Ann Inst Stat Math (2013) 65:457–472
DOI 10.1007/s10463-012-0377-6

Objective Bayesian analysis for CAR models

Cuirong Ren · Dongchu Sun

Received: 29 July 2011 / Revised: 19 June 2012 / Published online: 29 August 2012

Abstract Objective priors, especially reference priors, have been studied extensively for spatial data in the last decade. In this paper, we study objective priors for a CAR model. In particular, the properties of the reference prior and the corresponding posterior are studied. Furthermore, we show that the frequentist coverage probabilities of posterior credible intervals depend only on the spatial dependence parameter ρ, and not on the regression coefficient or the error variance. Based on the simulation study for comparing the reference and Jeffreys priors, the performance of two reference priors is similar and better than the Jeffreys priors. One spatial dataset is used for illustration.

Keywords Conditional autoregressive · Jeffreys prior · Reference prior · Integrated likelihood · Propriety of posterior

1 Introduction

Conditional autoregressive (CAR) models were introduced by Besag (1974) almost 40 years ago but have been extensively used for the analysis of spatial areal data only in the last two decades. This resurgence arises from the convenience of their employment in the context of Gibbs sampling and more general Markov Chain Monte

C. Ren (✉)
Department of Plant Science, South Dakota State University, Brookings, SD 57007, USA
e-mail: Cuirong.Ren@sdstate.edu

D. Sun
School of Finance and Statistics, East China Normal University,
500 Dongchuan Road, Shanghai 200241, China

D. Sun
Department of Statistics, University of Missouri-Columbia, Columbia, MO 65211, USA

Carlo (MCMC) methods for fitting certain classes of hierarchical spatial models. Since then, these models have been used to analyze data in many areas, such as epidemiology, demography, economy and geography.

The most common method used to estimate the parameters in the CAR model was maximum likelihood (e.g. Cressie and Chan 1989; Richardson et al. 1992; Cressie et al. 2005). Although Bayesian analyses of CAR models have been extensively used to estimate latent variables and spatially varying random effects in the context of hierarchical models, not much has been done on Bayesian analysis of CAR models to describe the observed data (with only rare exceptions, e.g., Bell and Broemeling 2000). This may be due to a lack of knowledge on adequate priors for these models and frequentist properties of the resulting Bayesian procedures. De Oliveira (2012) probably was the first to propose default Bayesian analyses for CAR models and to study some of their properties, but he only considered two versions of the Jeffreys prior, the independence Jeffreys and the Jeffreys-rule priors, for parameters.

It is well known that in spite of success in using Jeffreys priors for one-parameter problems (Welch and Peers 1963; Berger and Bernardo 1992) gave examples where the Jeffreys-rule prior provided inconsistent estimates in some multiparameter problems. Therefore, they suggested arranging the parameters in order according to the importance of inference, and proposed an algorithm to derive a reference prior. This algorithm depends on the Fisher information matrix and is based on the asymptotic normality of the posterior distribution. This reference prior has been commonly used in numerous applications in non-geostatistics context. For example, see Ye (1994), Sun and Ye (1995) and Berger et al. (1998).

In addition, based on the simulation studies in geostatistics contexts for point data, such as Berger et al. (2001) and Ren et al. (2012), in terms of frequentist performance, the reference priors have a reasonable performance, but Jeffreys-rule prior can be seriously inadequate. Therefore, it is very interesting and necessary to reconsider the reference prior for CAR models. Although De Oliveira (2012) finally recommended the independence Jeffreys prior as one default objective prior, we should realize that the independence Jeffreys prior does not always yield a proper posterior. This is another reason why we are studying the reference prior.

We propose the reference priors for the CAR model including one given in Remark 4 of De Oliveira (2012), but he had difficulty verifying properties such as posterior propriety. Perhaps the difficulty arises from the expression of the prior when it was expressed as the function of eigenvalues, which makes it very hard to find the limiting behavior of marginal likelihood and the posterior when the spatial parameter approaches the boundaries of its range. In our paper, the reference priors are expressed in terms of the traces of matrices (see Proposition 2 in Sect. 2). The advantage of this method is the ability to find the limiting behaviors. We will derive the results on propriety of the resulting posterior distributions. In addition, we will perform a simulation experiment to compare frequentist properties of inferences about the parameters based on the Jeffreys and the proposed reference priors.

The organization of the paper is as follows: In Sect. 2, we give a brief description of the CAR model, summarize two versions of the Jeffreys prior in De Oliveira (2012), and consider commonly used reference priors. The property of the reference priors and the propriety of the corresponding posterior distribution is given. In Sect. 3, it is shown that the frequentist coverage probabilities of Bayesian credible intervals under a large class of priors depends only on the spatial parameter. Numerical simulations are given to compare the frequentist coverage probabilities of Bayesian credible intervals for four objective priors, the Jeffreys-rule, independence Jeffreys and two Type I reference priors. Finally, one proposed reference prior is illustrated by an example. The summary and comments are also given.

2 Main results

2.1 A CAR model

We consider a Gaussian Markov random field, where the study area is partitioned into n regions, indexed by integers $1, 2, \ldots, n$. For region i, the variable of interest, y_i, is observed, and a set of p explanatory variables, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$, is pre-specified. For this class of models, spatial association is specified through a set of conditional distributions,

$$(y_i \mid y_j, j \neq i) \sim N\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \sum_{j=1}^{n} W_{ij}(\rho)(y_j - \boldsymbol{x}_j'\boldsymbol{\beta}), \delta_1\right), \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$ are unknown regression parameters, $\delta_1 > 0$ and $W_{ij}(\rho)$ are covariance parameters, with $W_{ii}(\rho) = 0$ for all i. It is from Besag (1974) that the joint distribution of $\boldsymbol{y}$ is uniquely determined by the full conditional distributions (1). Often, $W_{ij}(\rho)$ is a linear function of the weight C_{ij}. That is,

$$W_{ij}(\rho) = \rho C_{ij}, \quad \text{for all } i, j. \tag{2}$$

The matrix $\boldsymbol{C} = (C_{ij})_{n\times n}$ is often called a *Weight Matrix* or *Proximity Matrix*.

Frequently the proximity matrix $\boldsymbol{C}$ is symmetric and known. Common choices of $\boldsymbol{C}$ are as follows:

- *Adjacency matrix* $C_{ij} = 1$ if region i and region j share common boundary.
- *k-neighbor Adjacency matrix* $C_{ij} = 1$ if region j is one of the k nearest neighbors of region i.
- *Distance matrix* C_{ij} = the distance between centroids of regions i and j.

The case of *Adjacency* matrix was introduced in Clayton and Kaldor (1987) and the other two cases can be found in Cressie (1993) and Rue and Held (2005).

