Bayesian Empirical Likelihood Spatial Model applying Leroux Structure

Farzana Jahan [1] and Kerrie Mengersen [1]

[1] School of Mathematical Science, ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology, Australia.

Abstract

Bayesian empirical likelihood (BEL) methods have been applied to spatial data analysis for small area estimation by extending the Fay-Herriot model in recent times. But the area of Bayesian semiparametric empirical likelihood spatial models has not been explored by utilising some of the popular choices of spatial dependence structures. The present study is an attempt to develop a Bayesian semiparamtric model for spatial data analysis utilising the popular Leroux prior for spatial dependence. The proposed model will be evaluated using simulated data on cancer incidences. The model performance will be compared with existing parametric and semiparametric models.

1 Introduction

Bayesian Empirical Likelihood (BEL) was first described by Lazar (2003) which was a Bayesian perspective of Empirical Likelihood (EL) proposed by (Owen, 1988). The concept of EL has been utilised in Bayesian analysis in a few instances since then (Schennach, 2005; Mengersen, Pudlo, & Robert, 2013; Chib, Shin, & Simoni, 2017). BEL provides a flexible semi-parametric approach using data to approximate the likelihood part of Bayesian posterior combined with parametric prior distributions. In recent years, the BEL approach has been applied in spatial models (Chaudhuri & Ghosh, 2011; Porter, Holan, & Wikle, 2015a, 2015b). EL framework, motivated from independent data has been applied for spatially dependent data by utilising hierarchical framework, moving the spatial dependence structure in the parameter model of hierarchy which leaves no obstacles in applying EL structure in observation level (Porter et al., 2015a).

Chaudhuri and Ghosh (2011) introduced area-level and unit-level models using BEL which can handle both discrete and continuous data in the context of small area estimation. The article extended the traditional Fay-Herriot (FH) model using an EL framework with informative priors for the spatial random effects as hierarchical multivariate Gaussian and Dirichlet process mixture priors. Following this work, Porter et al. (2015a) provided a more complete version of the EL formulation in the context of Bayesian hierarchical models which was demonstrated using an example with spatial dependencies and which can reportedly be generalized to temporal or spatial-temporal dependencies. The extension to (Chaudhuri & Ghosh, 2011) is made by formulating a different set of lattice priors utilising the generalized Moran basis (Hughes & Haran, 2013) for the spatial dependencies (Porter et al., 2015a). The multivariate extension of the Bayesian semiparametric hierarchical empirical likelihood (BSHEL) model of (Porter et al., 2015a) can be found in (Porter et al., 2015b). However, there is still scope for work on BEL formulation of spatial models using other available choices of priors for spatial dependence.

A model which accounts for spatial dependence that has gained popularity in the last decade is the Leroux model (Leroux, Lei, & Breslow, 2000). Details of the estimation procedure of the Leroux model using maximum likelihood and a penalized quasi likelihood approach are presented by Leroux (2000). Leroux structured Conditional Autoregressive (CAR) priors have been applied to disease mapping in many instances (Ugarte, Adin, Goicoa, & Militino, 2014). A comparison of the performance of this model with other available spatial models in Bayesian analysis of spatial disease mapping can be found in Lee (2011). Lee (2011) compared the Leroux model with the intrinsic and convolution model, popularly known as the BYM model (Besag, York, & Mollie, 1993) and the Cressie model which applies a proper CARmodel (Ver Hoef & Cressie, 1993; Stern & Cressie, 2000) based on a simulation study and real datasets on cancer from Scotland. Lee (2011) found that the Leroux model performed best in terms of consistent results for different spatial correlation scenarios. Theoretical reasons in favor of this model were also obtained in Lee (2011). More recently, Rampaso, de Souza, and Flores (2016), included the Lu model, which considers a random adjacency matrix by modelling the spatial weights (Lu, Reilly, Banerjee, & Carlin, 2007) in addition to the models compared by (Lee, 2011) using disease mortality data in Brazil. The authors found that all the models yielded close results. Pros and cons of each model choice were presented for different scenarios based on a simulation study and the choice of most relevant model for a particular application was left up to researcher's judgment.

The aim of the present study is to extend the BSHEL proposed by Porter et al. (2015a) by utilising the Leroux CAR prior to model spatial random effects. A recap of the BEL models and BEL spatial models followed by formulation of the proposed model and a detailed algorithm to draw the posterior samples are discussed in Section 2. The application of the proposed model is discussed in section 3. The results of the application are presented in Section 4 followed by Conclusions and future works in section 5.

2 BEL Spatial Leroux Model

This section briefly discusses the background of Bayesian Empirical Likelihood(BEL), then describes BEL for spatial dependence and formulates a BEL spatial Leroux model.

2.1 Recap of Bayesian Empirical Likelihood

Empirical likelihood (EL) is a non-parametric method of statistical inference, which combines the reliability of non-parametric methods with the flexibility and effectiveness of the likelihood approach. Owen (1988) showed that the empirical likelihood ratio function can be used for constructing confidence intervals for a sample mean, for a class of M-estimates and for other differentiable statistical functionals. These methods were framed as a non-parametric extension of Wilks (1938) theorem for parametric likelihood ratio tests. EL can be constructed without any assumptions about the parametric family of distribution of data and can be used for hypothesis testing as well as construction of confidence region without an information/variance estimator.

In a Bayesian framework, the likelihood is used to update a prior distribution and yield posterior inference. Lazar (2003) argued that EL can be used in place of a density and, when multiplied by prior can yield the posterior distribution of interest in such an analysis. The author explored the characteristics of Bayesian inference using EL instead of a parametric density.

Reprising Lazar (2003), for data points $y_1, y_2, ..., y_n$ from some unknown distribution F, let some functional of F, say $\boldsymbol{\theta(F)}$ be the parameter of interest for inference, which can be determined by the estimating equation $f(y_i, \boldsymbol{\theta})$. Then the empirical likelihood ratio function can be defined as:

$$W_E(\boldsymbol{\theta}) = 2\sum_{i=1}^{n} log\{1 + \lambda^T f(y_i, \boldsymbol{\theta})\} \tag{1}$$

where the vector λ satisfies

$$\sum_{i=1}^{n} \frac{f(y_i, \boldsymbol{\theta})}{1 + \lambda^T f(y_i, \boldsymbol{\theta})} = 0.$$

To proceed with the Bayesian likelihood, the profile EL function $W_E(\boldsymbol{\theta})$ as shown in equation (1) is chosen as the likelihood part of the Bayesian theorem, and a prior on the functional $\boldsymbol{\theta}$ is imposed. Lazar (2003) reported that the Bayesian approach increases the flexibility of the EL approach as it allows allocation of the knowledge about the problem between the data component (EL, without any assumptions about the data generating distribution) and the prior beliefs resulting in shorter length of posterior intervals for the functional of interest. The paper tested the methods using simulated data sets.

