THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

Date August 16, 2009

WP # 0110MSS-496-2009

Maximum Lilkelihood and Restricted Maximum Likelihood Estimation for a Class of Gaussian Markov Random Fields

Victor De Oliveira
Department of Management Science and Statistics
University of Texas at San Antonio

ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631
210 458-4317 | BUSINESS.UTSA.EDU

Maximum Likelihood and Restricted Maximum Likelihood Estimation for a Class of Gaussian Markov Random Fields

Victor De Oliveira[1] and Marco A. R. Ferreira[2]

[1] *Department of Management Science and Statistics, The University of Texas at San Antonio San Antonio, TX 78249, USA*

[2] *Department of Statistics, University of Missouri–Columbia Columbia, MO 65211, USA*

February 4, 2010

Summary.

This work describes a Gaussian Markov random field model that includes several previously proposed models, and studies properties of their maximum likelihood (ML) and restricted maximum likelihood (REML) estimators in a special case. Specifically, for models where a particular relation holds between the regression and precision matrices of the model, we provide sufficient conditions for existence and uniqueness of ML and REML estimators of the covariance parameters, and provide a straightforward way to compute them. It is found that the ML estimator always exists while the REML estimator may not exist with positive probability. A numerical comparison suggests that for this model ML estimators of covariance parameters have, overall, better frequentist properties than REML estimators.

Keywords: Eigenvalues and eigenvectors; Profile likelihood; Restricted likelihood; Spatial data.
JEL Classifications: C11, C31

1. Introduction

Gaussian Markov random fields (GMRF) are important families of distributions for the modeling of spatial data, which have been extensively used in different areas of spatial statistics such as remote sensing (Chellappa and Jain, 1993), disease mapping (Cressie and Chan, 1989) and image analysis (Besag, York and Mollié, 1991). The practical use of GMRF for modeling large scale spatial phenomena has significantly increased after recent advances on the efficient simulation of GMRFs (Rue, 2001; Rue and Follestad, 2002); see Cressie (1993) and Rue and Held (2005) for detailed accounts on GMRFs.

Two of the most commonly used approaches for parameter estimation in GMRFs have been maximum likelihood (ML) and restricted maximum likelihood (REML); the former is described in Mardia and Marshall (1984), Cressie and Chan (1989) and Richardson, Guihenneuc and Lasserre (1992), while the latter is described in Zimmerman and Harville (1991) and Cressie and Lahiri (1996). Most of the known results about the behavior of ML and REML estimators are asymptotic in nature, and little is known about their behavior in small samples. In particular, little is known about conditions that guarantee existence and uniqueness of ML and REML estimators of GMRF parameters based on finite samples.

This work describes a GMRF model that includes several previously proposed models, and studies in detail properties of its ML and REML estimators in a special case. Specifically, for models where a particular relation holds between the regression and precision matrices of the model, we provide sufficient conditions for existence and uniqueness of ML and REML estimators of the covariance parameters, and provide a straightforward way to compute them. It is found that the ML estimators of covariance parameters always exist, but the REML estimators of covariance parameters do not exist with positive probability. In addition, the existence and positivity of the ML and REML estimators of covariance parameters depend on easy-to-check conditions involving a ratio of quadratic forms and some eigenvalue summaries of a matrix that determines the precision matrix of the model. Demidenko and Massam (1999) and Birkes and Wulff (2003) provided general results on existence of ML and REML estimators of covariance parameters for a large class of Gaussian variance components models. Although the GMRF model we consider here can in principle be framed as an instance of the class of models studied by Birkes and Wulff (2003), doing so is cumbersome due to the resulting awkward parameter space and the difficulty of checking the conditions of their results. We discuss this connection in Section 6.

We also study some (small sample) frequentist properties of ML and REML estimators of the covariance parameters, as well as how these properties depend on the strength of spatial association. Based on the above theoretical and empirical results it is found for this model that, overall, the ML estimators of covariance parameters have better frequentist properties than the REML estimators.