If we write $\boldsymbol{y} = (y_1, \ldots, y_n)'$ and $\boldsymbol{X} = (x_{ij})_{n\times p}$, (1) and (2) are equivalent to

$$\boldsymbol{y} \sim N_n(\boldsymbol{X\beta}, \delta_1(\boldsymbol{I}_n - \rho\boldsymbol{C})^{-1}), \tag{3}$$

here ρ is often called a 'spatial parameter'. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the ordered eigenvalues of $\boldsymbol{C}$. Because $\mathrm{tr}(\boldsymbol{C}) = 0$, it is clearly that $\lambda_1 < 0 < \lambda_n$. The range of the spatial parameter ρ is $(\lambda_1^{-1}, \lambda_n^{-1})$, including 0 as an interior point. The likelihood function of parameters $(\rho, \delta_1, \boldsymbol{\beta})$ in model (3) is given by:

$$L(\rho, \delta_1, \boldsymbol{\beta}; \boldsymbol{y}) = \frac{1}{(2\pi\delta_1)^{n/2}|\boldsymbol{\Sigma}_\rho|^{1/2}} \exp\left\{-\frac{(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}_\rho^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})}{2\delta_1}\right\}, \tag{4}$$

where $\boldsymbol{\Sigma}_\rho^{-1} = \boldsymbol{I}_n - \rho\boldsymbol{C}$.

2.2 Common objective priors

Denote the eigenvector of $\boldsymbol{C}$ in model (3) corresponding to eigenvalue λ_i by $\boldsymbol{u}_i$, $i = 1, 2, \ldots, n$. De Oliveira (2012) obtained two versions of the Jeffreys priors and present the properties and propriety for the marginal Jeffreys priors and the corresponding posterior distributions. One can find these results from Theorem 1, Lemma 2, and Corollaries 1 and 2 in De Oliveira (2012). We summarize the results in the following lemma.

Lemma 1 *Consider the CAR model* (3).

(a) *The Jeffreys-rule prior* $\pi^J(\rho, \delta_1, \boldsymbol{\beta})$ *is given by*

$$\pi^J(\rho, \delta_1, \boldsymbol{\beta}) \propto \frac{1}{\delta_1^{1+p/2}}\sqrt{|\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|[n\,\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{C})^2 - \{\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{C})\}^2]}. \tag{5}$$

(b) *The independence Jeffreys prior treating* (ρ, δ_1) *and* $\boldsymbol{\beta}$ *as independence,* π^{IJ}*, is*

$$\pi^{IJ}(\rho, \delta_1, \boldsymbol{\beta}) \propto \frac{1}{\delta_1}\sqrt{n\,\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{C})^2 - \{\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{C})\}^2}. \tag{6}$$

(c) *Suppose* λ_1 *and* λ_n *are simple eigenvalues. Then as* $\rho \to \lambda_1^{-1}$*, it holds that*

$$\pi^J(\rho) = \begin{cases} O\{(1-\lambda_1\rho)^{-1/2}\}, & \text{if } \boldsymbol{u}_1 \in \mathcal{C}(\boldsymbol{X}), \\ O\{(1-\lambda_1\rho)^{-1}\}, & \text{if } \boldsymbol{u}_1 \notin \mathcal{C}(\boldsymbol{X}), \end{cases}$$
$$\pi^{IJ}(\rho) = O\{(1-\lambda_1\rho)^{-1}\},$$

where $\mathcal{C}(\boldsymbol{X})$ *is the column space of* $\boldsymbol{X}$ *consisting of all linear combinations of column vectors of* $\boldsymbol{X}$*. The same results hold as* $\rho \to \lambda_n^{-1}$ *when* λ_1 *and* $\boldsymbol{u}_1$ *are replaced by* λ_n *and* $\boldsymbol{u}_n$*, respectively.*

(d) *The marginal Jeffreys-rule prior* $\pi^J(\rho)$ *is unbounded. Furthermore, it is integrable when both* $\boldsymbol{u}_1$ *and* $\boldsymbol{u}_n$ *are in* $\mathcal{C}(\boldsymbol{X})$*, while it is not integrable when either* $\boldsymbol{u}_1$ *or* $\boldsymbol{u}_n$ *is not in* $\mathcal{C}(\boldsymbol{X})$*.*

(e) *The marginal independence Jeffreys prior* $\pi^{IJ}(\rho)$ *is unbounded and not integrable.*

(f) π^J *yields a proper posterior distribution.*

(g) *When neither $\boldsymbol{u}_1$ nor $\boldsymbol{u}_n$ is in $\mathcal{C}(\boldsymbol{X})$, π^{IJ} yields a proper posterior distribution, while it yields an improper posterior distribution when either $\boldsymbol{u}_1$ or $\boldsymbol{u}_n$ is in $\mathcal{C}(\boldsymbol{X})$.*

The method for finding the reference priors based on Berger and Bernardo (1992) algorithm can be applied to model (4) by specifying the order of parameter. We summarize these in the following proposition and the proof is given in Appendix A.

Proposition 1 *Consider the CAR model* (3).

(a) *The reference priors with the orderings $\{(\rho, \delta_1), \boldsymbol{\beta}\}$ or $\{\boldsymbol{\beta}, (\rho, \delta_1)\}$ or $\{\rho, \delta_1, \boldsymbol{\beta}\}$ or $\{\rho, \boldsymbol{\beta}, \delta_1\}$ or $\{\boldsymbol{\beta}, \rho, \delta_1\}$ are all the same as π^{IJ}. Here the ordering $\{\rho, \delta_1, \boldsymbol{\beta}\}$ means that ρ is the most important or the parameter of interest, δ_1 is less important, and $\boldsymbol{\beta}$ is least important.*
(b) *The reference priors with the orderings $\{\delta_1, \rho, \boldsymbol{\beta}\}$ or $\{\delta_1, \boldsymbol{\beta}, \rho\}$ or $\{\boldsymbol{\beta}, \delta_1, \rho\}$ are*

$$\pi^R(\rho, \delta_1, \boldsymbol{\beta}) = \pi^R(\rho)/\delta_1, \tag{7}$$

where $\pi^R(\rho) \propto \sqrt{\mathrm{tr}(\Sigma_\rho C)^2}$.
(c) *Suppose λ_1 and λ_n are simple eigenvalues. Then as $\rho \to \lambda_1^{-1}$, it holds that*

$$\pi^R(\rho) = O\{(1 - \lambda_1 \rho)^{-1}\}.$$

The same results hold as $\rho \to \lambda_n^{-1}$ when λ_1 and $\boldsymbol{u}_1$ are replaced by λ_n and $\boldsymbol{u}_n$, respectively. Thus, when neither $\boldsymbol{u}_1$ nor $\boldsymbol{u}_n$ is in $\mathcal{C}(\boldsymbol{X})$, π^R yields a proper posterior distribution, while it yields an improper posterior distribution when either $\boldsymbol{u}_1$ or $\boldsymbol{u}_n$ is in $\mathcal{C}(\boldsymbol{X})$.