Schennach (2005) argued that there was lack of probabilistic interpretation of EL based Bayesian models, rather the validity of using EL in Bayesian inference came from Monte Carlo simulations (?, ?). Therefore, the author explored the Bayesian analysis of moment condition models through EL. Starting from a well-known parametric case of a Bayesian posterior of a parameter vector $\boldsymbol{\theta}$ in the presence of a vector of nuisance parameters, $\boldsymbol{v}$, the author showed the derivation of non-parametric version by allowing the dimension of $\boldsymbol{v}$ to go to infinity. The author showed that the non-parametric posterior is equivalent to a representation as an EL type function where the weights attributed to the sample points, which can be calculated by solving an entropy maximisation problem. That is,

$$p(\boldsymbol{\theta}|X) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} w_i^*(\boldsymbol{\theta})$$

where $p(\boldsymbol{\theta})$ is a given prior on $\boldsymbol{\theta}$ and $(w_1^*(\boldsymbol{\theta}), ..., w_n^*(\boldsymbol{\theta}))$ is the solution of

$$max \sum_{i=1}^{n} -w_i \text{log} w_i, \text{subject to} \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i g_i(x_i, \boldsymbol{\theta}) = 0.$$

This method is called Bayesian exponentially tilted empirical likelihood (BETEL) as the weights are computed as in Efron's exponential tilting (Efron, 1981). The author showed an application using real life data.

Mengersen et al. (2013) used EL in approximate Bayesian computation and proposed a new algorithm, Bayesian computation via empirical likelihood (BCel), by replacing the data likelihood by EL.

Chib et al. (2017) developed fully Bayesian framework for comparison and estimation in statistical models which are defined by moment restrictions only. The authors showed that using BETEL, it is possible to develop a fully Bayesian treatments of models with moment restrictions even if they are misspecified. The paper described two key concepts. The first part of the paper re-expressed the moment conditions $E^P[g(X, \boldsymbol{\theta})] = 0$ using a nuisance parameter as: $E^P[g^A(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{V})] = 0$, where V is an additional nuisance parameter, E^P is the expectation taken with respect to P and $g^{\boldsymbol{A}}(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{V}) = g(\boldsymbol{X}, \boldsymbol{\theta}) - \boldsymbol{V}$. The augmented moment condition model is misspecified if the set of probability measures implied by the above moment restrictions do not contain the true data generating process P. The authors made the prior-posterior analysis of BETEL from Schennach (2005) adapting to the presence of augmented parameter and misspecification of model which led to new results. The posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{V})$,which is called the BETEL posterior distribution was expressed in the form

$$\pi(\boldsymbol{\theta}, \boldsymbol{v}|\boldsymbol{x}_{1:n}) \propto \pi(\boldsymbol{\theta}, \boldsymbol{v}) \prod_{i=1}^{n} \frac{e^{\hat{\rho}(\boldsymbol{\theta},\boldsymbol{v})' g^{\boldsymbol{A}}(\boldsymbol{x}_i, \boldsymbol{\theta}, \boldsymbol{v})}}{\sum_{j=1}^{n} e^{\hat{\rho}(\boldsymbol{\theta},\boldsymbol{v})' g^{\boldsymbol{A}(\boldsymbol{x}_j, \boldsymbol{\theta}, \boldsymbol{v})}}}$$

which can be efficiently simulated by Markov Chain Monte Carlo (MCMC). The detailed algorithm is provided in the paper (Chib et al., 2017). Using an example of simple linear regression models, the authors illustrated the methods, evaluating the asymptotic properties of the BETEL function under the scenarios of correct specification and misspecification of the model.

The second concept developed by Chib et al. (2017) was a model selection criteria based on marginal likelihoods and the Bayes factor. The marginal likelihood was computed using the output of MCMC simulation from the BETEL posterior distribution. It was observed that the marginal likelihood criteria selects the model that is closer to the true data generating process in terms of Kullback Leiber divergence. The authors explained all their proposed methods with several examples.

2.2 Recap of Bayesian Empirical Likelihood for Spatial Model

BEL has recently been employed for spatial analysis (Chaudhuri & Ghosh, 2011; Porter et al., 2015a, 2015b) and has been found to provide precise estimation of small area effects.

Chaudhuri and Ghosh (2011) introduced area-level and unit-level models using BEL by extending the traditional Fay-Herriot (FH) (Fay III & Herriot, 1979) model using an EL framework. Following this work, Porter et al. (2015a) provided a general hierarchical Bayesian framework incorporating empirical likelihood methodology for the data model and developed a spatial FH model used for small area estimation (SAE). The FH model for SAE can be written as:

$$Y_i = \mu_i + \epsilon_i$$

$$\mu_i = \boldsymbol{X_i'}\boldsymbol{\beta} + \psi_i$$

where Y_i is a design unbiased estimate of μ_i, and ϵ_i is an unstructured error component, $\boldsymbol{X_i'}$ is the vector of covariate information for area i, $\boldsymbol{\beta}$ is the vector of fixed covariate effects, $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$ and ψ_i is spatially referenced random effect for area i.

The priors for the vector of spatial random effects $\boldsymbol{\psi}$ can be specified in three ways following Porter et al. (2015a) and Chaudhuri and Ghosh (2011).

The first two priors are specified by Chaudhuri and Ghosh (2011):
(i) independent and identically distributed (iid) Gaussian distribution with variance $\boldsymbol{A}$ following a Inverse Gamma (IG) distribution

$$\psi_i \sim N(0, A), A \sim IG(\alpha_1, \alpha_2)$$

(ii) Dirichlet process (DP) with a Gaussian base

$$\psi_i | G \sim G, G|A \sim DP(\alpha, \mathcal{G})$$

where $DP(\alpha, \mathcal{G})$ represents a Dirichlet process with precision parameter α and a base measure $G_0 \sim N(0, A)$.
(iii) Porter et al. (2015a) used the generalised Moran basis (Hughes & Haran, 2013) to specify the priors for ψ_i:

$$\pi(\boldsymbol{\psi}) \propto \tau^{q/2}\exp\{-\frac{1}{2}\tau\boldsymbol{\psi^* M'(B_+ - B)M\psi^*}\}$$

where, $\boldsymbol{B}$ is an adjacency matrix for a first order Intrinsic Gaussian Random Field (IGMRF) (rank $\boldsymbol{(B)} = n-1$), $\boldsymbol{B_+}$ is a diagonal matrix with $\{\boldsymbol{B_+}\}_{i,i} = \sum_{j \in ne(i)} b_{ij}, b_{ij} = 1$ if i and j are adjacent and 0 otherwise, where $j \in ne(i)$ means that area j is a neighbour of area i . $\boldsymbol{M}$ is the set of eigen vectors corresponding to the non-zero eigenvalues of the matrix $\boldsymbol{P_c B P_c}$ with $\boldsymbol{P_c = I - X(X'X)^{-1}X'}$ q is the number of non-zero eigenvalues of the matrix $\boldsymbol{P_c B P_c}$ and $\boldsymbol{\psi = M\psi^*}$. τ is the precision parameter whose prior can be specified as Inverse Gamma:

$$\tau \sim IG(\alpha_1, \alpha_2)$$

where α_1 and α_2 are chosen to be 1 in Porter et al. (2015a). The prior for fixed covariate effect is specified

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta^*}, g^{-1}A\boldsymbol{I_2})$$

where g represents the Zellner prior (Zellner, 1988), which is evaluated at a fixed point estimate 10 (Chaudhuri & Ghosh, 2011). In the prior specification of Porter et al. (2015a), A is replaced by τ^{-1}.

