The Canadian Journal of Statistics

Vol. xx, No. yy, 2019, Pages 1–28

La revue canadienne de statistique

Spatial generalized linear mixed models in small area estimation

Mahmoud Torabi*

Department of Community Health Sciences, University of Manitoba, S113 Medical Services Building, 750 Bannatyne Ave., Winnipeg, MB, Canada R3E 0W3.

Key words and phrases: Generalized linear mixed model; maximum likelihood estimation; parametric bootstrap; small area estimation; spatial model; Taylor expansion.

MSC 2010: 62D99; 62H11; 62P10; 62F40

Abstract: In survey sampling, policy decisions with respect to the allocation of resources to subgroups of a population depend on reliable predictors of their underlying parameters. However, in some subgroups, called small areas due to small sample sizes relative to the corresponding population sizes, the information needed to estimate reliable predictors is typically not available. Consequently, coarser scale data is used to predict the characteristics of the small areas. Mixed models are the primary tools in small area estimation (SAE) to also borrow information from alternative sources (e.g., previous surveys, administrative, and census datasets). In many circumstances, the small area predictors are associated with their locations. For instance, in the case of chronic disease or cancer (e.g., esophageal cancer), it is important for policy makers to understand the spatial patterns of disease in order to determine small areas with high risk of disease and establish prevention strategies. The literature concerning SAE assuming spatial random effects is sparse and what is available is mostly in the context of spatial linear mixed models. In this

© 2019 Statistical Society of Canada / Société statistique du Canada

paper, small area models are proposed in the class of spatial generalized linear mixed models to predict small area predictors as well as the second-order unbiased mean squared prediction errors estimation of small area predictors using Taylor expansion and parametric bootstrap approaches. Evaluation of the performance of the proposed approach is completed through simulation studies and by a real application of the models to an esophageal cancer dataset in Minnesota, USA. *The Canadian Journal of Statistics* xx: 1–28; 2019 © 2019 Statistical Society of Canada

Résumé: Insérer votre résumé ici. **Abstract in French**. *La revue canadienne de statistique* xx: 1–28; 2019 © 2019 Société statistique du Canada

Email address for corresponding author

1. INTRODUCTION

Sample surveys are administered with the intention of gaining information to provide reliable estimates for the finite population characteristics, such as totals or means. The total sample size determines the approaches used to compose these estimators (direct survey estimators). Over the past few decades, there has been a growing need to use the same sample survey data to get estimations for sub-populations, such as counties or gender-age groups. In the literature, these sub-populations for which reliable estimations are required are called small areas due to small sample sizes relative to the corresponding population sizes. However, these small areas result in traditional area-specific direct estimators with inadequate precision. Survey researchers are developing methods to provide more reliable estimations for small areas since policy decisions regarding

---

* *Author to whom correspondence may be addressed.*
E-mail: Mahmoud.Torabi@umanitoba.ca

the implementation of specific projects are often based on using estimations of the underlying characteristics.

Nowadays, small area estimation (SAE) is commonly used in the planning of health, social and other services, and for allocating government funds in Canada, the United States of Americans (USA), and many European countries. Model based estimators (Pfeffermann, 2013; Rao & Molina, 2015; Jiang, 2017, Chap. 4) have been proposed to borrow strength from other sources such as administrative, survey, and census datasets. Linear mixed models (LMM) (Searle, Casella, & McCulloch, 2006) and generalized linear mixed models (GLMM) (McCulloch & Searle, 2001) are the main models used for SAE. Specifically, small area models can be categorized into two broad types. The first type is area-level models, which relate small area direct estimates to the area-specific covariates. These models are typically used when unit-level data are unavailable. The second type is unit-level models, which relate the unit values of a study variable to the associated unit-level covariates with known area means and area-specific covariates. Rao & Molina (2015) gave an extensive review of model-based small area estimation under area-level and unit-level models.

Most of the models in SAE assume the small areas are independent from one another. However, the small area parameters are associated with their locations in many situations. For instance, when dealing with chronic disease or cancer, it is important for policy makers (and the public) to understand the spatial patterns of a disease in order to determine small areas with a high risk of disease so prevention strategies can be implemented. The literature concerning SAE assuming spatial random effects is limited. The spatio-temporal area-level models in SAE were studied by Singh, Shukla, & Kundu (2005). Specifically, they employed the simultaneous autoregressive (SAR) model to explain the spatial random effects, and to estimate the model parameters, they utilized the maximum likelihood estimation (MLE) approach. They heuristically derived the second-order mean squared prediction error (MSPE) of the empirical best linear unbiased prediction (EBLUP) of the small area mean, and using Taylor expansion they found the corresponding second-order unbiased estimator of the MSPE. Recently, Chandra, Salvati, & Chambers (2017) studied non-stationary spatial SAR models for small area count outcomes in the area-level model and heuristically provided MSPE estimation of prediction of small area parameters using Taylor expansion and parametric bootstrap methods. In the light of this work, Chandra & Salvati (2018) also studied spatial SAR models for small area binomial outcomes in the area-level model using penalized quasi-likelihood method to estimate the model parameters and by also providing MSPE estimation of prediction of small area parameters based on Taylor expansion. It is well-known that the conditional spatial dependence parameter defined through the SAR model is not consistent (Schabenberger & Gotway, 2004). A spatial area-level model was studied by Molina, Salvati, & Pratesi (2009), where they found the EBLUP of the small area mean and used bootstrap MSPE estimators for the spatial EBLUP. Recently, Torabi & Jiang (2019) studied the spatial area-level model and employed a conditional autoregressive (CAR) model to explain the spatial random effects. They used the generalized weighted least squares procedure to estimate the regression coefficients. The variance components, including CAR parameters were estimated through the restricted maximum likelihood (REML). Here, the CAR parameters are consistent unlike with the SAR model. They also rigorously obtained the MSPE of the EBLUP of the small area means and also found the estimator of MSPE of the EBLUP of small area means. To the best of our knowledge, in the context of SAE, in the literature there are no studies that have rigorously obtained the MSPE of the empirical best prediction (EBP) of the small area predictors and the estimator of the MSPE of the EBP of the small area predictors through spatial GLMMs.