2. The Model

Consider a collection of sites or regions indexed by the integers $1, 2, \ldots, n$, forming a lattice (regular or irregular) within a geographical domain of interest. This lattice is assumed to be endowed with a neighborhood system, $\{N_k : k = 1, \ldots, n\}$, where N_k denotes the collection of sites that are neighbors of site k. This neighborhood system satisfies that for any $k, l = 1, \ldots, n$, $k \in N_l$ if and only if $l \in N_k$ and $k \notin N_k$.

For each site, k, it is observed the variable of interest, Y_k, and a set of p explanatory variables, $\mathbf{x}_k = (x_{k1}, \ldots, x_{kp})'$. The random vector of observed responses, $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, would be modeled by the joint distribution

$$\mathbf{Y} \sim \mathrm{N}_n(X\boldsymbol{\beta}, \sigma^2\Sigma(\phi)) \qquad \text{with} \qquad \Sigma(\phi)^{-1} = I_n + \phi H, \tag{1}$$

where $X = (\mathbf{x}_1 \cdots \mathbf{x}_n)'$ is a known $n \times p$ design matrix of rank p, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$ are unknown regression parameters, $\sigma > 0$ is a scale parameter and $\phi \geq 0$ is a 'spatial' parameter, I_n is the $n \times n$ identity matrix and H is given by

$$(H)_{kl} = \begin{cases} h_k & \text{if } k = l \\ -g_{kl} & \text{if } k \in N_l \\ 0 & \text{otherwise.} \end{cases}$$

The weight $g_{kl} > 0$ is a 'measure of similarity' between sites k and l, $g_{kl} = g_{lk}$, and $h_k = \sum_{l \in N_k} g_{kl}$. For every $\phi \geq 0$, $\Sigma(\phi)^{-1}$ is diagonally dominant which together with the fact that all its diagonal elements are positive imply that the matrix is positive definite (Harville, 1997 p. 279-280). The random vector $\mathbf{Y}$ has then a probability density function. The matrix H, assumed known, allows the modeling of different patterns of spatial association by the specification of different neighborhood systems and weights $\{g_{kl}\}$. We denote the model parameters by $\boldsymbol{\eta} = (\boldsymbol{\beta}', \sigma^2, \phi) \in \mathbb{R}^p \times (0, \infty) \times [0, \infty)$.

The parameter ϕ controls the strength of association between the components of $\mathbf{Y}$ and determines the main properties of model (1). When $\phi = 0$, the components of $\mathbf{Y}$ become independent random variables with $Y_k \sim \mathrm{N}(\mathbf{x}_k'\boldsymbol{\beta}, \sigma^2)$, while when $\phi \to \infty$ model (1) approaches the intrinsic autoregressive model (Besag et al. 1991; Besag and Kooperberg, 1995), which is an improper distribution that has been extensively used in spatial statistics to model latent processes and spatial random effects. In addition, ϕ also controls conditional correlations among neighboring sites. If ρ_{kl}^c denotes the conditional correlation of Y_k and Y_l given $\{Y_j, j \neq k, l\}$, with $k \neq l$, then for the model (1) we have $\rho_{kl}^c = \phi g_{kl}/\big((1+\phi h_k)(1+\phi h_l)\big)^{0.5}$ when $k \in N_l$, and 0 otherwise.

An equivalent specification of model (1) can be stated in terms of its full conditional distributions, which are given by

$$(Y_k \mid Y_l, l \neq k) \sim \mathrm{N}\left(\mathbf{x}_k'\boldsymbol{\beta} + \frac{\phi}{1+\phi h_k}\sum_{l \in N_k} g_{kl}(Y_l - \mathbf{x}_l'\boldsymbol{\beta}) \ , \ \frac{\sigma^2}{1+\phi h_k}\right), \qquad k = 1, \ldots, n.$$

Then, the full conditional mean of Y_k is equal to the sum of its marginal mean $\mathbf{x}_k'\boldsymbol{\beta}$ and a correction term depending on the deviations of the observed neighboring values from their respective marginal means. Also, the full conditional variance of Y_k decreases with the number of neighbors. These are natural properties for many kinds of spatial data.