For estimation of the fixed effects and the random effects, BEL was used in spite of using a parametric distribution for Y_i. The estimating equations for $(\boldsymbol{\beta, \psi})$ (Porter et al., 2015a) are:

$$\sum_{i=1}^{n} w_i(y_i - \mu_i) = 0 \tag{2}$$

$$\sum_{i=1}^{n}\{w_i(y_i-\mu_i)^2/\sigma_i^2\}-1=0. \tag{3}$$

Details of the MCMC sampling algorithm using a random walk Metropolis-Hastings (MH) approach can be found in Porter et al. (2015a).

2.3 Bayesian Semiparametric Leroux Empirical Likelihood Spatial Model

The approach of the present study is to extend the BEL semiparametric model proposed by Chaudhuri and Ghosh (2011) and Porter et al. (2015a) by applying the structure of spatial random effects ψ_i proposed in (Leroux, 2000). The proposed model in the study is named Bayesian Semiparametric Leroux Empirical Likelihood (BSLEL) Spatial Model.

According to the Leroux model specification, the spatial random effect $\boldsymbol{\psi}$, under intrinsic autoregression, has the following distribution (Leroux, 2000)

$$\boldsymbol{\psi} \sim MVN(\mathbf{0},\sigma^2\boldsymbol{D}) \tag{4}$$

where σ^2 is the overall variance , D is a singular covariance matrix. Leroux (2000) proposed the generalized inverse of D, D^- as:

$$\boldsymbol{D}^- = (1-\rho)I+\rho\boldsymbol{R} \tag{5}$$

where, I is an identity matrix, $\boldsymbol{R}$ is the intrinsic autoregression matrix which represents the neighbourhood structure of the regions with typical element R_{ij} as,

$$R_{ij} = \begin{cases} n_i, & i=j \\ -\boldsymbol{I}(i\sim j), & i\neq j \end{cases}$$

where n_i is the numbers of neighbours of region i and $\boldsymbol{I}(i \sim j)$ is an indicator function taking the value 1 when i and j are adjacent.

The term ρ is introduced as a spatial dependence parameter, $\rho \in [0,1]$, whose two extreme cases give rise to the independence model ($\rho = 0$) and intrinsic autoregression ($\rho = 1$). The Leroux Model proposed by Leroux et al. (2000) is more flexible than the earlier models, namely, the BYM model (Besag et al., 1993) and the Cressie model (Ver Hoef & Cressie, 1993). The Leroux model has overcome the shortcoming of the Cressie model which depended on the number of neighbours even when there was no spatial dependence by inclusion of the spatial dependence parameter ρ and setting $\rho = 0$ makes the variance term independent of the neighbourhood structure (σ^2). The Leroux model also captures the intrinsic autoregression case when $\rho = 1$. The flexibility of the Leroux model has been one of the reasons of popularity of this model in spatial data analysis. This has also motivated the formulation of BEL model using the Leroux prior structure in the present study.

The conditional moments corresponding to (1) can be expressed as weighted averages of local moments (Leroux et al., 2000):

$$E(\psi_i|\psi_{-i}) = \frac{1-\rho}{1-\rho+\rho n_i}\times 0 + \frac{\rho n_i}{1-\rho+\rho n_i}\frac{1}{n_i}\sum_{j\sim i}\psi_j \tag{6}$$

$$var(\psi_i|\psi_{-i}) = \frac{1-\rho}{1-\rho+\rho n_i}\times \sigma^2 + \frac{\rho n_i}{1-\rho+\rho n_i}\frac{\sigma^2}{n_i} \tag{7}$$

where ψ_{-i} denotes the random effect vector with the i-th component deleted.

The BSLEL model utilizes the parametric prior of the Leroux model for the spatial random effects and a Gaussian prior for the covariate fixed effect and EL to estimate the parameters of the SAE model without specifying the data distribution of Y_i. Thus this model formulation is different from those of Porter et al. (2015a) and Chaudhuri and Ghosh (2011) in terms of prior specification for the spatial random effects which results in a completely new algorithm to obtain the posterior distributions of interest.

The empirical likelihood (EL) estimating equations utilised for estimating (β, ψ) are in equations (2) and (3) with $\mu_i = x_i^{'}\beta + \psi_i$. The constrained optimization of EL estimating equations is well established (Owen, 2001). Some R-packages perform the computation as well (gmm, emplik etc.). A random walk Metropolis-Hastings sampling algorithm is proposed to fit the proposed model following the suggestions of Porter et al. (2015a).

2.3.1 Prior Distributions of BSLEL

The prior distribution of random effect $\boldsymbol{\psi}$ is taken to be a Leroux structure prior given by equation (3). Let us define a precision parameter associated with the variance of the random effect as $\tau = \frac{1}{\sigma^2}$. Then the distribution of random effect ψ can be written as,

$$\boldsymbol{\psi} \sim MVN(\vec{0}, \tau^{-1}\boldsymbol{D}) \tag{8}$$

where $\boldsymbol{D}$ is singular covariance matrix with generalised inverse given by equation (4). For a specific value of ρ and an intrinsic autoregression matrix R, the prior density of $\boldsymbol{\psi}$ can be specified as:

$$\pi(\boldsymbol{\psi}|\tau) \propto exp(-\frac{1}{2}\boldsymbol{\psi}'\boldsymbol{D}^{-}\boldsymbol{\psi}\tau) \tag{9}$$