In the context of SAE, we propose a unified approach for Normal and non-Normal responses with spatial patterns. More precisely, we produce the EBP of small area predictors in Section 2, and in Section 3 we obtain the corresponding MSPE. In Section 4, we use Taylor series expansion and parametric bootstrap procedures to provide the second-order unbiased estimators of MSPE of EBP of small area predictors. We implement our method using a real dataset of esophageal cancer mortality in Minnesota, USA, from 1991 to 1998 in Section 5. The performance of the proposed approach is evaluated in Section 6 using two simulation studies with Poisson and binomial responses. Finally, Section 7 provides some concluding remarks and the Appendix contains the proof of Theorem 1. Supplementary Materials contain other techincal details, an extra simulation study in the case of Normal reponse, and R codes and corresponding files for the simualtions and application conducted in this paper.

2. STATISTICAL MODEL

The basic area-level model in small area estimation can be described as follows. Let y_i be the variable of interest for the ith small area $(i = 1, ..., m)$. The y_i are assumed to be independent conditional on latent variable η_i with the exponential family probability density or mass function

$$f(y_i|\eta_i, \phi) = \exp[\{y_i\eta_i - a(\eta_i)\}/\phi + b(y_i, \phi)], \tag{1}$$

where $a(\cdot)$ and $b(\cdot)$ are known functions, and ϕ is the known scale parameter. The latent variable η_i is modeled as

$$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u},$$

and $\eta_i = h[E(y_i|\mathbf{u})]$, where h is a strictly increasing function, $\mathbf{x}_i^\top$ is the $i-$th row of known matrix $\mathbf{X}(m \times p)$, $\boldsymbol{\alpha}_1 := \boldsymbol{\beta}(p \times 1)$ is a vector of unknown regression coefficient, $\mathbf{z}_i^\top$ is the $i-$th row of the identity matrix $\mathbf{Z}(m \times m)$, and $\mathbf{u} = (u_1, ..., u_m)^\top$ are spatial random effects from a multivariate Normal distribution $\mathbf{u}|\boldsymbol{\alpha}_2 \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_u(\boldsymbol{\alpha}_2))$. The objective in small area estimation is to make inferences on the small area predictors η_i or its variant.

To that end, we first need to obtain the conditional density of the latent variable η_i which can be written as

$$g(\eta_i|y_i, \boldsymbol{\alpha}) \propto \exp\left\{\frac{-\eta_i^2}{2\sigma_{\eta_i}^2} + \frac{\eta_i(\mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma_{\eta_i}^2} + [y_i\eta_i - a(\eta_i)]/\phi\right\}, \tag{2}$$

where $\sigma_{\eta_i}^2 = \mathbf{z}_i^\top \boldsymbol{\Sigma}_u \mathbf{z}_i$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top$. A Normal approximation, using the Laplace approximation (Rue, Martino, & Chopin, 2009) centred around the point $\eta_i^0 = \arg\max_{\eta_i} f(y_i|\eta_i, \phi)$, to the probability function (2) is constructed by linearizing the likelihood part of (2) at a fixed point η_i^0. The feasibility of this Normal approximation is evaluated through simulation studies in Section 6. We write the following for each small area $i(= 1, ..., m)$:

$$[y_i\eta_i - a(\eta_i)] \approx [y_i\eta_i^0 - a(\eta_i^0)] + (\eta_i - \eta_i^0)[y_i - a'(\eta_i^0)] - \frac{1}{2}(\eta_i - \eta_i^0)^2 a''(\eta_i^0), \tag{3}$$

where the first and second derivatives can be written in a closed form. Inserting (3) into (2), the conditional density of η_i has a Normal approximation with conditional mean $E(\eta_i|y_i, \boldsymbol{\alpha})$ and conditional variance $var(\eta_i|y_i, \boldsymbol{\alpha})$, given by

$$E(\eta_i|y_i, \boldsymbol{\alpha}) = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\Sigma}_u \mathbf{Z}^\top \mathbf{R}^{-1}[l(\mathbf{y}, \boldsymbol{\eta}^0) - \mathbf{X}\boldsymbol{\beta}] \tag{4}$$

and

$$var(\eta_i|y_i,\boldsymbol{\alpha}) = \mathbf{z}_i^\top[\boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_u\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\Sigma}_u]\mathbf{z}_i, \tag{5}$$

where $\mathbf{R} = \mathbf{Z}\boldsymbol{\Sigma}_u\mathbf{Z}^\top + \mathbf{P}, \mathbf{P}$ is a diagonal matrix with entries $\mathbf{P}_{i,i} = \phi/a''(\eta_i^0)$, $\boldsymbol{\eta}^0 = (\eta_1^0, ..., \eta_m^0)^\top$, and $l(y_i, \eta_i^0) = [y_i - a'(\eta_i^0) + \eta_i^0 a''(\eta_i^0)]/a''(\eta_i^0), (i = 1, ..., m)$.

When $\boldsymbol{\alpha}$ is known, the best predictor of η_i is given by $\tilde{\eta}_i^B(\boldsymbol{\alpha}, y_i) = \tilde{\eta}_i^B = E(\eta_i|y_i,\boldsymbol{\alpha})$ defined in (4). Moreover, the only sensible prediction variance for η_i is given by $E(\tilde{\eta}_i^B - \eta_i)^2 = var(\eta_i|y_i,\boldsymbol{\alpha}) =: g_{1i}(\boldsymbol{\alpha})$. By estimating the model parameters $\boldsymbol{\alpha}$, called $\hat{\boldsymbol{\alpha}}$, the EBP of η_i is given with replacing $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}$ in (4) as

$$\hat{\eta}_i^{EB} = \tilde{\eta}_i^B(\hat{\boldsymbol{\alpha}}, y_i)\{1 + O_p(m^{-1})\}, \tag{6}$$

noting that we estimate the model parameters $\boldsymbol{\alpha}$ using maximum likelihood estimation (MLE) approach via data cloning (DC) (Lele, Nadeem, & Schmuland, 2010). We breifly explain the DC approach below.

The DC method provides the MLE using the Bayesian computational approach. To understand the logic behind this approach, imagine a hypothetical situation where the observations $\mathbf{y} = (y_1, \ldots, y_m)^\top$ are repeated independently by K different individuals, and all these individuals happen to obtain exactly the same set of observations $\mathbf{y}$. These K repeated data sets are de-noted by $\mathbf{y}^{(K)} = (\mathbf{y}^\top, \mathbf{y}^\top, \ldots, \mathbf{y}^\top)^\top$. As a result, the likelihood function for the combination of the data from these K independent experiments is given by $\{L(\boldsymbol{\alpha}; \mathbf{y})\}^K = L^K(\boldsymbol{\alpha}; \mathbf{y})$ where $L(\boldsymbol{\alpha}; \mathbf{y})$ is the likelihood of the original observed $\mathbf{y}$. As shown by Lele, Nadeem, & Schmuland (2010), the likelihood function $L^K(\boldsymbol{\alpha}; \mathbf{y})$ has two important features: 1) the location of the maximum of this function is exactly equal to the location of the maximum of $L(\boldsymbol{\alpha}; \mathbf{y})$; 2) the Fisher information matrix based on this likelihood is K times the Fisher information matrix based on $L(\boldsymbol{\alpha}; \mathbf{y})$.