2.3 The "exact" reference priors

If we specify (ρ, δ_1) as the parameter of interest and $\boldsymbol{\beta}$ as the nuisance parameter in applying the reference prior method, then $\pi_*^R(\boldsymbol{\beta}|\rho, \delta_1) = 1$ since this is the conditional Jeffreys-rule (or reference) prior in model (4) when (ρ, δ_1) is assumed to be known. Thus, by factoring the prior distribution, we have

$$\pi_*^R(\rho, \delta_1, \boldsymbol{\beta}) = \pi_*^R(\boldsymbol{\beta} \mid \rho, \delta_1)\pi_*^R(\rho, \delta_1) = \pi_*^R(\rho, \delta_1),$$

where $\pi_*^R(\rho, \delta_1)$ is computed using the Jeffreys-rule prior based on the following marginal model defined via the following integrated likelihood

$$\begin{aligned} L_*(\rho, \delta_1; \boldsymbol{y}) &= \int_{\mathbb{R}^p} L(\rho, \delta_1, \boldsymbol{\beta}; \boldsymbol{y}) \pi_*^{Ri}(\boldsymbol{\beta} \mid \rho, \delta_1)\, \mathrm{d}\boldsymbol{\beta} \\ &\propto \delta_1^{-(n-p)/2} |\boldsymbol{\Sigma}_\rho|^{-1/2} |\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|^{-1/2} \exp\left\{-\frac{S^2}{2\delta_1}\right\}, \end{aligned} \tag{8}$$

where $S^2 = \boldsymbol{y}'\boldsymbol{R}_\Sigma \boldsymbol{y}$, and $\boldsymbol{R}_\Sigma = \boldsymbol{\Sigma}_\rho^{-1} - \boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}$. L_* is named as Type I integrated likelihood, which is used to distinguish it from L_{**} called as Type

II integrated likelihood in the following. The corresponding priors are called Type I (II) reference priors if they are derived from Type I (II) integrated likelihood.

Based on the result in Harville (1974), there is a particular transformation of the data which has sampling distribution proportional to (8), and hence it is legitimate to calculate the associated Jeffreys-rule prior from (8). The results in Part (a) in the following proposition can be derived from a result in Berger et al. (2001) and found from Remark 4 in De Oliveira (2012), but it is expressed in terms of the traces of matrices.

Proposition 2 *Consider the model with sampling distribution* (4).

(a) *Type I reference prior distribution with orderings* $\{(\rho, \delta_1), \boldsymbol{\beta}\}$ *and* $\{\rho, \delta_1, \boldsymbol{\beta}\}$ *is given by:*

$$\pi_*^{R1}(\rho, \delta_1, \boldsymbol{\beta}) \propto \frac{1}{\delta_1} \pi_*^{R1}(\rho), \tag{9}$$

where

$$\pi_*^{R1}(\rho) \propto [(n-p)\mathrm{tr}(\boldsymbol{\Sigma}_\rho \boldsymbol{R}_\Sigma \boldsymbol{\Sigma}_\rho \boldsymbol{C})^2 - \{\mathrm{tr}(\boldsymbol{\Sigma}_\rho \boldsymbol{R}_\Sigma \boldsymbol{\Sigma}_\rho \boldsymbol{C})\}^2]^{1/2}. \tag{10}$$

(b) *Type I reference prior distribution with ordering* $\{\delta_1, \rho, \boldsymbol{\beta}\}$ *is given by:*

$$\pi_*^{R2}(\rho, \delta_1, \boldsymbol{\beta}) \propto \frac{1}{\delta_1} \pi_*^{R2}(\rho), \tag{11}$$

where

$$\pi_*^{R2}(\rho) \propto \{\mathrm{tr}(\boldsymbol{\Sigma}_\rho \boldsymbol{R}_\Sigma \boldsymbol{\Sigma}_\rho \boldsymbol{C})^2\}^{1/2}. \tag{12}$$

With the above proposition, one can obtain the following conclusions. The proof is given in Appendix B.

Proposition 3 *Suppose* λ_1 *and* λ_n *are simple eigenvalues. Then as* $\rho \to \lambda_1^{-1}$*, for* $i = 1, 2$*, there exists a positive constant* d *such that*

$$\pi_*^{Ri}(\rho) \le \begin{cases} d, & \text{if } \boldsymbol{u}_1 \in \mathcal{C}(\boldsymbol{X}), \\ d(1-\lambda_1\rho)^{-1}, & \text{if } \boldsymbol{u}_1 \notin \mathcal{C}(\boldsymbol{X}). \end{cases} \tag{13}$$

The same results hold as $\rho \to \lambda_n^{-1}$ *when* λ_1 *and* $\boldsymbol{u}_1$ *are replaced by* λ_n *and* $\boldsymbol{u}_n$*, respectively.*

Furthermore, we can show that the reference prior of δ_1 is a scale invariant prior $1/\delta_1$. We then derive Type II marginal likelihood of ρ,

$$L_{**}(\rho; \boldsymbol{y}) = \int_0^\infty \int_{\mathbb{R}^p} L(\rho, \delta_1, \boldsymbol{\beta}; \boldsymbol{y}) \frac{1}{\delta_1} \, \mathrm{d}\boldsymbol{\beta} \, \mathrm{d}\delta_1 \propto |\boldsymbol{\Sigma}_\rho|^{-1/2} |\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|^{-1/2} (S^2)^{-(n-p)/2}. \tag{14}$$

From Theorem 1 (the equivalence theory) in Ren et al. (2012), we have that $L_{**}(\rho; \boldsymbol{y})$ is essentially a proper density of the $n - p - 1$ dimensional random variable and therefore it is legitimate to find the Fisher information matrix of ρ based on it. Similar to Proposition 5 in Ren et al. (2012), one can obtain Type II reference prior, which is equal to π_*^{R1}.

Theorem 1 *Consider the model with sampling distribution* (4).

(a) *The marginal exact reference prior $\pi_*^{Ri}(\rho), i = 1, 2$ is bounded when both $\boldsymbol{u}_1$ and $\boldsymbol{u}_n$ are in $\mathcal{C}(\boldsymbol{X})$.*
(b) *The posterior distribution is proper under the prior π_*^{Ri}, for $i = 1, 2$.*

The result in Theorem 1 (a) follows from Proposition 3 and the rest of the proof is similar to Corollary 1 in De Oliveira (2012), so it is omitted.

Remark 1 All these objective priors belong to the following class of improper priors for $(\rho, \delta_1, \boldsymbol{\beta}) \in \boldsymbol{\Omega} = (\lambda_1^{-1}, \lambda_n^{-1}) \times (0, \infty) \times \mathbb{R}^p$ of the form, which was also introduced by De Oliveira (2012),

$$\pi(\rho, \delta_1, \boldsymbol{\beta}) \propto \frac{\pi(\rho)}{\delta_1^a}, \quad (\rho, \delta_1, \boldsymbol{\beta}) \in \boldsymbol{\Omega} = (\lambda_1^{-1}, \lambda_n^{-1}) \times (0, \infty) \times \mathbb{R}^p, \tag{15}$$

where a is a real value and $\pi(\rho)$ is the marginal prior of ρ with support $(\lambda_1^{-1}, \lambda_n^{-1})$.