The prior distribution of fixed effect $\boldsymbol{\beta}$ can be specified following the suggestion of Chaudhuri and Ghosh (2011) and Porter et al. (2015a) as,

$$\boldsymbol{\beta} \sim MVN(\tilde{\boldsymbol{\beta}}, g^{-1}\tau^{-1}I_p) \tag{10}$$

where g represents the Zellner prior (Zellner, 1988), which is evaluated at a fixed point estimate 10 (Chaudhuri & Ghosh, 2011), τ is the precision parameter and p is the number of covariates in the study and I is an identity matrix of dimension $p \times p$. We specify $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_{WLS}$, the weighted least squares estimate of $\boldsymbol{\beta}$ following the suggestion of Porter et al. (2015a). Then the prior density of β can be written as,

$$\pi(\boldsymbol{\beta}|\tau) \propto exp(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta_{WLS}})^{'}g\tau(\boldsymbol{\beta} - \boldsymbol{\beta_{WLS}})) \tag{11}$$

The prior distribution of precision parameter τ can be taken as an inverse gamma distribution (IG) (Porter et al., 2015a) as,

$$\tau \sim IG(\alpha_1, \alpha_2) \tag{12}$$

The prior density of τ can be written as,

$$\pi(\tau) \propto \tau^{1+\alpha_1} exp(-\frac{\alpha_2}{\tau}) \tag{13}$$

The samples form the posterior distributions are drawn utilising the above stated prior and EL weights instead of data likelihood applying a Markov Chain Monte Carlo (MCMC) algorithm described in next section.

2.3.2 MCMC Sampling Algorithm

This sampling algorithm is motivated by the one provided by Porter et al. (2015a). Hence there are similarities in the algorithm steps with differences in some models and equations induced by the new prior specification for the spatial random effects ψ_i.

1. **Obtaining starting values**
   Using the gmm package and the EL estimating equations (1) and (2), the maximum empirical likelihood estimators (MELE) for $\boldsymbol{\beta}$ are obtained. The initial values for $\boldsymbol{\beta}$ are chosen randomly from the prior distribution, weights w_i are set to $1/n$ and σ_i^2 is replaced using the sample variance

of the data for estimation purposes. Using the MELE of $\boldsymbol{\beta}$ and setting $\boldsymbol{\psi} = \vec{\mathbf{0}}$ gives the starting values of $\mu_i = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\psi}$. We calculate the empirical likelihood weights for the data which will satisfy the constraint:

$$W_\mu = \{\sum_{i=1}^{n} w_i = 1; w_i > 0 \ \forall \ i; \sum_{i=1}^{n} w_i m_j(y_i, \boldsymbol{\mu}) = 0 \ \forall \ j)\} \tag{14}$$

where $m_j(y_i, \boldsymbol{\mu})$ are the estimating equations ($j = 1, 2$) presented in (2) and (3). The calculation of the EL weights w_i can be made by constrained optimization of the above equations following the recommendations of Owen (2013).

2. **Sampling spatial random effects, $\boldsymbol{\psi}$**
To sample $\boldsymbol{\psi}$, we use a multivariate normal proposal for a block of size B. For the ith observation in the kth block, ψ_i^* can be chosen from the multivariate normal proposal as $\boldsymbol{\psi_k^*} \sim MVN(0, \boldsymbol{\Sigma_k})$ where the proposal covariance $\boldsymbol{\Sigma_k}$ (the $B \times B$ matrix for k-th block of size= B) can be tuned by using pilot chains (Gelman, Carlin, Stern, & Rubin, 1995). The proposed values are then utilised in the estimating equations below to generate weights w_i^*:

$$\sum_{i=1}^{n} w_i^* \{Y_i - \boldsymbol{x_i'}\boldsymbol{\beta} - \boldsymbol{\psi^*}\} = 0 \tag{15}$$

$$\sum_{i=1}^{n} \{w_i^* (Y_i - \boldsymbol{x_i'}\boldsymbol{\beta} - \boldsymbol{\psi^*})^2 / \sigma_i^2\} - 1 = 0 \tag{16}$$

To generate the set of weights, the updated estimate of $\boldsymbol{\beta}$ is used in the above equation using the initial values of w_i estimated in step 1. The elements of $\boldsymbol{\psi^*}$ in the block k are set as $\boldsymbol{\psi_k^*}$, keeping other values as the previous values of $\boldsymbol{\psi_i^*}$. If w_i^* does not satisfy equation (14), the previous values of the $\boldsymbol{\psi^*}$ are kept unchanged and we move to the next block. Otherwise, we perform a Metropolis-Hastings step with the following posterior density ratio

$$\gamma_{\boldsymbol{\psi}} = \frac{p(Y|\boldsymbol{\psi^*}, \beta)\pi(\boldsymbol{\psi^*}|\tau)}{p(Y|\boldsymbol{\psi^*_{(t-1)}}, \boldsymbol{\beta})\pi(\boldsymbol{\psi^*_{(t-1)}}|\tau)}$$

$$\gamma_{\boldsymbol{\psi}} = \frac{\prod_{i=1}^{n}(w_i^*) exp(-\frac{1}{2}\boldsymbol{\psi^{*\prime}}\boldsymbol{D^-}\boldsymbol{\psi^*}\tau)}{\prod_{i=1}^{n}(w^*_{i\,(t-1)}) exp(-\frac{1}{2}\boldsymbol{\psi^*_{(t-1)}}\boldsymbol{D^-}\boldsymbol{\psi^*_{(t-1)}}\tau)} \tag{17}$$

where, $t = 1, 2, 3, ...$ is the number of iterations, $\boldsymbol{\psi^*_{(t-1)}}$ is the value of $\boldsymbol{\psi}$ in the previous iteration. When $t = 1$, in the first iteration, $\boldsymbol{\psi^*_{(t-1)}} = \boldsymbol{\psi^*_0} = \vec{\mathbf{0}}$ and $w^*_{i\,(t-1)}$ is the EL weights generated at iteration t utilising equations (15) and (16).

To compute the ratio $\gamma_{\boldsymbol{\psi}}$, we need to estimate $\boldsymbol{D^-}$ using estimated value of ρ. To estimate the ρ, the local conditional moments from equations (6) and (7) can be used assuming local conditional distributions to be Gaussian (Leroux, 2000). An alternative to using a parametric distribution here to estimate ρ is to employ a generalized empirical likelihood (GEL) method using the two conditional moments. These can be computed using the gmm package in R (Donald, Imbens, & Newey, 2003). A third approach is to use fixed value of ρ (for example, ρ=0.99).

Notice that this step of sampling spatial random effects is similar to that of Porter et al. (2015a), except the prior specification used in posterior ratio $\gamma_{\boldsymbol{\psi}}$ uses the Leroux structure, which adds one more layer of complexity in terms of estimation of the variance covariance matrix $\boldsymbol{D}$ for the spatial random effects.

3. **Sampling the fixed effect, $\boldsymbol{\beta}$**
Using a multivariate normal proposal, $\boldsymbol{\beta}^* \sim MVN(\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma_\beta})$ with proposal covariance tuned on the basis of the pilot chains (Gelman et al., 1995), we sample $\boldsymbol{\beta}$ using a random-walk MH step. The estimating equations to generate weights w_i^* utilising the proposed values $\boldsymbol{\beta^*}$ in equations (15) and (16).