Assuming the model is identifiable and that there is a unique global mode for the likelihood function, we can write the posterior distribution of $\boldsymbol{\alpha}$ conditional on the data $\mathbf{y}^{(K)}$ as

$$\pi_K(\boldsymbol{\alpha}|\mathbf{y}^{(K)}) = \frac{L^K(\boldsymbol{\alpha}; \mathbf{y})\pi(\boldsymbol{\alpha})}{C(\mathbf{y}^{(K)})}, \tag{7}$$

where $\pi(\boldsymbol{\alpha})$ is the the prior distribution and $C(\mathbf{y}^{(K)}) = \int L^K(\boldsymbol{\alpha}; \mathbf{y})\pi(\boldsymbol{\alpha})d\boldsymbol{\alpha}$ is the normalizing constant. Following Lele, Nadeem, & Schmuland (2010), we can show that under some mild regularity conditions, as K becomes large, the posterior distribution of $\sqrt{K}\boldsymbol{\Sigma}^{-1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})|\mathbf{y}^{(K)}$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}$ which is the identity matrix with the dimension of $\boldsymbol{\alpha}$; noting that $\hat{\boldsymbol{\alpha}}$ is the MLE and $\boldsymbol{\Sigma}$ is the inverse of the Fisher information matrix for the MLE. Lele, Nadeem, & Schmuland (2010) also provided various checks to determine the adequate number of clones K based on

(7).

3. MEAN SQUARED PREDICTION ERROR APPROXIMATION

We now need to obtain the measure of variability of the $\hat{\eta}_i^{EB}$. To that end, we assume the following regularity conditions (referred to as RC later on) on the estimator $\hat{\boldsymbol{\alpha}}$ and the predictor $\tilde{\eta}_i^{B}(\boldsymbol{\alpha}, y_i)$ for large m :

1) The dimension of $\boldsymbol{\alpha}$ is bounded and the estimator $\hat{\boldsymbol{\alpha}}$ satisfies that $(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \mathbf{O}_p(m^{-1/2})$ and $E|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}| = \mathbf{O}(m^{-1/2})$.

2) We have $\eta_i = O_p(1)$ and $\tilde{\eta}_i^{B}(\boldsymbol{\alpha}, y_i) = O_p(1)$ for $i = 1, ..., m$. In addition, the estimator $\tilde{\eta}_i^{B}(\boldsymbol{\alpha}, y_i)$ is continuously differentiable with respect to $\boldsymbol{\alpha}$, and $\frac{\partial \tilde{\eta}_i^{B}(\boldsymbol{\alpha}, y_i)}{\partial \boldsymbol{\alpha}} = O_p(1)$.

Theorem 1. *Under the RC (1–2) and the model (1), a second-order approximation to the MSPE of the $\hat{\eta}_i^{EB}$ can be written as*

$$\text{MSPE}(\hat{\eta}_i^{EB}) = g_{1i}(\boldsymbol{\alpha}) + g_{2i}(\boldsymbol{\alpha}) + o(m^{-1}), \tag{8}$$

where $g_{1i}(\boldsymbol{\alpha})$ is given by (5) and $g_{2i}(\boldsymbol{\alpha}) = tr\Big\{E[(\frac{\partial \tilde{\eta}_i^{B}(\boldsymbol{\alpha}, y_i)}{\partial \boldsymbol{\alpha}})(\frac{\partial \tilde{\eta}_i^{B}(\boldsymbol{\alpha}, y_i)}{\partial \boldsymbol{\alpha}})^{\top}]E[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^{\top}]\Big\}$.

The proof is deferred to Appendix.

4. MEAN SQUARED PREDICTION ERROR ESTIMATION

4.1. Linearization Method

Since the approximated MSPE (8) is a function of unknown parameters $\boldsymbol{\alpha}$, it is not computable. We now obtain the estimation of MSPE$(\hat{\eta}_i^{EB})$ which is second-order unbiased in the sense that

$$E[\text{mspe}(\hat{\eta}_i^{EB})] = \text{MSPE}(\hat{\eta}_i^{EB}) + o(m^{-1}). \tag{9}$$

As shown in Theorem 1, the order of $g_{2i}(\boldsymbol{\alpha})$ is $O(m^{-1})$, so one can estimate $g_{2i}(\boldsymbol{\alpha})$ by $g_{2i}(\hat{\boldsymbol{\alpha}})$ unbiasedly up to second-order. To estimate $g_{1i}(\boldsymbol{\alpha})$, the naive estimator $g_{1i}(\hat{\boldsymbol{\alpha}})$ has a second-order bias due to $g_{1i}(\boldsymbol{\alpha}) = O(1)$. We can then use a Taylor expansion about $\boldsymbol{\alpha}$ for $g_{1i}(\boldsymbol{\alpha})$ as follows:

$$g_{1i}(\hat{\boldsymbol{\alpha}}) = g_{1i}(\boldsymbol{\alpha}) + (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^\top \frac{\partial g_{1i}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} + \frac{1}{2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^\top \frac{\partial^2 g_{1i}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) + o_p(m^{-1}),$$

yielding that

$$E[g_{1i}(\hat{\boldsymbol{\alpha}})] = g_{1i}(\boldsymbol{\alpha}) + g_{11i}(\boldsymbol{\alpha}) + g_{12i}(\boldsymbol{\alpha}) + o(m^{-1}),$$

where

$$g_{11i}(\boldsymbol{\alpha}, y_i) = [\frac{\partial g_{1i}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}]^\top E(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$$

and

$$g_{12i}(\boldsymbol{\alpha}) = \frac{1}{2} tr\Big\{ (\frac{\partial^2 g_{1i}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top}) E[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^\top] \Big\},$$

which are of the order $O(m^{-1/2})$ and $O(m^{-1})$ under the RC, respectively.