For the case when no explanatory variables are available, several GMRFs proposed in the literature can be shown to be reparameterizations of model (1), for instance those in Leroux, Lei and Breslow (1999), Sun, Tsutakawa and Speckman (1999) [their model 1A] and Dryden, Ippoliti and Romagnoli (2002); see Ferreira and De Oliveira (2007) for details. Pettitt, Weir and Hart (2002) studied a model slightly more general than the one considered here as it allows both positive and negative association among the Y_ks. On the other hand, the model considered in Clayton and Kaldor (1987) and Cressie and Chang (1989) has precision matrix that is not necessarily diagonally dominant, so it is not a special case of model (1).

The matrix H is symmetric, non-negative definite and satisfies $H\mathbf{1}_n = \mathbf{0}_n$. From its spectral decomposition (Harville, 1997 p. 537), $H = TDT'$ where $T = (\mathbf{t}_1 \cdots \mathbf{t}_n)$ has orthonormal columns given by the normalized eigenvectors of H, that is $\mathbf{t}_i'\mathbf{t}_j = \delta_{ij}$, and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{n-1} > \lambda_n = 0$ are the ordered eigenvalues of H. Then

$$|I_n + \phi H| = \prod_{k=1}^{n-1}(1 + \phi\lambda_k),$$

and the likelihood function of the parameters $\boldsymbol{\eta}$ based on the observed data $\mathbf{y}$ is given by

$$L(\boldsymbol{\eta}; \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \prod_{k=1}^{n-1}(1+\phi\lambda_k)^{\frac{1}{2}} \ \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\beta})'\Sigma(\phi)^{-1}(\mathbf{y} - X\boldsymbol{\beta})\right\}.$$

Throughout the article we make the following two assumptions which hold in most practical applications:

(A1) The matrix X has rank $p \leq n-1$ and $\mathbf{1}_n$ belongs to the subspace generated by the columns of X (e.g. $\mathbf{x}_1 = \mathbf{1}_n$) .

(A2) The matrix H has rank $n-1$. A necessary and sufficient condition for this to hold is that 0 is an eigenvalue of H with multiplicity 1 (Harville, 1997 Lemma 21.1.1).

The first assumption states that the model has an overall mean. The second assumption may not hold for matrices H that are reducible (which occurs when the neighborhood system is disconnected), but this rarely occurs in practice.

3. Maximum Likelihood

The ML estimator of $\boldsymbol{\eta}$, provided it exists, is given by $\hat{\boldsymbol{\eta}} = \arg\max\ \log\big(L(\boldsymbol{\eta}; \mathbf{y})\big)$. For any fixed $\phi \geq 0$ the ML estimators of $\boldsymbol{\beta}$ and σ^2 are given, respectively, by

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\phi) &= (X'\Sigma(\phi)^{-1}X)^{-1}X'\Sigma(\phi)^{-1}\mathbf{y},\\ \hat{\sigma}^2(\phi) &= \frac{1}{n}S^2(\phi),\end{aligned}$$

where

$$S^2(\phi) = \mathbf{y}'(I_n - A(\phi))'\Sigma(\phi)^{-1}(I_n - A(\phi))\mathbf{y},$$

and

$$A(\phi) = X(X'\Sigma(\phi)^{-1}X)^{-1}X'\Sigma(\phi)^{-1}.$$

Also, the profile log–likelihood of ϕ is given, up to an additive constant, by

$$\begin{aligned}l_p(\phi; \mathbf{y}) &= \log\big(L(\hat{\boldsymbol{\beta}}(\phi), \hat{\sigma}^2(\phi), \phi; \mathbf{y})\big)\\ &= \frac{1}{2}\left(\sum_{k=1}^{n-1}\log(1+\phi\lambda_k) - n\log(S^2(\phi)) - n\log(2\pi)\right).\end{aligned}$$