The corresponding marginal prior for ρ is denoted by the same notation as the prior for $(\rho, \delta_1, \boldsymbol{\beta})$. For example, from (5) in Proposition 2, the Jeffreys-rule prior can be written as

$$\pi^J(\rho, \delta_1, \boldsymbol{\beta}) = \frac{\pi^J(\rho)}{\delta_1^{1+p/2}}, \quad \text{where } \pi^J(\rho) \propto \sqrt{|\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|[n\,\mathrm{tr}(\boldsymbol{\Sigma}_\rho \boldsymbol{C})^2 - \{\mathrm{tr}(\boldsymbol{\Sigma}_\rho \boldsymbol{C})\}^2]}.$$

Under the prior (15), a standard calculation yields

$$\begin{aligned} L_{**a}(\rho; \boldsymbol{y}) &= \int_0^\infty \int_{\mathbb{R}^p} L(\rho, \delta_1, \boldsymbol{\beta}; \boldsymbol{y}) \frac{1}{\delta_1^a} d\boldsymbol{\beta} d\delta_1 \\ &= \frac{\Gamma((n-p)/2)}{(2\pi)^{(n-p)/2}} |\boldsymbol{\Sigma}_\rho|^{-1/2} |\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|^{-1/2} (S^2)^{-((n-p)/2+a-1)}. \end{aligned} \tag{16}$$

Therefore, the joint posterior distribution of $(\rho, \delta_1, \boldsymbol{\beta})$ is proper if and only if

$$0 < \int_{\lambda_1^{-1}}^{\lambda_n^{-1}} L_{**a}(\rho; \boldsymbol{y}) \pi(\rho)\, \mathrm{d}\rho < \infty. \tag{17}$$

3 Comparison of the reference and Jeffreys priors

We first introduce a method that will simplify the computation tremendously. As Paulo (2005) pointed out, evaluating any objective prior at a particular value of parameters is a computationally intensive task since it involves computing each of the matrices' entries in the likelihood functions. Therefore, there is a considerable need to develop new methods that make computation less expensive and more feasible. We use this computational method to investigate the frequentist coverage of equal-tailed credible intervals for one parameter of interest ρ when either the Jeffreys or reference prior is used. The closer to the nominal level this frequentist coverage is, the 'better' the prior is. Finally, we will analyze a real dataset.

3.1 Frequentist coverage probabilities

Suppose we are interested in $\tau = \tau(\boldsymbol{\xi})$, a function of the parameter $\boldsymbol{\xi} = (\rho, \delta_1, \boldsymbol{\beta})$. Note that τ could be a function of ρ only. For example, τ could be ρ.

For the fixed $\boldsymbol{\xi} = \boldsymbol{\xi}^* \equiv (\rho^*, \delta_1^*, \boldsymbol{\beta}^*)$, we simulate the data based on $\boldsymbol{y} \mid \boldsymbol{\xi}^*$. For any $\alpha \in (0, 1)$, let $\tau_\alpha(\boldsymbol{y})$ be the α-posterior quantile of τ given $\boldsymbol{y}$. That is,

$$P(\tau^* < \tau_\alpha(\boldsymbol{y}) \mid \boldsymbol{y}) = \alpha, \quad \forall \alpha \in (0, 1). \tag{18}$$

Here the probability is computed based on the marginal posterior distribution of τ given $\boldsymbol{y}$. We then consider the frequentist coverage of the one-sided credible interval $(\tau_L, \tau_\alpha(\boldsymbol{y}))$, i.e.,

$$P_{\boldsymbol{\xi}^*}(\tau^* < \tau_\alpha(\boldsymbol{y})), \tag{19}$$

where τ_L is the low boundary of τ and the probability is based on the distribution of $\boldsymbol{y}$ given $\boldsymbol{\xi}^*$. We hope this coverage is close to α.

It seems that the coverage probability depends on the $\tau_\alpha(\boldsymbol{y})$, which is often hard to compute itself. Alternatively, we note that

$$\tau^* < \tau_\alpha(\boldsymbol{y}) \quad \text{if and only if } F(\tau^* \mid \boldsymbol{y}) < \alpha,$$

where $F(\tau \mid \boldsymbol{y})$ is the marginal cumulative posterior distribution of τ given $\boldsymbol{y}$. Then

$$P_{\boldsymbol{\xi}^*}(\tau^* < \tau_\alpha(\boldsymbol{y})) = P_{\boldsymbol{\xi}^*}(F(\tau^* \mid \boldsymbol{y}) < \alpha). \tag{20}$$

This formula shows that the frequentist coverage probabilities depend only on posterior cumulative distribution function $F(\tau^* \mid \boldsymbol{y})$ at the true values. Of course, it might depend on the entire parameters $\boldsymbol{\xi}^*$. Finding $F(\tau^* \mid \boldsymbol{y})$ requires only integration, and there is no need in finding the posterior quantiles in simulations. Another nice feature of the method is that once $F(\tau^* \mid \boldsymbol{y})$ is computed, it can be used to find the coverage probabilities for any α.

Theorem 2 *Assume that the prior* (15) *is used. If τ is a function of ρ, then the frequentist coverage probabilities in* (20) *depends only on ρ^* and is independent of $(\delta_1, \boldsymbol{\beta})$.*

Proof Let $\pi(\tau \mid \boldsymbol{y})$ be the posterior density of τ given $\boldsymbol{y}$. Note that

$$\pi(\tau \mid \boldsymbol{y}) = \frac{\int_{\tau=\tau(\rho)} L_{**a}(\rho)\pi(\rho)\,\mathrm{d}\rho}{\int_{\lambda_1^{-1}}^{\lambda_n^{-1}} L_{**a}(\rho)\pi(\rho)\,\mathrm{d}\rho} = \frac{\int_{\tau=\tau(\rho)} \frac{\pi(\rho)}{|\boldsymbol{\Sigma}_\rho|^{1/2}|\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|^{1/2}(S^2)^{(n-p)/2+a-1}}\,\mathrm{d}\rho}{\int_{\lambda_1^{-1}}^{\lambda_n^{-1}} \frac{\pi(\rho)}{|\boldsymbol{\Sigma}_\rho|^{1/2}|\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|^{1/2}(S^2)^{(n-p)/2+a-1}}\,\mathrm{d}\rho}. \tag{21}$$

Clearly, $\pi(\tau \mid \boldsymbol{y})$ depends on $\boldsymbol{y}$ only through S^2. Since $\boldsymbol{X}'\boldsymbol{R}_{\boldsymbol{\Sigma}} = \boldsymbol{0}$, we have

$$S^2 = \boldsymbol{y}\boldsymbol{R}_{\boldsymbol{\Sigma}}\boldsymbol{y} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*)'\boldsymbol{R}_{\boldsymbol{\Sigma}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*).$$