Theorem 2. *Under the RC (1–2) and the model (1), a second-order correct unbiased estimator of the MSPE of* $\hat{\eta}_i^{EB}$ *can be written as*

$$\text{mspe}(\hat{\eta}_i^{EB}) = g_{1i}(\hat{\boldsymbol{\alpha}}) - g_{11i}(\hat{\boldsymbol{\alpha}}) - g_{12i}(\hat{\boldsymbol{\alpha}}) + g_{2i}(\hat{\boldsymbol{\alpha}}). \tag{10}$$

It is well-known that the mspe based on the linearization method can take a negative value (Prasad & Rao, 1990); in this case, one can replace the negative value with a very small positive value. In Supplemnetray Materials (Section 1), we derive the EB prediction and corresponding second-order unbiased MSPE estimation of EBP of small area predictors for some popular distributions in exponential family (Normal, Poisson, and binomial) based on the above results.

4.2. Parametric Bootstrap Approach

We now obtain a nearly unbiased estimator of MSPE$(\hat{\eta}_i^{EB})$, in the sense of (9), using the parametric bootstrap approach. We first generate $\mathbf{u}^* = (u_1^*, ..., u_m^*)^\top$ from a multivariate Normal distribution with mean $\mathbf{0}$ and variance-covariance $\boldsymbol{\Sigma}_u(\hat{\boldsymbol{\alpha}}_2)$ to establish $\eta_i^* = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \mathbf{z}_i^\top \mathbf{u}^*, (i = 1, ..., m)$. A bootstrap sample is then generated from $y_i^*|(\eta_i^*, \hat{\boldsymbol{\alpha}}) \sim f(y_i^*|\eta_i^*, \hat{\boldsymbol{\alpha}}), i = 1, ..., m,$ noting that we construct the estimator $\hat{\boldsymbol{\alpha}}^*$ from the bootstrap sample $(y_1^*, ..., y_m^*)$ with the same method used to obtain the estimator $\hat{\boldsymbol{\alpha}}$. We then obtain the EB prediction of η_i^* using the bootstrap dataset $\{(y_i^*, x_i); i = 1, ..., m\}$ as $\hat{\eta}_i^{EB*} = \tilde{\eta}_i^B(\hat{\boldsymbol{\alpha}}^*, y_i^*)$ for $i = 1, ..., m$. Hence, the bootstrap MSPE estimator of $\hat{\eta}_i^{EB*}$ is given by

$$\text{mspe}_{boot1}(\hat{\eta}_i^{EB}) = E_*\{(\hat{\eta}_i^{EB*} - \eta_i^*)^2\} \equiv \hat{w}_i, \tag{11}$$

where E_* denotes the bootstrap expectation.

We also provide a double bootstrap (Hall & Maiti, 2006) by drawing a second-phase bootstrap sample from a given bootstrap sample using the bootstrap model parameters given above. Following as above we draw the second-phase bootstrap sample, from each first-phase bootstrap sample, to get second-phase bootstrap MSPE as $\text{MSPE}_{**}(\hat{\eta}_i^{EB**}) = E_{**}\{(\hat{\eta}_i^{EB**} - \eta_i^{**})^2\}$, where E_{**} denotes the second-phase bootstrap expectation. We then have the following bootstrap MSPE estimators proposed by Hall & Maiti (2006):

$$\text{mspe}_{boot2}(\hat{\eta}_i^{EB}) \approx \begin{cases} 2\hat{w}_i - \hat{v}_i & \hat{w}_i \geq \hat{v}_i \\ \hat{w}_i exp\{-(\hat{v}_i - \hat{w}_i)/\hat{v}_i\} & \hat{w}_i < \hat{v}_i \end{cases} \tag{12}$$

and

$$\mathrm{mspe}_{boot3}(\hat{\eta}_i^{EB}) \approx \hat{w}_i^2/\hat{v}_i, \tag{13}$$

where $\hat{v}_i = E_*[E_{**}\{(\hat{\eta}_i^{EB**} - \eta_i^{**})^2\}]$. In practice, we approximate $\hat{w}_i$ by drawing a large number, B_1, of independent bootstrap samples. Similarly, we approximate $\hat{v}_i$ by drawing a large number, B_2, of second-phase independent bootstrap samples from each first-phase bootstrap sample.

5. APPLICATION

We use a non-Normal response data to evaluate performance of the proposed approach. The data consists of the number of deaths due to esophagus cancer in the years from 1991 to 1998 at the 87 counties (small areas) in Minnesota, USA (Jin, Carlin, & Banerjee, 2005; Torabi, 2014). A spatial Poisson model is used as this disease is assumed to be rare enough relative to the population in each county. The model is then given by

$$y_i \sim Poisson(\lambda_i), \qquad i = 1, ..., 87, \tag{14}$$

$$log(\lambda_i) = \log(E_i) + \beta + \mathbf{z}_i^\top \mathbf{u},$$

where y_i is the observed number of death due to esophagus cancer in county i, E_i is the corresponding expected age-adjusted number of deaths, β is an intercept, $\mathbf{z}_i^\top$ is the ith raw of the identity matrix $\mathbf{Z}, \mathbf{u}$ are from proper CAR model (see Sectoin 1 of the Supplementry Materials for more details of this model) with parameters $\boldsymbol{\alpha}_2 = (\lambda_u, \sigma_u^2)$. Note that the expected number of deaths (E_i) is calculated by $E_i = \sum_{j=1}^{J} N_{ij} y_j / N_j$ where N_{ij} is the population at risk for the ith county and age group j, N_j is the population at risk for the age group j based on the USA Census 2000 dataset, and similarly y_j is the number of deaths for the age group j. The range of esophagus cancer cases during 7 years (1991–1998) is from 0 to 319 with mean and median of 15.99 and 8.00, respectively. The range of standardized mortality ratio (y_i/E_i) is from 0 to 2 with mean and median of 0.91 and 1.00, respectively.