It then holds that the ML estimator of $\boldsymbol{\eta}$ exists and is unique if and only if $l_p(\phi; \mathbf{y})$ has a unique maximum in $[0, \infty)$, say $\hat{\phi}$, in which case the ML estimator of $\boldsymbol{\eta}$ is given by $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\beta}}(\hat{\phi}), \hat{\sigma}^2(\hat{\phi}), \hat{\phi})$. Since $l_p(\phi; \mathbf{y})$ is continuously differentiable with respect to ϕ, when $\hat{\phi}$ exists it must hold that $\hat{\phi}$ is either zero or a root of the profile likelihood equation (recall that $\phi \in [0, \infty)$)

$$\begin{aligned}0 &= \frac{\partial}{\partial\phi}l_p(\phi; \mathbf{y})\\ &= \frac{1}{2}\left(\sum_{k=1}^{n-1}\Big(\frac{\lambda_k}{1+\phi\lambda_k}\Big) - n\frac{\frac{\partial}{\partial\phi}S^2(\phi)}{S^2(\phi)}\right)\\ &= \frac{1}{2}\left(\sum_{k=1}^{n-1}\Big(\frac{\lambda_k}{1+\phi\lambda_k}\Big) - n\frac{\mathbf{y}'(I_n - A(\phi))'H(I_n - A(\phi))\mathbf{y}}{S^2(\phi)}\right)\\ &= \frac{1}{2}\left(\sum_{k=1}^{n-1}\Big(\frac{\lambda_k}{1+\phi\lambda_k}\Big) - n\Big(\frac{Q(\phi; \mathbf{y})}{1+\phi Q(\phi; \mathbf{y})}\Big)\right), \qquad (2)\end{aligned}$$

where

$$Q(\phi; \mathbf{y}) = \frac{\mathbf{y}'(I_n - A(\phi))'H(I_n - A(\phi))\mathbf{y}}{\mathbf{y}'(I_n - A(\phi))'(I_n - A(\phi))\mathbf{y}};$$

the justification for the third identity in (2) is given by Lemma 6.1 in the Appendix. The solution (in ϕ) of equation (2) is usually found by iterative numerical methods. But before this is attempted it is convenient to be able to analytically determine whether (2) has a unique solution and whether that solution maximizes $l_p(\phi; \mathbf{y})$, which in general are difficult tasks. In what follows we consider a special case for which such tasks can be undertaken.

3.1. A Special Case

We first introduce some notation. Let $C(X)$ denote the p-dimensional subspace of $\mathbb{R}^n$ spanned by the columns of X, $A = A(0) = X(X'X)^{-1}X'$ is the orthogonal projection matrix onto $C(X)$, and for $\mathcal{M}$ a subspace of $\mathbb{R}^n$ and V an $n \times n$ symmetric non-negative definite matrix, $V(\mathcal{M}) = \{V\mathbf{x} \,:\, \mathbf{x} \in \mathcal{M}\}$. In this section we study the problem of determination and computation of the ML estimator of $\boldsymbol{\eta}$ in models for which $A(\phi)$ does not depend on ϕ. There are several equivalent conditions, first given by Zyskind (1967, Theorem 2), that allow such simplification. Some of these are summarized in the following result.

LEMMA 3.1. *Consider the GMRF model in (1), and let $V = \Sigma(\phi)^{-1}$. Then the following statements are equivalent:*

(i) $C(X)$ coincides with the subspace spanned by some p eigenvectors of V.

(ii) $V(C(X)) \subseteq C(X)$.

(iii) $H(C(X)) \subseteq C(X)$.

