2006. Chronic schistosomiasis: an incidental finding in sigmoid volvulus. Journal of Clinical Pathology 59, 111.

Opsomer, J.D., Claeskens, G., Ranalli, M.G., Kauermann, G. and Breidt, F.J., 2008. Nonparametric small area estimation using penalized spline regression. Journal of the Royal Statistical Society B 70, 265-286.

Prasad, N.G.N., Rao, J.N.K., 1990. The estimation of the mean squared error of small-area estimators. Journal of the American Statistical Association 85, 163-171.

Pratesi, M., Salvati, N., 2008. Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Statistical Methods and Applications 17, 114-131.

Pratesi, M., Ranalli, M.G., Salvati, N., 2009. Nonparametric M-quantile regression using penalised splines. Journal of Nonparametric Statistics 21, 287-304.

Rao, J.N.K., 2011. Impact of frequentist and Bayesian methods on survey sampling practice: a selective appraisal. Statistical Science 26 (2), 240–256.

Rao, J.N.K., 2003. Small Area Estimation. Wiley, New York.

Rao, J. N. K. and Molina, I., 2015. Small Area Estimation. John Wiley & Sons. Inc., New Jersey, 2nd edition.

R Development Core Team (2013), R: a language and environment for statistical computing. Vienna, Austria: the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at http://www.R-project.org/.

Ruppert D., Wand M.P., Carroll R.J., 2003. Semiparametric Regression. Cambridge University Press, New York.

Saei, A. and Chambers, R., 2003. Small area estimation under linear and generalized linear mixed models with time and area effects, Methodology working paper No. M03/15, Southampton Statistical Sciences Research Institute, University of Southampton, UK.

Saei, A. and McGilchrist, C., 1998. Longitudinal threshold models with random components. The Statistician Series D 47, 365-375.

Schall, R., 1991. Estimation in generalized linear models with random effects. Biometrika 78, 719-727.

Torabi, M. and Shokoohi, F., 2015. Non-parametric generalized linear mixed models in small area estimation. The Canadian Journal of Statistics 43(1), 82-96.

Tzavidis, N., Salvati, N., Pratesi, M. and Chambers, R., 2008. M-quantile models with application to poverty mapping. Statistical Methods and Applications 17(3), 393-411.

Ugarte, M.D., Militino, A.F. and Goicoa, T., 2009. Benchmarked estimates in small areas using linear mixed models with restrictions. TEST 18(2), 342-364.

Wahba, G. and Wang, Y., 1990. When is the optimal regularization parameter insensitive to the choice of the loss function. Communications in Statistics – Theory and Methods 19, 1685-1700.

Wand, M., 2003. Smoothing and mixed models. Computational Statistics 18, 223–249.