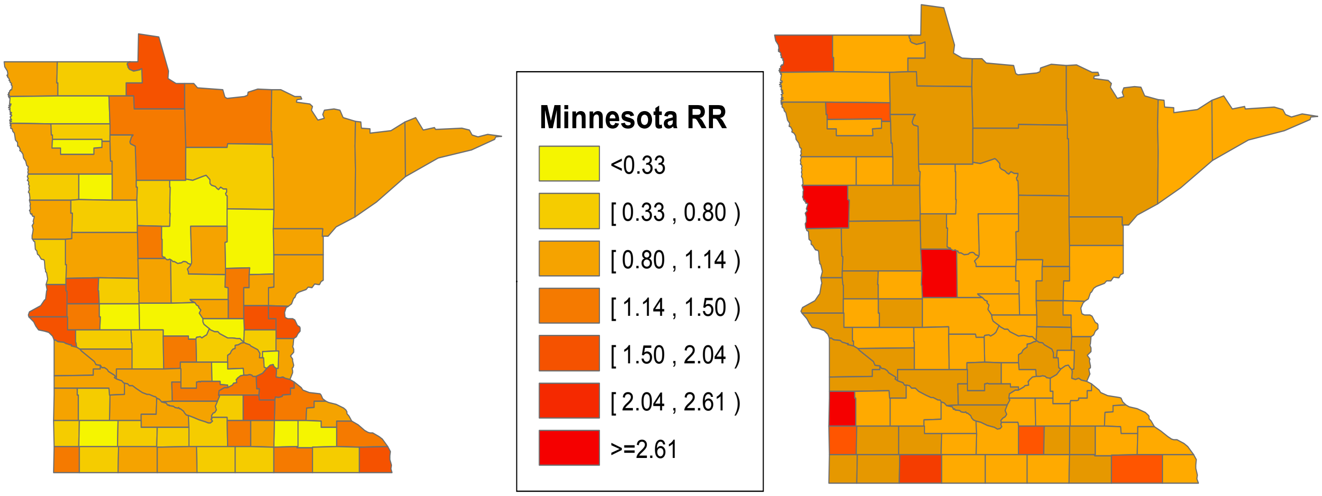
We first fit the model (14) to the dataset and provide the model parameters estimate and corresponding standard errors (Table 1). It appears that our data has spatial patterns as the spatial model parameters are statistically significant. We also fit the model with non-spatial random effects ($\lambda_u = 0$) to this dataset. As shown in Table 1, we now get a non-significant value for the dispersion parameter. To evaluate the impact of ignoring the spatial random effects, we further investigate it in the simulation study. We then provide the prediction of mortality ratio as well as raw ratio (y_i/E_i) of esophagus cancer in each county (Figure 1) with corresponding MSPE estimation of log-ratio of esophagus cancer (Figure 2) using the Taylor expansion and parametric bootstrap approaches; noting that in

TABLE 1: Model parameters estimates and corresponding standard errors using MLE approach for the spatial and non-spatial Poisson models.

Parameter	Spatial Model		Non-spatial Model	
	Estimate	Standard Error	Estimate	Standard Error
β	-0.041	0.054	-0.062	0.043
σ_u^2	0.012	0.005	0.032	0.019
λ_u	0.290	0.029	–	–

this paper we consider $B_1 = 1000$ and $B_2 = 100$ as the first- and second-phase of bootstrap samples, respectively.

As shown in Figure 1, our prediction ratios provide smooth estimates of raw (naive) ratios. The effect of model-based approach to produce smoothed estimates (Figure 1 b) in some counties in the central and northern parts of the Minnesota is very clear compared to the naive method (Figure 1 a). In particular, some of these counties have relatively small population sizes compared to the general population which clearly demonstrate the advantage of using spatial mixed model. As shown in Figure 2, the MSPE estimation of the EB prediction for the Taylor expansion approach (eq. (10)) is smaller than the other approaches, and first-phase bootstrap approach (eq. (11)) is relatively smaller than corresponding values of the second-phase bootstrap methods (eqs. (12) and (13)). We further investigate performance of these estimators in the simulation studies.



(a) (b)

FIGURE 1: Raw (a) and EB prediction (b) of mortality ratio of esophagus cancer in Minnesota, spatial Poisson model.

6. SIMULATION STUDY

In this section, we carry out simulation studies to evaluate performance of the proposed approach. Two simualtion studeis in the cases of Poisson and binomial responses are presented in this section, and third simualton study in the case of Normal response is deffered to Section 2 of the Supplementary Materials.

6.1. Spatial Poisson Model

We conduct a simulation study to evaluate performance of the proposed approach in the Poisson mixed model set-up. The spatial structure of the model is also

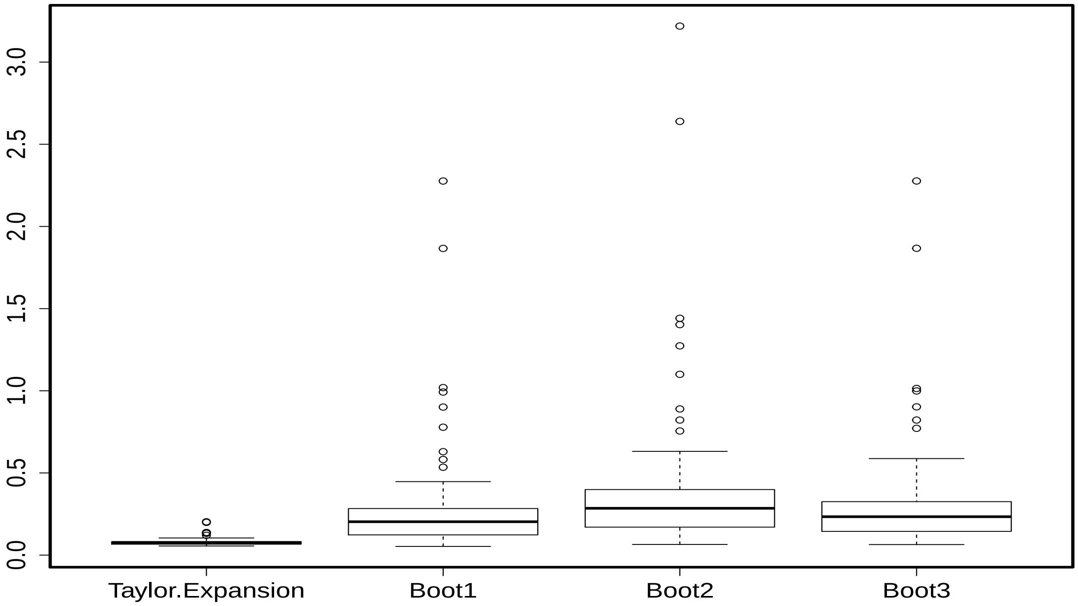


FIGURE 2: Boxplots of square root of mspe of log-ratio of esophagus cancer in Minnesota, spatial Poisson model.