In addition if any of the above conditions hold, then $A(\phi) = A$ for every $\phi \geq 0$.

Conditions (i) and (ii) in Lemma 3.1 (and many other equivalent conditions) have a long history in the statistical literature of linear models for guaranteeing that ordinary least squares and best linear unbiased estimators of regression parameters agree; see e.g. Zyskind (1967) and Puntanen and Styan (1989). These conditions hold for model (1) when the mean response is constant (i.e. $X = \mathbf{1}_n$) since $H\mathbf{1}_n = \mathbf{0}_n$. Other situations are considered in Section 5. Then for models in which $H(C(X)) \subseteq C(X)$ we have

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y} \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{n}S^2(\hat{\phi}),$$

where $\hat{\phi}$ is the ML estimator of ϕ. In this case it holds that $A(\phi)$ does not depend on ϕ, so $Q(\phi;\mathbf{y})$ does not depend on ϕ either and is equal to $Q(0;\mathbf{y}) = Q(\mathbf{y})$ with

$$Q(\mathbf{y}) = \frac{\mathbf{y}'(I_n - A)H(I_n - A)\mathbf{y}}{\mathbf{y}'(I_n - A)\mathbf{y}},$$

as $I_n - A$ is symmetric and idempotent. In all that follows recall that $\lambda_n = 0$ and note that assumption (A2) implies that $C(H) \not\subset C(X)$, which in turn guarantees that $Q(\mathbf{y}) > 0$ with probability one.

THEOREM 3.1. *Consider the GMRF model in (1) where it holds that $H(C(X)) \subseteq C(X)$, and let $\bar{\lambda} = \frac{1}{n}\sum_{k=1}^{n}\lambda_k$. Then:*

(i) If $Q(\mathbf{y}) \geq \bar{\lambda}$, then $\hat{\phi} = 0$.

(ii) If $Q(\mathbf{y}) < \bar{\lambda}$, then $\hat{\phi} > 0$ is the unique solution to the equation

$$\sum_{k=1}^{n-1}\left(\frac{\lambda_k}{1+\phi\lambda_k}\right) - n\left(\frac{Q(\mathbf{y})}{1+\phi Q(\mathbf{y})}\right) = 0. \tag{3}$$

Proof. As stated before, the ML estimator of $\boldsymbol{\eta}$ exists and is unique if and only if $l_p(\phi;\mathbf{y})$ has a unique maximum in $[0,\infty)$. Also note from Lemma 3.1 that $\frac{\partial}{\partial\phi} l_p(\phi;\mathbf{y})$ is given by (2) with $Q(\phi;\mathbf{y})$ replaced by $Q(\mathbf{y})$.

(i) It follows from (2) that for any $\phi > 0$

$$\begin{aligned} \frac{\partial}{\partial\phi} l_p(\phi;\mathbf{y}) &\leq \frac{n}{2}\left(\frac{1}{n}\sum_{k=1}^{n}\Big(\frac{\lambda_k}{1+\phi\lambda_k}\Big) - \Big(\frac{\bar\lambda}{1+\phi\bar\lambda}\Big)\right) \\ &< 0, \end{aligned}$$

where the last relation follows from Jensen's inequality applied to the uniform distribution supported at $\{\lambda_k : k = 1,\dots,n\}$ and the strictly concave function $x(1+\phi x)^{-1}$, $x > 0$ (ϕ fixed). Since $l_p(\phi;\mathbf{y})$ is continuous at $\phi = 0$, the profile log–likelihood of ϕ is strictly decreasing on $[0,\infty)$ so the ML estimator of ϕ is $\hat\phi = 0$.