Note that $\tilde{\boldsymbol{\epsilon}} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*)/\sqrt{\delta_1^*} \sim N_n(\boldsymbol{0}, \boldsymbol{\Sigma}_\rho^*)$, where $\boldsymbol{\Sigma}_\rho^* = (\boldsymbol{I}_n - \rho^*\boldsymbol{C})^{-1}$. Then $S^2 = \delta_1^*\tilde{\boldsymbol{\epsilon}}'\boldsymbol{R}_{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\epsilon}}$ and

$$\pi(\tau \mid \boldsymbol{y}) = \frac{\int_{\tau=\tau(\rho)} \frac{\pi(\rho)}{|\boldsymbol{\Sigma}_\rho|^{1/2}|\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|^{1/2}(\tilde{\boldsymbol{\epsilon}}'\boldsymbol{R}_{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\epsilon}})^{(n-p)/2+a-1}}\,d\rho}{\int_{\lambda_1^{-1}}^{\lambda_n^{-1}} \frac{\pi(\rho)}{|\boldsymbol{\Sigma}_\rho|^{1/2}|\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|^{1/2}\left(\tilde{\boldsymbol{\epsilon}}'\boldsymbol{R}_{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\epsilon}}\right)^{(n-p)/2+a-1}}\,\mathrm{d}\rho}. \tag{22}$$

$\tilde{\boldsymbol{\epsilon}}$ depends only on ρ^*, so does the frequentist distribution of $\pi(\tau \mid \boldsymbol{y})$. □

Theorem 2 shows that the frequentist coverage probabilities of Bayesian credible intervals for many functions of parameters under a large class of priors will depend only on true parameter ρ. Therefore, in the simulation study we could fix $(\delta_1, \boldsymbol{\beta})$ at any value. For simplicity, we choose $(\delta_1^*, \boldsymbol{\beta}^*) = (1, \boldsymbol{0})$. Since one does not need considering choices of nuisance parameters, it can tremendously simplify and speed up computation.

We see that finding the marginal posterior cumulative distribution of η requires only an integration. In fact, define

$$g(\rho) = \frac{\pi(\rho)}{|\boldsymbol{\Sigma}_\rho|^{1/2}|\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|^{1/2}(S^2)^{(n-p)/2+a-1}}.$$

We have

$$F(\rho^* \mid \boldsymbol{y}) \equiv P(\rho < \rho^* \mid \boldsymbol{y}) = \frac{\int_{\lambda_1^{-1}}^{\rho^*} g(\rho)\,d\rho}{\int_{\lambda_1^{-1}}^{\lambda_n^{-1}} g(\rho)\rho}. \tag{23}$$

Thus, if, for example, we take a random sample of size m, $(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m)$, from the model (4) with the parameter $\boldsymbol{\xi}^* = (\rho^*, 1, \boldsymbol{0})$, then the frequentist coverage probability $P_{\boldsymbol{\xi}^*}(\rho^* < \rho_\alpha(\boldsymbol{y}))$ can be estimated by

$$\frac{\#\{\boldsymbol{y}_i, i = 1, \ldots, m : F(\rho^* \mid \boldsymbol{y}_i) < \alpha\}}{m}.$$

Table 1 Frequentist coverage of Bayesian equal-tailed 95 % credible intervals for ρ

	$p = 1$			$p = 6$		
	$\rho = 0.05$	$\rho = 0.12$	$\rho = 0.25$	$\rho = 0.05$	$\rho = 0.12$	$\rho = 0.25$
Reference (π_*^{R1})	0.960	0.957	0.981	0.976	0.957	0.976
Reference (π_*^{R2})	0.962	0.958	0.977	0.967	0.956	0.978
Independence Jeffreys	0.954	0.954	0.976	0.927	0.851	0.990
Jeffreys-rule	0.961	0.957	0.961	0.880	0.856	0.758

The method we develop only involves evaluating an integration and specifying the values of nuisance parameters. It does not require MCMC simulations in finding the posterior distributions.

3.2 Simulation study

This section presents the results of a small simulation experiment to investigate the frequentist coverage of equal-tailed credible intervals for one parameter of interest, ρ, by the above method, when one of four priors is used. These priors are the Jeffreys-rule, independence Jeffreys, and two "exact" reference priors.

Consider a setup similar to De Oliveira (2012). The models are defined on a 10×10 regular lattice with first order neighborhood system and $\boldsymbol{C}$ the adjacency matrix. Thus ρ must belong to the interval $(-0.260554, 0.260554)$. We consider two different mean functions $\mathbb{E}\{\boldsymbol{y}(\boldsymbol{s})\}$, namely the constant ($p = 1$) or $10 + s_{i1} + s_{i2} + s_{i1}s_{i2} + s_{i1}^2 + s_{i2}^2$ ($p = 6$), and three different values of ρ: 0.05, 0.12, or 0.25 (negative estimates of the spatial parameter are rare cases in practice, if they appear at all, so only positive values of ρ are considered). While De Oliveira (2012) considered different value of δ_1, from Theorem 2, the coverage for ρ does not depend on the choice of δ_1. One can find the simulated results in Tables 1 and 2 of De Oliveira (2012) are almost same for different choices of δ_1. Therefore, we choose $\delta_1 = 1$. 3,000 replications are generated for each choice of ρ and compute the equal-tailed 95 % credible intervals for ρ.

Table 1 shows frequentist coverage of equal-tailed credible intervals for ρ corresponding to four default priors, and large sample 95 % confidence intervals for ρ. When $p = 1$, the performance for two reference, Jeffreys-rule, independence Jeffreys priors is reasonable, and the coverage of confidence intervals are close to the nominal 0.95. When $p = 6$, the performance for two reference priors is reasonable and the coverage of confidence intervals are close to the nominal 0.95, while the coverage of the credible intervals based on Jeffreys-rule prior are below nominal. The coverage of the credible intervals based on independence Jeffreys are below nominal in some cases when ρ is small.

3.3 Real data analysis

In practice, we may have the following form of CAR models:

$$\boldsymbol{y} \sim N_n(\boldsymbol{X\beta}, \delta_1(\boldsymbol{I} - \rho\boldsymbol{C})^{-1}\boldsymbol{D}), \tag{24}$$

where $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_n)$ and $d_i, i = 1, \ldots, n$, are known positive values. With an appropriate transformation, (24) becomes (3). For example, if we make the transformation $\tilde{\boldsymbol{y}} = \boldsymbol{D}^{-1/2}\boldsymbol{y}$, then we will obtain that $\tilde{\boldsymbol{y}} \sim N_n(\tilde{\boldsymbol{X}}\boldsymbol{\beta}, \delta_1(\boldsymbol{I}_n - \rho\tilde{\boldsymbol{C}})^{-1})$, which is the form of (3). Here $\tilde{\boldsymbol{X}} = \boldsymbol{D}^{-1/2}\boldsymbol{X}$ and $\tilde{\boldsymbol{C}} = \boldsymbol{D}^{1/2}\boldsymbol{C}\boldsymbol{D}^{-1/2}$.

The data in our example are from the paper by Cressie and Chan (1989), where they used CAR models to analyze sudden infant death syndrome (SIDS). They considered CAR models for two sets of data: the number of SIDS from 1 July 1974 to 30 June 1978 and from 1 July 1979 to 30 June 1984. As an illustration, we only consider the 1974–1978 data here.