based on the Minnesota county map (Section 5). We assume that the data are obtained from the following model:

$$y_i \sim Poisson(\lambda_i), \qquad i = 1, ..., 87,$$

$$\log(\lambda_i) = \log(n_i) + \beta + \mathbf{z}_i^\top \mathbf{u},$$

where $n_i = 30$ is an offset, β is an intercept, $\mathbf{z}_i^\top$ is the ith raw of the identity matrix $\mathbf{Z}, \mathbf{u}$ are generated from the proper CAR model with parameters $\boldsymbol{\alpha}_2 = (\lambda_u, \sigma_u^2)^\top$. We first generate $R = 1000$ independent samples $\mathbf{u}^{(r)}, (r = 1, ..., R)$, from the proper CAR model with parameters $\lambda_u = 0.6, \sigma_u^2 = 0.0001$, and then generate $y_i^{(r)} \sim Poisson(\lambda_i^{(r)}), (i = 1, \ldots, 87; r = 1, \ldots, R),$ where $log(\lambda_i^{(r)}) = \log(n_i) + \beta + \mathbf{z}_i^\top \mathbf{u}^{(r)}$ with $\beta = 0.001$. For each simulated run, we find the MLE of the model parameters to provide the prediction of the small area log-rates $\eta_i^{(r)} = \beta + \mathbf{z}_i^\top \mathbf{u}^{(r)}, (r = 1, \ldots, R)$, using $\hat{\eta}_i^{EB(r)} = \hat{\beta}^{(r)} + \mathbf{z}_i^\top E(\mathbf{u}^{(r)}|\mathbf{y})$ as in (6). We also calculate the empirical MSPE (EMSPE) of $\hat{\eta}_i^{EB}$ as

$$\text{EMSPE}(\hat{\eta}_i^{EB}) = \frac{1}{R}\sum_{r=1}^{R}[\hat{\eta}_i^{EB(r)} - \eta_i^{(r)}]^2,$$

and the relative bias (RB) of an estimator of the MSPE, say mspe, as

$$\text{RB}[\text{mspe}(\hat{\eta}_i^{EB})] = \left\{\frac{1}{R}\sum_{r=1}^{R}\text{mspe}^{(r)}(\hat{\eta}_i^{EB}) - \text{EMSPE}(\hat{\eta}_i^{EB})\right\}/\text{EMSPE}(\hat{\eta}_i^{EB}),$$

where $\hat{\eta}_i^{EB(r)}$ and $\eta_i^{(r)}$, and $\text{mspe}^{(r)}(\hat{\eta}_i^{EB})$ are the values of $\hat{\eta}_i^{EB}, \eta_i,$ and $\text{mspe}(\hat{\eta}_i^{EB})$ for the rth simulation batch, respectively. Note that $\text{mspe}(\hat{\eta}_i^{EB})$ is calculated for the both Taylor expansion and bootstrap approaches.

We also evaluate the performance of the non-spatial model $(\lambda_u = 0)$ in this simulation study. The result of EMSPE of $\hat{\eta}_i^{EB}$ is reported in Figure 3. As shown in Figure 3, the values of EMSPE are smaller for the proposed approach compared to the non-spatial model. The results of absolute RB (ARB) of mspe of $\hat{\eta}_i^{EB}$ for the Taylor expansion and bootstrap approaches are also reported in Figure 4. The proposed approach using Taylor expansion performs very well in terms of ARB $(< \%8)$; noting that the first-phase bootstrap also performs better than the

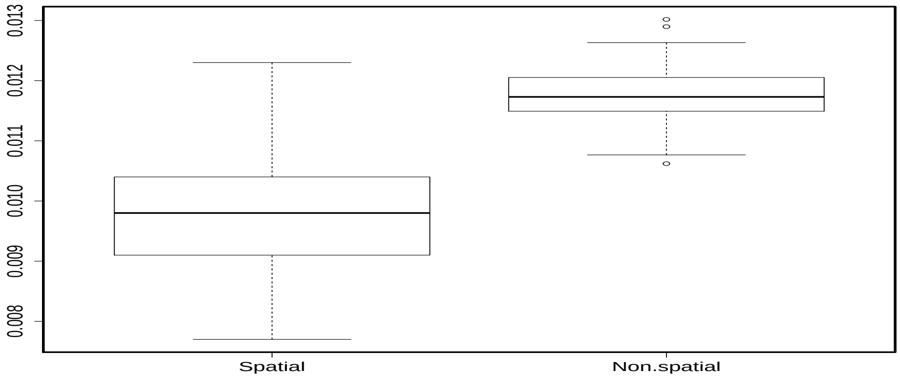


FIGURE 3: Boxplots of the EMSPE of $\hat{\eta}_i^{EB}$, spatial and non-spatial Poisson models.

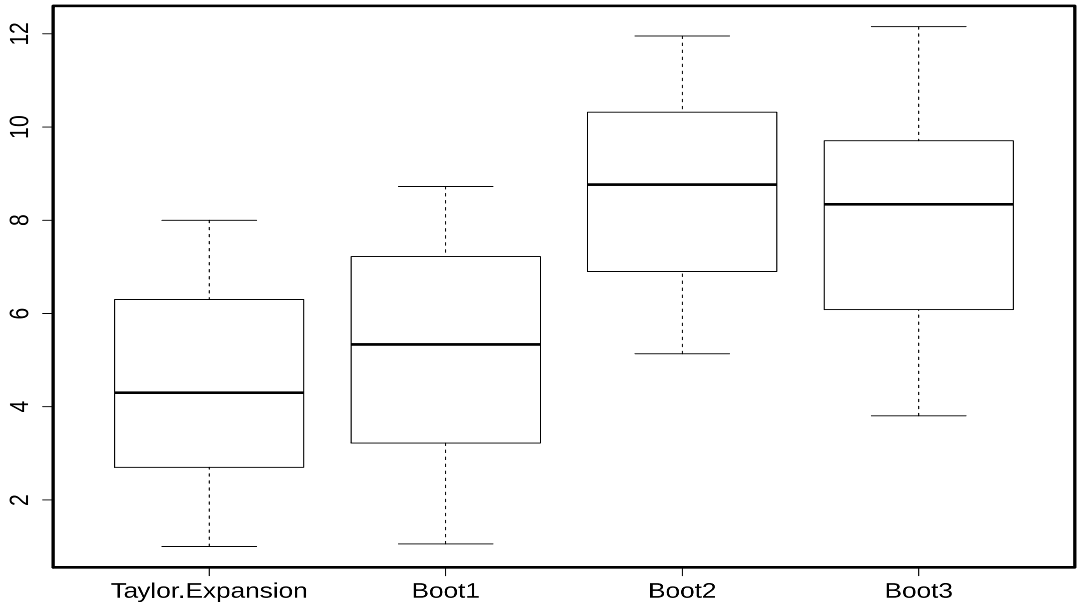


FIGURE 4: Boxplots of percent absolute RB of mspe of $\hat{\eta}_i^{EB}$, spatial Poisson model.

both second-phase bootstrap methods in terms of RB of mspe of $\hat{\eta}_i^{EB}$.