If the generated weights w_i^* verify the constraints (14), a Metropolis Hastings step is performed having posterior density ratio as:

$$\gamma_{\boldsymbol{\beta}} = \frac{p(Y|\boldsymbol{\psi^*}, \boldsymbol{\beta^*})\pi(\boldsymbol{\beta}^*|\tau)}{p(Y|\boldsymbol{\psi^*_{(t-1)}}, \boldsymbol{\beta^*_{(t-1)}})\pi(\boldsymbol{\beta^*_{(t-1)}}|\tau)}$$

$$\gamma_{\boldsymbol{\beta}} = \frac{\prod_{i=1}^{n} w_i^* exp(-(\frac{1}{2}(\boldsymbol{\beta^*} - \boldsymbol{\beta_{WLS}})'(\boldsymbol{\beta^*} - \boldsymbol{\beta_{WLS}})g\tau)}{\prod_{i=1}^{n} w_{i\,(t-1)}^* exp(-(\frac{1}{2}(\boldsymbol{\beta^*_{(t-1)}} - \boldsymbol{\beta_{WLS}})(\boldsymbol{\beta^*_{(t-1)}} - \boldsymbol{\beta_{WLS}})g\tau)} \tag{18}$$

where $\boldsymbol{\beta^*_{(t-1)}}$ and $w_{i\,(t-1)}^*$ are the values of $\boldsymbol{\beta^*}$ and w_i^* in the $(t-1)$th iteration respectively. For $t = 1$, $\boldsymbol{\beta^*_{(0)}}$ and $w_{i\,(0)}^*$ are replaced by the initial estimates of $\boldsymbol{\beta}$ and weights w_i generated using the initial $\boldsymbol{\beta}$. Also g is the Zellner prior (?, ?), set to 10 following Chaudhuri and Ghosh (2011) and $\boldsymbol{\beta_{WLS}}$ is the weighted least squares estimate of $\boldsymbol{\beta}$ (Porter et al., 2015a).

The proposed values are accepted if $\gamma_{\boldsymbol{\beta}>u_{\boldsymbol{\beta}}}$ where $u_{\boldsymbol{\beta}} \sim Unif(0,1)$. If equation (14) is not satisfied, previous values of $\boldsymbol{\beta}$, $\boldsymbol{\beta^*_{(t-1)}}$ are kept unchanged and the above MH step is not performed.

Notice that, this step is also similar with that proposed by Porter et al. (2015a). The only difference lies in using the values of $\boldsymbol{\psi^*}$ from the step 2, which was estimated considering the Leroux model structure.

4. **Sampling τ**
Following the suggestion of Porter et al. (2015a), τ will be sampled with a Gaussian proposal as $\tau^{*'} \sim N(\tau, \Sigma_\tau)$, with a proposal variance to be tuned based on pilot chains and accepted according to a Metropolis-Hastings step with posterior density ratio as:

$$\gamma_\tau = \frac{\pi(\boldsymbol{\beta^*}|\tau^*)\pi(\boldsymbol{\psi^*}|\tau^*)\pi(\tau^*)}{\pi(\boldsymbol{\beta^*_{(t-1)}}|\tau^*_{(t-1)})\pi(\boldsymbol{\psi^*_{(t-1)}}|\tau^*_{(t-1)})\pi(\tau^*_{(t-1)})}$$

$$\gamma_\tau = \frac{exp(-\frac{1}{2}\boldsymbol{\psi_i^{*'}} \boldsymbol{D^-} \boldsymbol{\psi_i^*}\tau^*)exp(-(\frac{1}{2}(\boldsymbol{\beta^*} - \boldsymbol{\beta'_{WLS}})(\boldsymbol{\beta^*} - \boldsymbol{\beta_{WLS}}))g\tau^*(\tau^*)^{1+\alpha_1}exp(-\frac{\alpha_2}{\tau^*})}{exp(-\frac{1}{2}\boldsymbol{\psi^*_{(t-1)}}\boldsymbol{D^-}\boldsymbol{\psi^*_{(t-1)}}\tau^*_{(t-1)})exp(-(\frac{1}{2}(\boldsymbol{\beta^*_{(t-1)}} - \boldsymbol{\beta_{WLS}})'(\boldsymbol{\beta^*_{(t-1)}} - \boldsymbol{\beta_{WLS}})g\tau^*_{(t-1)})\tau^{*\,1+\alpha_1}_{(t-1)}exp(-\frac{\alpha_2}{\tau^*_{(t-1)}})} \tag{19}$$

τ^* is accepted if $\gamma_\tau > u_\tau$, $u_\tau \sim Unif(0,1)$. The prior of τ is taken to be an Inverse Gamma distribution as $IG(\alpha_1, \alpha_2)$.

Notice that, this step is also similar to that given by (Porter et al., 2015a) with changes in the posterior density ratio due to the estimation of $\boldsymbol{\psi^*}$ in step 2.

5. Steps 2-4 are repeated until convergence.

After the convergence, the samples drawn for each of the parameters of interest $\boldsymbol{\psi}$, $\boldsymbol{\beta}$ and τ are stored as draws from the desired posterior distribution samples and used to obtain inferences.

2.3.3 Implementation

The BSLEL model was fitted using R. The MCMC iterations to be performed using MH steps coded in R.

For calculation of the EL weights in all the steps, we utilised the codes of Owen (2013) which are made available in http://statweb.stanford.edu/ owen/empirical/scel.R applying the corresponding estimating equations of the BSLEL model. For estimating the initial values of the fixed effects $\boldsymbol{\beta}$, the generalised empirical likelihood method was used and computation was performed via the R-package gmm which allows estimation of the maximum empirical likelihood estimator from a set of estimating equations and their gradients using the gel command.

The initial values of the parameters were randomly generated from the respective prior distributions. The computation is time consuming because for performing each MH step in each iteration in steps 2-4, optimization of the EL weights was undertaken and the constraint (8) was checked before performing the M-H step. We did not apply any parallelisation, but this can be implemented in a straightforward manner in order to reduce the computational time.

3 Application

For application of the proposed model, we use simulated data on lung cancer incidence (females) on 2147 small areas (SA2s according to 2011 ABS ASGS boundaries) across Australia. The simulated data reflect real cancer patterns, which were not available for privacy reasons. The dataset was compiled by the Cancer Council Queensland.

The covariate information is taken to be the urban or remoteness of each area obtained from the Australian Bureau of Statistics (ABS). The ABS provides information regarding level of urban/remoteness for each SA2 in five categories as: 1 = major city, 2 = inner regional, 3 = outer regional, 4 = remote, 5 = very remote. We reconstructed these categories into 2 classes: 1= major cities, 0= regional or remote.