Let $\{S_i : i = 1, \ldots, n\}$ and $\{m_i : i = 1, \ldots, n\}$ denote the number of SIDS and the number of live births in the n counties of North Carolina, 1974–1987, respectively. In this example, $n = 99$. Originally, there are 100 counties, that is, we have 100 observations. Since the standardized residual of Anson County is unacceptably high, they deleted this county from the data analysis. They modeled the Freeman–Tukey (square-root) transformation

$$y_i = \left(\frac{1000 S_i}{m_i}\right)^{1/2} + \left(\frac{1000(S_i + 1)}{m_i}\right)^{1/2}. \tag{25}$$

Denote the location of the ith county seat by (u_i, v_i). Thus the distance for the ith and jth counties, denoted by d_{ij}, is defined by

$$d_{ij} = \{(u_i - u_j)^2 + (v_i - v_j)^2\}^{1/2}.$$

Define $\{N_i : i = 1, \ldots, n\}$ as the set of neighborhoods. $j \in N_i$ if $d_{ij} \le 30$ miles.

They considered the following model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{26}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$, many different forms of large-scale variation $\boldsymbol{X}\boldsymbol{\beta}$ and

$$\boldsymbol{e} \sim N(\boldsymbol{0}, \delta_1(\boldsymbol{I} - \rho\boldsymbol{C})^{-1}\boldsymbol{D}).$$

Here δ_1 is the variance, $\boldsymbol{D} = \text{diag}(m_1^{-1}, \ldots, m_n^{-1})$ and $\boldsymbol{C} = (c_{ij})$ where if $d_{ij} \le 30$ miles, $c_{ij} = C(k)d_{ij}^{-k}(m_j/m_i)^{1/2}$ and $c_{ij} = 0$ otherwise. Here k is specified to be 0, 1 or 2 according to how fast c_{ij} will decrease with distance d_{ij}. For comparability across different values of k, the constant of proportionality $C(k)$ is used and defined as $(\min\{d_{ij} : j \in N_i, i = 1, \ldots, n\})^k$.

Based on the appropriate statistics for model selection, they found the expected value $\boldsymbol{X}\boldsymbol{\beta}$ with the following form is preferred:

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_{5,1} \\ 1 & X_{5,2} \\ \cdots & \cdots \\ 1 & X_{5,n} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

Table 2 Summaries of the marginal posterior distributions (2.5 % quantile, median, 97.5 % quantile) for small-scale-variation parameters ($\hat{\rho}$, $\hat{\delta}_1$): 1974–1978. Data by the reference prior and MLE

	$k=0$	$k=1$	$k=2$
ρ			
Reference	0.021	0.118	0.011
CI	[−0.249, 0.173]	[−0.872, 0.855]	[−0.952, 0.945]
MLE	0.113	0.640	0.336
CI	[−0.01, 0.18]	[−0.24, 0.90]	[−0.98, 0.99]
δ_1			
Reference	1,232.7	1,238.3	1,252.2
MLE	1,138.7	1,169.2	1,214.6
β_1			
Reference	1.611	1.601	1.595
MLE	1.644	1.644	1.644
β_2			
Reference	0.036	0.036	0.036
MLE	0.035	0.035	0.035

where $X_{5,i}$ is the (Freeman–Tukey transformed) nonwhite live-birth rate, defined by:

$$X_{5,i} = \left(\frac{1000\bar{\omega}_i}{m_i}\right)^{1/2} + \left(\frac{1000(\bar{\omega}_i + 1)}{m_i}\right)^{1/2},$$

and $\bar{\omega}_i$ is the number of nonwhite live births in the ith county of North Carolina.

We make a transformation for the model (26), so the new model will be given in the following:

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{e}},$$

where $\tilde{\boldsymbol{y}} = \boldsymbol{D}^{-1/2}\boldsymbol{y}$, $\tilde{\boldsymbol{X}} = \boldsymbol{D}^{-1/2}\boldsymbol{X}$ and $\tilde{\boldsymbol{e}} = \boldsymbol{D}^{-1/2}\boldsymbol{e}$. Thus, $\tilde{\boldsymbol{e}} \sim N_n(\boldsymbol{0}, \delta_1(\boldsymbol{I} - \rho\tilde{\boldsymbol{C}})^{-1})$ where $\tilde{\boldsymbol{C}} = (\tilde{c}_{ij})$ with $\tilde{c}_{ij} = C(k)d_{ij}^k$ if $j \in N_i$ and $\tilde{c}_{ij} = 0$ otherwise.

In this example, a ratio-of-uniforms method, which can be found in Wakefield et al. (1991), for sampling ρ from its marginal posterior distribution is used for simulation. This method is quite efficient in computation. Because the expected value of parameters could not exist when the reference prior is applied, we report these estimates by quantiles in Table 2. Based on the simulation results, two reference priors' performance is almost same, so π^{R1} is used in this example. For comparison, we also present the estimate results by MLE in Cressie and Chan (1989).

We conclude that ρ is not significantly different from 0 based on the confidence intervals for three cases, which are same as the conclusions obtained by Cressie and Chan (1989). Our estimates for ρ are close to zero and much smaller than what they got by MLE. The confidence interval based on the reference prior is wider than MLE's when $k = 0$ and 1, but is narrower than MLE's when $k = 2$. Finally, the estimates for the error variance δ_1 and the coefficients $\boldsymbol{\beta}$ are close for both methods.

3.4 Comments

Based on the simulation results from this section, together with the results from De Oliveira (2012), we could summarize as follows:

(a) The frequentist properties of credible intervals computed using the independence Jeffreys and a reference priors are comparable.
(b) The computation of independence Jeffreys prior is simpler than the computation of reference priors. The latter guarantees posterior propriety in all cases, while this is not the case for the former.
(c) Neither the marginal independence Jeffreys prior nor the reference priors of the spatial parameter are integrable.

4 Appendix A: Proof of Proposition 1

We only prove the result for the ordering $\{\boldsymbol{\beta}, \rho, \delta_1\}$ in Part (a). For model (4), it is not difficult to obtain the Fisher information matrix as follows:

$$\boldsymbol{\Sigma}(\rho, \delta_1, \boldsymbol{\beta}) = \frac{1}{2}\begin{pmatrix} \boldsymbol{\Sigma}_2(\rho, \delta_1) & \boldsymbol{O} \\ \boldsymbol{O} & \frac{2}{\delta_1}\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X} \end{pmatrix},$$

where

$$\boldsymbol{\Sigma}_2(\rho, \delta_1) = \begin{pmatrix} \operatorname{tr}((\boldsymbol{I}_n - \rho\boldsymbol{C})^{-1}\boldsymbol{C})^2 & \frac{1}{\delta_1}\operatorname{tr}((\boldsymbol{I}_n - \rho\boldsymbol{C})^{-1}\boldsymbol{C}) \\ \frac{1}{\delta_1}\operatorname{tr}((\boldsymbol{I}_n - \rho\boldsymbol{C})^{-1}\boldsymbol{C}) & \frac{n}{\delta_1^2} \end{pmatrix}.$$