6.2. Spatial Binomial Model

We also conduct a simulation study to evaluate performance of the proposed approach in the binomial mixed model set-up. The spatial structure of the model is also based on the Minnesota county map (Section 5). We assume that the data are obtained from the following model:

$$y_i \sim Binomial(n_i, p_i), \qquad i = 1, ..., 87,$$

$$\log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 x_i + \mathbf{z}_i^\top \mathbf{u},$$

where $n_i = 1, \beta_0 = \beta_1 = 0.1,$ and x_i is generated from uniform distribution between 0 and 1 and fixed during the simulation study. We first generate $R = 1000$ independent samples $\mathbf{u}^{(r)}, (r = 1, ..., R),$ from the proper CAR model with parameters $\lambda_u = 0.5, \sigma_u^2 = 0.0001,$ and then generate $y_i^{(r)} \sim Binomial(n_i, p_i^{(r)}), (i = 1, \dots, 87; r = 1, \dots, R),$ where $p_i^{(r)} = \frac{e^{\beta_0 + \beta_1 x_i + z_i^\top \mathbf{u}^{(r)}}}{1 + e^{\beta_0 + \beta_1 x_i + z_i^\top \mathbf{u}^{(r)}}}$. For each simulated run, we find the MLE of the model parameters to provide the prediction of the small area logit $\eta_i^{(r)} = \beta_0 + \beta_1 x_i + \mathbf{z}_i^\top \mathbf{u}^{(r)}, (r = 1, \dots, R),$ using $\hat{\eta}_i^{EB(r)} = \hat{\beta}_0^{(r)} + \hat{\beta}_1^{(r)} x_i + \mathbf{z}_i^\top E(\mathbf{u}^{(r)}|\mathbf{y})$ as in (6). We also calculate the EMSPE($\hat{\eta}_i^{EB}$) and the RB[mspe($\hat{\eta}_i^{EB}$)] similar to the spatial Poisson model (Section 6.1).

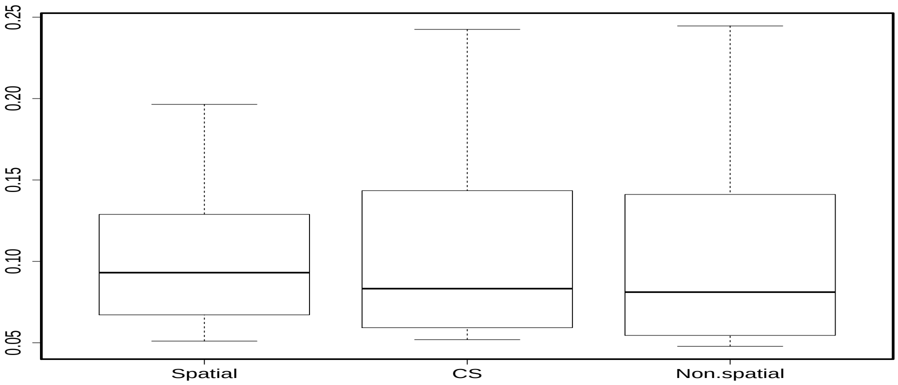


FIGURE 5: Boxplots of the EMSPE of $\hat{\eta}_i^{EB}$ in the cases of spatial, non-spatial, and CS Binomial models.

We also evaluate the performance of the non-spatial model $(\lambda_u = 0)$ and the SAR spatial model, developed for the binomial model by Chandra & Salvati (2018), called CS model, in this simulation study. The results of EMSPE of $\hat{\eta}_i^{EB}$ are reported in Figure 5. As shown in Figure 5, the range of EMSPE values is smaller for the proposed approach compared to the non-spatial and CS models while the performance of the CS and non-spatial models is very similar. In particular, the EMSPE values of non-spatial and CS models are $\%23$ larger than the corresponding values of the spatial model in some areas, noting that the CS model is based on the spatial SAR unlike the proposed model which is based on the spatial CAR. The results of ARB of mspe of $\hat{\eta}_i^{EB}$ for the Taylor expansion and bootstrap approaches are also reported in Figure 6. The proposed approach using Taylor expansion performs very well in terms of ARB $(< \%8)$; noting that the first-phase bootstrap also performs better than the both second-phase bootstrap methods in terms of RB of mspe of $\hat{\eta}_i^{EB}$.

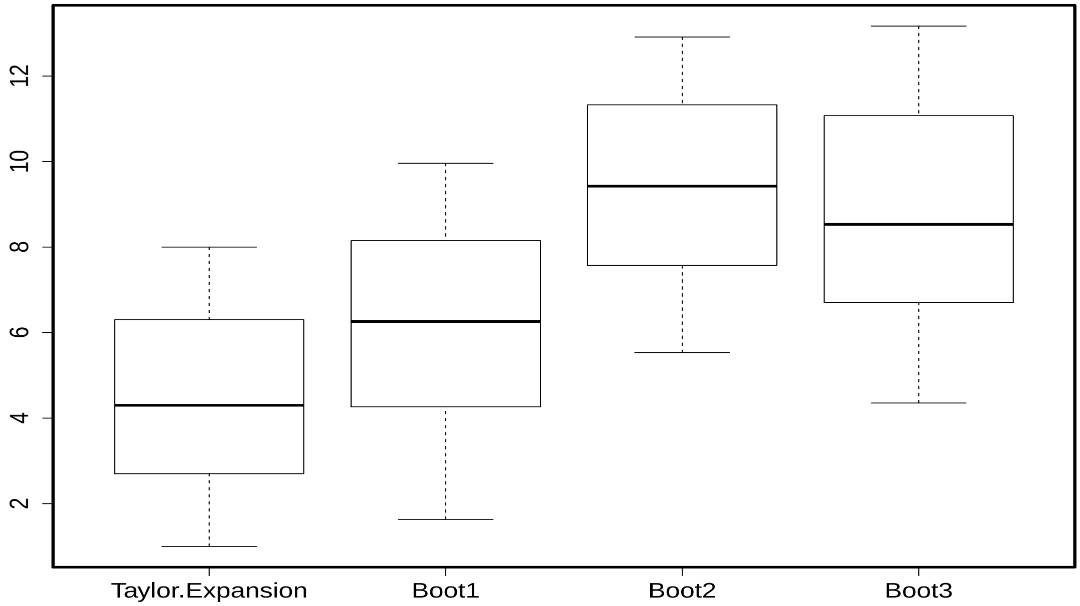


FIGURE 6: Boxplots of percent absolute RB of mspe of $\hat{\eta}_i^{EB}$, spatial Binomial model.