For the application of the BSLEL model, the response variable Y_i denotes the number of incidence of lung cancer among females in ith area, ($i = 1, 2, ..., 2147$). The covariates x_i is 1, if ith area is a major city and 0 if ith area is not major city(regional or remote).

In the present study, we will use different fixed values for ρ and compare the model performance for each of the ρ. To begin with, we use $\rho = 0.99$. Then the generalized inverse of the variance covariance matrix D^{-} can be estimated using equation (1). For the MCMC iterations in steps 2-4, we decided a burn in period of 5000 iterations and we ran 50000 iterations to obtain the posterior sample for the parameters of interest. For drawing posterior samples form spatial random effect $\boldsymbol{\psi}$, MH steps are carried out in block (step 1 in section 2.3.1). The block formation can be done in several ways (Chib & Ramamurthy, 2010) and the block size B can be chosen accordingly. Porter et al. (2015a) chose B to be 15. For a larger number of areas in the dataset, smaller block size will result in computationally expensive algorithms. Since the proposed model to be applied on data of 2147 small areas, we choose block size $B = 200$. However other choices of the block size can be made and a comparison can be conducted in terms of computational time and accuracy.

For comparision of the outputs of the proposed model, a parametric Leroux model using CARBayes package in R is fitted using same value of ρ. A BSHEL model proposed by (Porter et al., 2015a) is also fitted for this data and a comparison of the model outputs is shown in the next section. The priors of the proposed model were as detailed in the section 2.3.

The first two prior specifications proposed by (Chaudhuri & Ghosh, 2011) are not considered in the present study for comparison as BSHEL model has already been shown to have similar and in some cases improved performance (Porter et al., 2015a). So the present case study compared the performance of proposed BSLEL model with the BSHEL model (Porter et al., 2015a).

4 Results and Discussion

The proposed BSLEL spatial model is fitted to simulated incidence data on lung cancer (females) for 2147 small areas in Australia. As described in section 3, the only covariate considered for this case study is urban/remote status of an area. For comparison of the model outputs, the BSHEL (Porter et al., 2015a) model is also fitted to the same case study. Table 1 shows summary of posterior samples obtained by the two models for fixed covariate effects and precision parameters. Table 1 also shows the Mean Squared prediction error of each model calculated by carrying out five leave-one-out validations. Since there are 2147 components of each sample of random effects, these are not displayed in the table.

From the outputs displayed in Table 1, it can be noted that the two models produce similar posterior estimates. The mean squared prediction error (MSPE) are also very close for both the models, MSPE of proposed model is a lower than that of BSHEL model proposed by Porter et al.

(2015a). However, these are some preliminery results and more analysis will be made in order to compare the two models.

Table 1: Summary of posterior samples of intercept (β_0), slope coefficient (β_1) and precision parameter (τ)

		BSLEL	**Model**		
Parameter	mean	sd	Median	2.5% Quantile	97.5% Quantile
β_0(Intercept)	1.200614	0.0999831	1.1998872	1.0066630	1.3976578
β_1(Covariate effect)	0.121195	0.1009070	0.1213618	-0.0771596	0.3186009
τ(Precision)	1.630714	0.7354304	2.3210458	0.8534575	2.3369149
		MSPE	10.7191584		
		BSHEL	**Model**		
parameter	mean	sd	Median	2.5% Quantile	97.5% Quantile
β_0(Intercept)	1.2383851	0.0974191	1.2673235	1.0238560	1.3761125
β_1(Covariate effect)	0.1504099	0.0943995	0.1839951	-0.0610027	0.2940694
τ(Precision)	1.8374508	0.6942962	2.3236267	0.8537023	2.3371443
		MSPE	10.9371801		

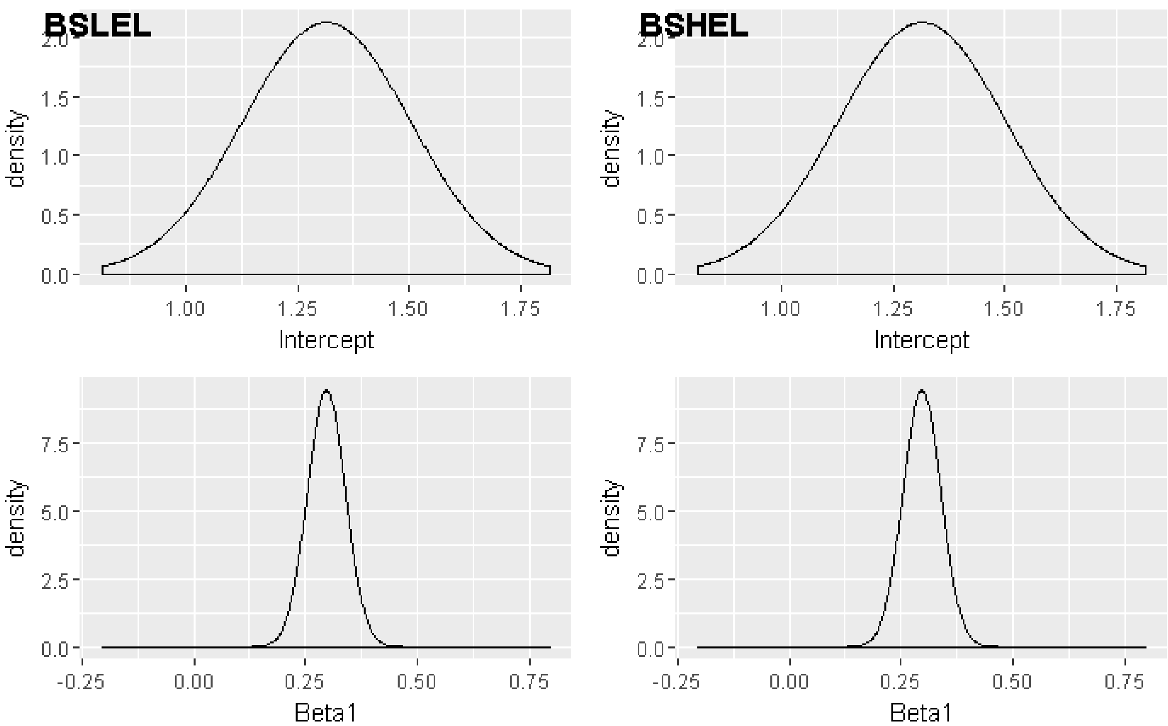

Figure 1: Posterior Densities for Fixed effects for proposed BSLEL model and BSHEL Model

The fitted values of cancer incidence can be used to obtained Standardized incidence Ratios (SIR) for lung cancer (females) utilising the expected incidence for each small area available in the simulated data set. The standardised incidence ratio (SIR) is a standard parameter to compare the disease incidence of a cohort relative to the population. The SIRs for raw data and modelled data are obtained by dividing the observed incidence and fitted incidence by the expected incidence respectively. We estimated SIR for raw data and modelled incidence for the proposed model BLSEL. If the SIR is equal or close to 1 for an area, it means that the area has average risk of cancer incidence and it is same as the national average of Australia, while a SIR value greater than

1 implies higher risk of cancer incidence and less than 1 implies lower risk of cancer incidence than the national average (Cramb, Mengersen, & Baade, 2011). The spatial map showing the raw and modelled SIR dividing in categories as: very low, low, average, high and very high (Cramb et al., 2011) are shown in Figure 2 and 3 respectively. The amount of prediction error associated with the proposed model can be seen in figure 3, as there are a few wrong predictions.