Let $[\rho_k^L, \rho_k^U]$, $[\delta_{1k}^L, \delta_{1k}^U]$, and $[\boldsymbol{\beta}_k^L, \boldsymbol{\beta}_k^U]$ be the compact sets so that as $k \to \infty$, $[\rho_k^L, \rho_k^U] \to (\lambda_1^{-1}, \lambda_n^{-1})$, $[\delta_{1k}^L, \delta_{1k}^U], \to (0, \infty)$, and $[\boldsymbol{\beta}_k^L, \boldsymbol{\beta}_k^U] \to \mathbb{R}^p$. In the following steps, we use the results in Lemma 2.1 in Datta and Ghosh (1996) and follow Berger and Bernardo's (1992) reference prior algorithm. First, we construct the conditional prior δ_1 given $(\rho, \boldsymbol{\beta})$ on $[\delta_{1k}^L, \delta_{1k}^U]$,

$$\pi_k(\delta_1 \mid \rho, \boldsymbol{\beta}) \propto \sqrt{\frac{n}{2\delta_1^2}} \propto \frac{1}{\delta_1}.$$

Next, we can construct the conditional prior for ρ given $\boldsymbol{\beta}$ on $[\rho_k^L, \rho_k^U]$,

$$\pi_k(\rho \mid \boldsymbol{\beta}) \propto \exp\left\{\frac{1}{2}\int_{\delta_{1k}^L}^{\delta_{1k}^U} \log\frac{\frac{1}{4}|\boldsymbol{\Sigma}_2(\rho, \delta_1)|}{\frac{n}{2\delta_1^2}}\pi_k(\delta_1 \mid \rho, \boldsymbol{\beta})\, d\delta_1\right\} \propto |\boldsymbol{\Sigma}_2(\rho, 1)|^{1/2},$$

where we use the fact $|\boldsymbol{\Sigma}_2(\rho, \delta_1)| = |\boldsymbol{\Sigma}_2(\rho, 1)|/\delta_1^2$.

Last, we can construct the prior of $\boldsymbol{\beta}$ within each compact set:

$$\pi_k(\boldsymbol{\beta}) \propto \exp\left\{\frac{1}{2}\int_{\rho_k^L}^{\rho_k^U}\int_{\delta_{1k}^L}^{\delta_{1k}^U} \log \frac{\frac{1}{4\delta_1^p}|\boldsymbol{\Sigma}_2(\rho,\delta_1)|\,|\boldsymbol{X}'\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{X}|}{\frac{1}{4}|\boldsymbol{\Sigma}_2(\rho,\delta_1)|}\right.$$
$$\left. \times\, \pi_k(\boldsymbol{\beta} \mid \rho, \delta_1)\pi_k(\delta_1 \mid \rho)\, d\rho\, \mathrm{d}\delta_1\right\} \propto 1.$$

Finally, for some interior point of $(\rho_0, \delta_{10}, \boldsymbol{\beta}_0)$ of $(\rho, \delta_1, \boldsymbol{\beta})$, the joint reference prior is:

$$\pi^R(\rho, \delta_1, \boldsymbol{\beta}) = \lim_{k\to\infty} \frac{\pi_k(\boldsymbol{\beta})\pi_k(\rho \mid \boldsymbol{\beta})\pi_k(\delta_1 \mid \rho, \boldsymbol{\beta})}{\pi_k(\boldsymbol{\beta}_0)\pi_k(\rho_0 \mid \boldsymbol{\beta}_0)\pi_k(\delta_{10} \mid \rho_0, \boldsymbol{\beta}_0)} \propto \frac{1}{\delta_1}|\boldsymbol{\Sigma}_2(\rho, 1)|^{1/2}.$$

5 Appendix B: Proof of Proposition 3

Here we only verify π_*^{R2} because

$$\mathrm{tr}(\boldsymbol{\Sigma}_\rho \boldsymbol{R}_\Sigma \boldsymbol{\Sigma}_\rho \boldsymbol{C})^2 - \frac{1}{n-p}\{\mathrm{tr}(\boldsymbol{\Sigma}_\rho \boldsymbol{R}_\Sigma \boldsymbol{\Sigma}_\rho \boldsymbol{C})\}^2 \le \mathrm{tr}(\boldsymbol{\Sigma}_\rho \boldsymbol{R}_\Sigma \boldsymbol{\Sigma}_\rho \boldsymbol{C})^2.$$

Let $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n)$, where $\boldsymbol{u}_i$ are the eigenvectors of $\boldsymbol{C}$. Denoting $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ by $\boldsymbol{\Lambda}$ where $\lambda_i, i = 1, \ldots, n$ are the eigenvalues of $\boldsymbol{C}$, we have $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$. Thus,

$$\boldsymbol{U}'\boldsymbol{R}_{\boldsymbol{\Sigma}}\boldsymbol{U} = (\boldsymbol{I}_n - \rho\boldsymbol{\Lambda}) - (\boldsymbol{I}_n - \rho\boldsymbol{\Lambda})\boldsymbol{U}'\boldsymbol{X}\{\boldsymbol{X}'\boldsymbol{U}(\boldsymbol{I}_n - \rho\boldsymbol{\Lambda})\boldsymbol{U}'\boldsymbol{X}\}^{-1}\boldsymbol{X}'\boldsymbol{U}(\boldsymbol{I}_n - \rho\boldsymbol{\Lambda}).$$

For $\boldsymbol{X}$, since $\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{R}$ (so-called QR decomposition) and $\boldsymbol{X}$ is a full-column rank matrix from the assumption, where $\boldsymbol{Q}$ is an $n \times p$ column orthonormal matrix and $\boldsymbol{R}$ is a $p \times p$ upper triangular matrix, so $\boldsymbol{R}$ is nonsingular. With some algebra, we have

$$\boldsymbol{U}'\boldsymbol{R}_{\boldsymbol{\Sigma}}\boldsymbol{U} = (\boldsymbol{I}_n - \rho\boldsymbol{\Lambda}) - (\boldsymbol{I}_n - \rho\boldsymbol{\Lambda})\boldsymbol{U}'\boldsymbol{Q}\{\boldsymbol{Q}'\boldsymbol{U}(\boldsymbol{I}_n - \rho\boldsymbol{\Lambda})\boldsymbol{U}'\boldsymbol{Q}\}^{-1}\boldsymbol{Q}'\boldsymbol{U}(\boldsymbol{I}_n - \rho\boldsymbol{\Lambda}).$$