7. DISCUSSION

Generalized linear mixed models (GLMMs), assuming small areas are independent from each other, have been previously used to derive empirical best prediction (EBP) of small area predictors as well as the corresponding second-order unbiased estimator of mean squared prediction errors (MSPE) of the EBP of small area predictors. However, there are many applications that the small area predictors are related to their locations. For instance, accurate statistical information concerning the wellbeing of people at regional level is needed to target the policies or programs aimed at reducing poverty in poorer regions; the estimation of poverty at regional or local level is then a really important task for policy making (Marhuenda, Molina, & Morales, 2013). As another application, among many others, is when health agencies (e.g., policy making) need to know the spatial patterns of a rare disease (e.g., chronic disease or cancer) to identify small areas with high risk of disease to implement the prevention.

We have proposed a unified approach for Normal and non-Normal responses with spatial patterns in the context of small area estimation. In particular, we have provided EBP of small area predictors and derived second order approximation to the MSPE of EBP of small area predictors. We have also obtained second-order unbiased MSPE estimation of EBP of small area predictors by Taylor expansion as well as parametric bootstrap approaches. We have shown by simulation studies (and a real data application of esophagus cancer dataset in Minnesota) that the proposed approach works very well in terms of small area predictors and their precisions.

APPENDIX

Proof of Theorem 1

We can write

$$\text{MSPE}(\hat{\eta}_i^{EB}) = E\{(\hat{\eta}_i^{EB} - \eta_i)^2\}$$

$$= E\{(\tilde{\eta}_i^{B} - \eta_i)^2\} + E\{(\hat{\eta}_i^{EB} - \tilde{\eta}_i^{B})^2\}$$

$$= g_{1i}(\boldsymbol{\alpha}) + E\{(\hat{\eta}_i^{EB} - \tilde{\eta}_i^B)^2\}\ ,$$

noting that $E[(\tilde{\eta}_i^B - \eta_i)(\hat{\eta}_i^{EB} - \tilde{\eta}_i^B)] = E[(\hat{\eta}_i^{EB} - \tilde{\eta}_i^B)E(\tilde{\eta}_i^B - \eta_i|y_i)] = 0$. It is also noted that

$$\hat{\eta}_i^{EB} = \tilde{\eta}_i^B(\boldsymbol{\alpha}, y_i) + [\frac{\partial \tilde{\eta}_i^B(\boldsymbol{\alpha}^\star, y_i)}{\partial \boldsymbol{\alpha}}]^\top(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}),$$

where $\boldsymbol{\alpha}^\star$ is between $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$. Thus, we obtain

$$E\{(\hat{\eta}_i^{EB} - \tilde{\eta}_i^B)^2\} = E[\{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^\top \frac{\partial \tilde{\eta}_i^B(\boldsymbol{\alpha}, y_i)}{\partial \boldsymbol{\alpha}}\}^2] + o(m^{-1})$$

$$= tr\Big\{E[(\frac{\partial \tilde{\eta}_i^B(\boldsymbol{\alpha}, y_i)}{\partial \boldsymbol{\alpha}})(\frac{\partial \tilde{\eta}_i^B(\boldsymbol{\alpha}, y_i)}{\partial \boldsymbol{\alpha}})^\top(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^\top]\Big\} + o(m^{-1})$$

$$= tr\Big\{E[(\frac{\partial \tilde{\eta}_i^B(\boldsymbol{\alpha}, y_i)}{\partial \boldsymbol{\alpha}})(\frac{\partial \tilde{\eta}_i^B(\boldsymbol{\alpha}, y_i)}{\partial \boldsymbol{\alpha}})^\top]E[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^\top]\Big\} + o(m^{-1}),$$

which completes the proof of Theorem 1.

ACKNOWLEDGEMENTS

Constructive comments and suggestions of the Editor, an Associate Editor, and a referee, which led to an improved version of the manuscript, are greatly appreciated. This work was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

SUPPLEMENTARY MATERIALS

The Supplementary Materials contain three sections. Section 1 provides the detail of derivations of the EB prediction and corresponding second-order unbiased MSPE estimation of EBP of small area predictors for three popular responses in exponential family (Normal, Poisson, and binomial). Section 2 provides a simulation study in the case of spatial Normal model. Section 3 provides R codes and corresponding "readme" files for the simulations and application conducted in this paper.

BIBLIOGRAPHY

[1] Chandra, H. & Salvati, N. (2018). Small area estimation of proportions under a spatial dependent aggregated level random effects model. *Communications in Statistics – Theory and Methods*, 47, 1234–1255.

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

[3] Hall, P. & Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *Journal of Royal Statistical Society, Series B*, 68, 221–238.

[4] Jiang, J. (2017). *Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems*. Chapman & Hall/CRC, Boca Raton.

[5] Jin, X., Carlin, B., & Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics*, 6, 950–961.

[6] Lele, S.R., Nadeem, K., & Schmuland, B. (2010). Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105, 1617–1625.

[7] Marhuenda, Y., Molina, I., & Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics and Data Analysis*, 58, 308–325.

[8] McCulloch, C.E. & Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*. John Wiley and Sons, New York.

[9] Molina, I., Salvati, N., & Pratesi, M. (2009). Bootstrap for estimating the MSE of the Spatial EBLUP. *Computational Statistics*, 24, 441–458.

[10] Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28, 40–68.

[11] Prasad, N.G.N. & Rao, J.N.K. (1990). The estimation of mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163–171.

[12] Rao, J.N.K. & Molina, I. (2015). *Small Area Estimation. 2nd ed.*, John Wiley, New York.

[13] Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of Royal Statistical Society, Series B*, 71, 319–392.

[14] Schabenberger, O. & Gotway, C.A. (2004). *Statistical Methods for Spatial Data Analysis*. Chapman and Hall/CRC, Boca Raton.

[15] Searle, S.R., Casella, G., & McCulloch, C.E. (2006). *Variance Components*. John Wiley and Sons, New York.

[16] Singh, B.B., Shukla, G.K., & Kundu, D. (2005). Spatio-temporal models in small area estimation. *Survey Methodology*, 31, 183–195.

[17] Torabi, M. (2014). Spatial generalized linear mixed models with multivariate CAR models for areal data. *Spatial Statistics*, 10, 12–26.

[18] Torabi, M. & Jiang, J. (2019). Estimation of mean squared prediction error of empirically spatial predictor of small area means under a linear mixed model, *Technical Report*.

Received 14 December 2017

Accepted