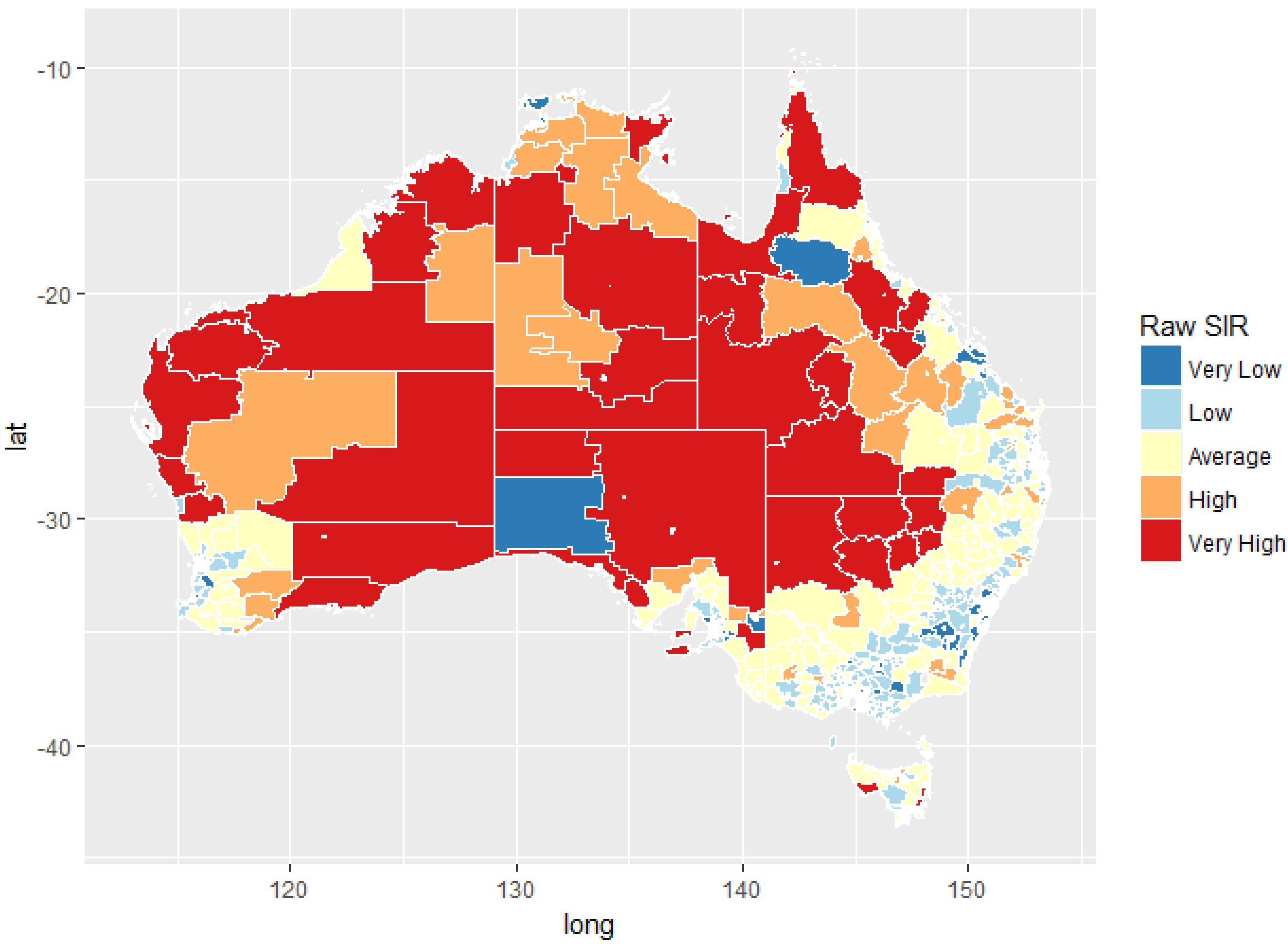

Figure 2: Raw SIR for Lung Cancer (females)