If $\boldsymbol{u}_1 \in \mathcal{C}(\boldsymbol{X})$, that is, $\boldsymbol{u}_1 \in \mathcal{C}(\boldsymbol{Q})$, one can find $\boldsymbol{t}_2, \ldots, \boldsymbol{t}_p$ in $\mathcal{C}(\boldsymbol{Q})$ such that $\boldsymbol{u}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_p$ are orthonormal. Thus, denoting $(\boldsymbol{u}_1\, \boldsymbol{t}_2 \ldots \boldsymbol{t}_p)$ by $\boldsymbol{Q}^*$, one can find a nonsingular $p \times p$ matrix $\boldsymbol{T}$ such that $\boldsymbol{Q} = \boldsymbol{Q}^*\boldsymbol{T}$. With some algebra, one can obtain

$$\boldsymbol{Q}^{*\prime}\boldsymbol{U} = \begin{pmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{0} & \tilde{\boldsymbol{Q}} \end{pmatrix},$$

so if we denote $\boldsymbol{I}_{n-1} - \rho\mathrm{diag}(\lambda_2, \ldots, \lambda_n)$ by $\tilde{\boldsymbol{\Lambda}}_2$, we obtain

$$\boldsymbol{U}'\boldsymbol{R}_{\boldsymbol{\Sigma}}\boldsymbol{U} = \begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{R}_{\tilde{\boldsymbol{\Lambda}}_2} \end{pmatrix},$$

where

$$\boldsymbol{R}_{\tilde{\boldsymbol{\Lambda}}_2} = \tilde{\boldsymbol{\Lambda}}_2 - \tilde{\boldsymbol{\Lambda}}_2\tilde{\boldsymbol{Q}}(\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{\Lambda}}_2\tilde{\boldsymbol{Q}})^{-1}\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{\Lambda}}_2.$$

Thus, we obtain

$$\pi_*^{R2}(\rho) \propto \{\mathrm{tr}(\tilde{\boldsymbol{\Lambda}}_2^{-1}\boldsymbol{R}_{\tilde{\boldsymbol{\Lambda}}_2}\tilde{\boldsymbol{\Lambda}}_2^{-1}\boldsymbol{\Lambda}_2)^2\}^{1/2},$$

where $\boldsymbol{\Lambda}_2 = \mathrm{diag}(\lambda_2, \ldots, \lambda_n)$. Note that $\tilde{\boldsymbol{Q}}$ and $\boldsymbol{\Lambda}_2$ both do not depend on λ_1, so one can obtain the result for $\pi_*^{R2}(\rho)$ as $\rho \to \lambda_1^{-1}$.

Now, assume that $\boldsymbol{u}_1 \notin \mathcal{C}(\boldsymbol{X})$. Note that

$$\boldsymbol{C} = \frac{1}{\rho}(\boldsymbol{I}_n - \boldsymbol{\Sigma}_\rho^{-1}),$$

and $\boldsymbol{R}_\Sigma\boldsymbol{\Sigma}_\rho$ is an idempotent matrix. With some algebra, one can obtain

$$\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{R}_\Sigma\boldsymbol{\Sigma}_\rho\boldsymbol{C})^2 = \frac{1}{\rho^2}\{\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{R}_\Sigma\boldsymbol{\Sigma}_\rho)^2 - 2\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{R}_\Sigma\boldsymbol{\Sigma}_\rho) + \mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{R}_\Sigma)\}.$$

By idempotency, we have $\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{R}_\Sigma) \le n$. Therefore, $\pi_*^{R2}(\rho)$ is at most proportional to $\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{R}_\Sigma\boldsymbol{\Sigma}_\rho)^2$ as $\rho \to \lambda_1^{-1}$.

$\boldsymbol{R}_\Sigma \le \boldsymbol{\Sigma}_\rho^{-1}$ and $\boldsymbol{\Sigma}_\rho$ is positive definite, so we have

$$\mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{R}_\Sigma\boldsymbol{\Sigma}_\rho)^2 \le \mathrm{tr}(\boldsymbol{\Sigma}_\rho\boldsymbol{\Sigma}_\rho^{-1}\boldsymbol{\Sigma}_\rho)^2 = \mathrm{tr}(\boldsymbol{\Sigma}_\rho)^2.$$

Since $\mathrm{tr}(\boldsymbol{\Sigma}_\rho)^2 = \sum_{i=1}^n \frac{1}{(1-\rho\lambda_i)^2}$. $\mathrm{tr}(\boldsymbol{\Sigma}_\rho)^2$ is proportional to $1/(1-\rho\lambda_1)^2$ as $\rho \to \lambda_1^{-1}$. The result then follows.

Acknowledgments D. Sun's research was supported partially by the National Science Foundation grants SES-0720229, SES-1024080, and DMS-1007874. The authors would like to thank the editor, the associate editor, and anonymous referees for constructive comments to the earlier versions of the paper.

References

Bell, B. S., Broemeling, L. D. (2000). A Bayesian analysis of spatial processes with application to disease mapping. *Statistics in Medicine, 19*, 957–974.

Berger, J. O., Bernardo, J. M. (1992). On the development of reference priors (Disc: p49–60). In *Bayesian statistics 4: Proceedings of the fourth Valencia international meeting* (pp. 35–49).

Berger, J. O., Philippe, A., Robert, C. (1998). Estimation of quadratic functions: Noninformative priors for non-centrality parameters. *Statistica Sinica, 8*, 359–376.

Berger, J. O., De Oliveira, V., Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association, 96*, 1361–1374.

Besag, J. G. P. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological, 36*, 3–66.

Clayton, D., Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics, 43*, 671–681.

Cressie, N. A. C. (1993). *Statistics for spatial data* (rev ed.). New York: Wiley.

Cressie, N. A. C., Chan, N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association, 84*, 393–401.

Cressie, N., Perrin, O., Thomas-Agnan, C. (2005). Likelihood-based estimation for Gaussian mrfs. *Statistical Methodology, 2*, 1–16.

Datta, G. S., Ghosh, M. (1996). On the invariance of noninformative priors. *The Annals of Statistics, 24*, 141–159.

De Oliveira, V. (2012). Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics, 64*, 107–133.

Harville, D. A. (1974). Bayeisan inference for variance components using only error contrasts. *Biometrika, 61*, 383–385.

Paulo, R. (2005). Default priors for Gaussian processes. *The Annals of Statistics, 33*, 556–582.

Ren, C., Sun, D., He, Z. (2012). Objective Bayesian analysis for a spatial model with nugget effects. *Journal of Statistical Planning and Inference, 142*, 1933–1946.

Richardson, S., Guihenneuc, C., Lasserre, V. (1992). Spatial linear models with autocorrelated error structure. *The Statistician, 41*, 539–557.

Rue, H., Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Boca Raton: Chapman and Hall/CRC.

Sun, D., Ye, K. (1995). Reference prior Bayesian analysis for normal mean products. *Journal of the American Statistical Association, 90*, 589–597.

Wakefield, J. C., Gelfand, A. E., Smith, A. F. M. (1991). Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing, 1*, 129–133.

Welch, B., Peers, H. W. (1963). Reference prior Bayesian analysis for normal mean products. *Journal of the Royal Statistical Society B, 25*, 318–329.

Ye, K. (1994). Bayesian reference prior analysis on the ratio of variances for the balanced one-way random effect model. *Journal of Statistical Planning and Inference, 41*, 267–280.