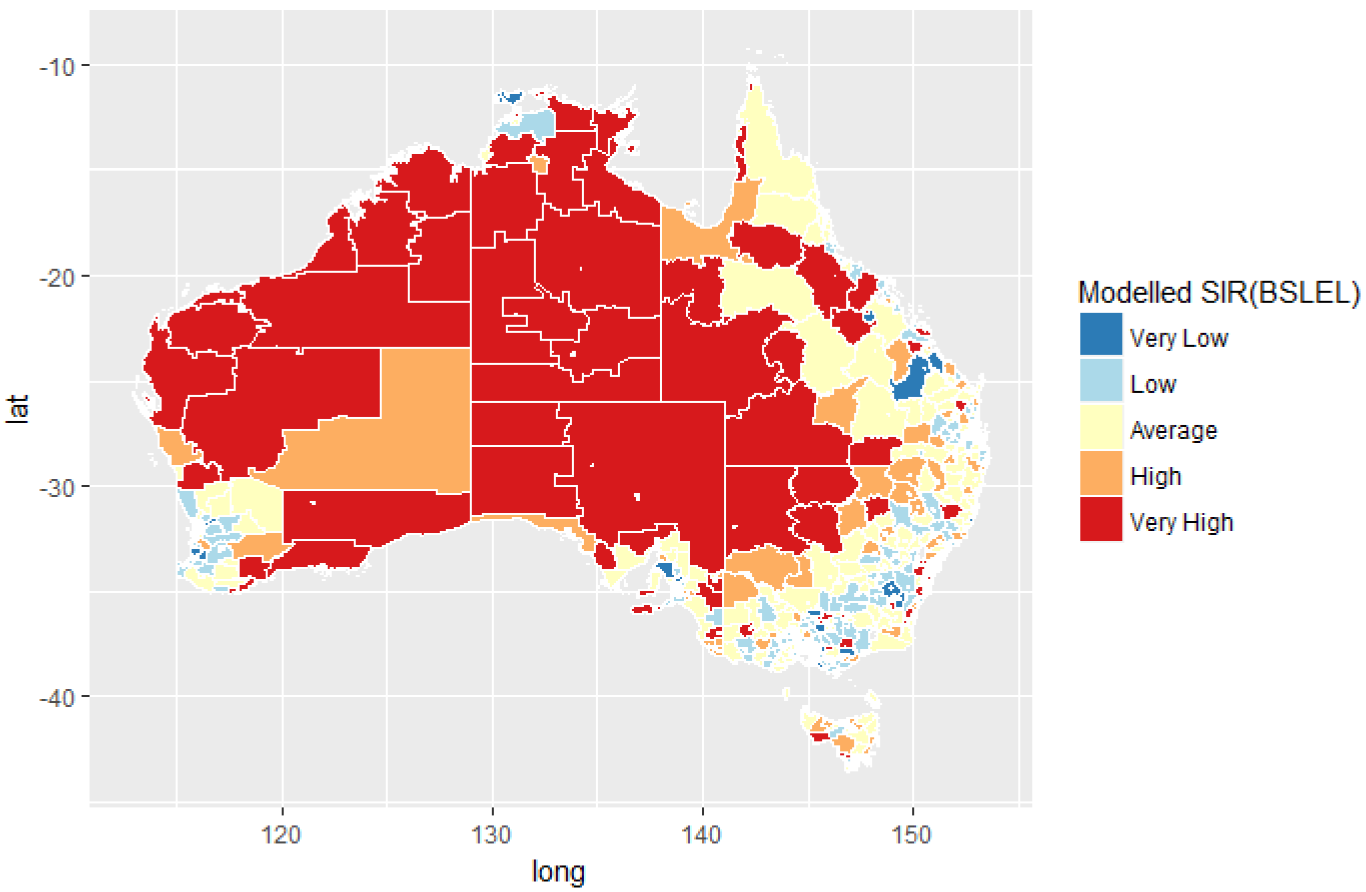

Figure 3: Modelled SIR using BSLEL Model for Lung Cancer (females)

5 Conclusions and Future Work

This is to mention here that the results shown in section 4 are preliminary results. Since, we will continue this analysis and also will fit the proposed model to other cancer datasets to evaluate the model performance, we are not making any conclusions from the results at this stage. This work is ongoing and there are some future works mentioned below.

For reducing the computational time of the proposed algorithm in section 2.3.2, more effective algorithms may be developed using Hamiltonian Monte Carlo proposed by Chaudhuri, Mondal, and Yin (2017).

After the evaluation of the proposed Bayesian Semiparametric Leroux Empirical Likelihood (BSLEL) spatial model, we will extend the proposed model to estimate the spatial dependence parameter ρ simultaneously with the other parameters of interest. Theoretical justification and application of the extended model will be made in an upcoming paper.

References

Besag, J., York, J., & Mollie, A. (1993). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*, 1–59.

Chaudhuri, S., & Ghosh, M. (2011). Empirical likelihood for small area estimation. *Biometrika*, 473–480.

Chaudhuri, S., Mondal, D., & Yin, T. (2017). Hamiltonian monte carlo sampling in bayesian empirical likelihood computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(1), 293–320.

Chib, S., & Ramamurthy, S. (2010). Tailored randomized block mcmc methods with application to dsge models. *Journal of Econometrics*, *155*(1), 19–38.

Chib, S., Shin, M., & Simoni, A. (2017). Bayesian estimation and comparison of moment condition models. *Journal of the American Statistical Association*(just-accepted).

Cramb, S. M., Mengersen, K. L., & Baade, P. D. (2011). Atlas of cancer in queensland:geographical variation in incidence and survival, 1998 to 2007. *Viertel Centre for Research in Cancer Control, Cancer Council Queensland.Brisbane, Queensland*.

Donald, S. G., Imbens, G. W., & Newey, W. K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, *117*(1), 55–93.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *canadian Journal of Statistics*, *9*(2), 139–158.

Fay III, R. E., & Herriot, R. A. (1979). Estimates of income for small places: an application of james-stein procedures to census data. *Journal of the American Statistical Association*, *74*(366a), 269–277.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Hughes, J., & Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(1), 139–159.

Lazar, N. A. (2003). Bayesian empirical likelihood. *Biometrika*, *90*(2), 319–326.

Lee, D. (2011). A comparison of conditional autoregressive models used in bayesian disease mapping. *Spatial and Spatio-Temporal Epidemiology*, *2*(2), 79–89.

Leroux, B. G. (2000). Modelling spatial disease rates using maximum likelihood. *Statistics in Medicine*, *19*(17-18), 2321–2332.

Leroux, B. G., Lei, X., & Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials* (pp. 179–191). Springer.

Lu, H., Reilly, C. S., Banerjee, S., & Carlin, B. P. (2007). Bayesian areal wombling via adjacency modeling. *Environmental and Ecological Statistics*, *14*(4), 433–452.

Mengersen, K. L., Pudlo, P., & Robert, C. P. (2013). Bayesian computation via empirical likelihood. *Proceedings of the National Academy of Sciences*, *110*(4), 1321–1326.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*(2), 237–249.

Owen, A. B. (2001). *Empirical likelihood*. Chapman and Hall/CRC.

Owen, A. B. (2013). Self-concordance for empirical likelihood. *Canadian Journal of Statistics*, *41*(3), 387–397.

Porter, A. T., Holan, S. H., & Wikle, C. K. (2015a). Bayesian semiparametric hierarchical empirical likelihood spatial models. *Journal of Statistical Planning and Inference*, *165*, 78–90.

Porter, A. T., Holan, S. H., & Wikle, C. K. (2015b). Multivariate spatial hierarchical bayesian empirical likelihood methods for small area estimation. *Stat*, *4*(1), 108–116.

Rampaso, R. C., de Souza, A. D. P., & Flores, E. F. (2016). Bayesian analysis of spatial data using different variance and neighbourhood structures. *Journal of Statistical Computation*

and Simulation, *86*(3), 535–552.
Schennach, S. M. (2005). Bayesian exponentially tilted empirical likelihood. *Biometrika*, *92*(1), 31–46.
Stern, H. S., & Cressie, N. (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine*, *19*(1718), 2377–2397.
Ugarte, M. D., Adin, A., Goicoa, T., & Militino, A. F. (2014). On fitting spatio-temporal disease mapping models using approximate bayesian inference. *Statistical Methods in Medical Research*, *23*(6), 507–530.
Ver Hoef, J. M., & Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology*, *25*(2), 219–240.
Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60–62.
Zellner, A. (1988). Bayesian analysis in econometrics. *Journal of Econometrics*, *37*(1), 27–50.