(ii) Note that the equation $\frac{\partial}{\partial\phi} l_p(\phi;\mathbf{y}) = 0$ is equivalent to the equation $m(\phi,\boldsymbol{\lambda}) = Q(\mathbf{y})$, where $m(\phi,\boldsymbol{\lambda})$ is the function defined in equation (6) in the Appendix. Using parts (i), (ii) and (iv) of Lemma 6.2 with $\mathbf{v} = \boldsymbol{\lambda}$ (recall that $\lambda_n = 0$), we have that $0 < Q(\mathbf{y}) < \bar\lambda$ implies that $\frac{\partial}{\partial\phi} l_p(\phi;\mathbf{y}) = 0$ has a unique solution on $(0,\infty)$, which is the unique ML estimator of ϕ. □

4. Restricted Maximum Likelihood

It has been found for a variety of Gaussian models that the method of maximum likelihood usually produces biased estimators for variance and covariance parameters, and this bias may be substantial in situations involving small samples or models where the mean response is not constant. In these situations the method of restricted maximum likelihood has been advocated by many as a better estimation approach, at least in regard to producing less biased variance and covariance estimators. The method consists of maximizing the likelihood function of the variance and covariance parameters based on a set of $n-p$ linearly independent 'error contrasts', which has a joint Gaussian distribution with null mean vector (so it does not depend on $\boldsymbol{\beta}$). This is called a restricted likelihood function. As first noted by Harville (1974), log–restricted likelihoods of the variance and covariance parameters based on any two sets of $n-p$ linearly independent error contrasts differ only by an additive constant, and for model (1) are equal to

$$l^r(\sigma^2,\phi;\mathbf{y}) = -\frac{1}{2}\left((n-p)\log(\sigma^2) - \log(|\Sigma(\phi)^{-1}|) + \log(|X'\Sigma(\phi)^{-1}X|) + \frac{1}{\sigma^2}S^2(\phi) + (n-p)\log(2\pi)\right).$$

The restricted maximum likelihood (REML) estimator of (σ^2,ϕ), provided it exists, is given by $(\tilde\sigma^2,\tilde\phi) = \arg\max\, l^r(\sigma^2,\phi;\mathbf{y})$. For any fixed $\phi \geq 0$, the REML estimator of σ^2 is given by

$$\tilde\sigma^2(\phi) = \frac{1}{n-p}S^2(\phi).$$

We again consider in detail the special case when $H(C(X)) \subseteq C(X)$, which from Lemma 3.1 amounts to assume that $C(X)$ coincides with the subspace of $\mathbb{R}^n$ spanned by some p eigenvectors of H. Let $\mathbf{t}_{i_1},\dots,\mathbf{t}_{i_p}$ be these eigenvectors, which are orthonormal with corresponding eigenvalues $\lambda_{i_1},\dots,\lambda_{i_p}$, and $W = (\mathbf{t}_{i_1}\cdots\mathbf{t}_{i_p})$. We then have that

$$W'W = I_p \quad , \quad HW = WD_w \quad \text{and} \quad X = WB, \tag{4}$$

where $D_w = \mathrm{diag}(\lambda_{i_1}, \dots, \lambda_{i_p})$ and B is some $p \times p$ non-singular matrix. From (4) follows that

$$\begin{aligned} X'\Sigma(\phi)^{-1}X &= B'W'(I_n + \phi H)WB \\ &= B'(I_p + \phi D_w)B, \end{aligned}$$

so the profile log–restricted likelihood of ϕ is given, up to an additive constant, by

$$\begin{aligned} l_p^r(\phi; \mathbf{y}) &= l^r(\tilde{\sigma}^2(\phi), \phi; \mathbf{y}) \\ &= \frac{1}{2}\left(\sum_{k=1}^{n-1} \log(1+\phi\lambda_k) - \log(|X'\Sigma(\phi)^{-1}X|) - (n-p)\log(S^2(\phi)) - (n-p)\log(2\pi)\right) \\ &= \frac{1}{2}\left(\sum_{k\in J} \log(1+\phi\lambda_k) - (n-p)\log(S^2(\phi)) - (n-p)\log(2\pi)\right), \end{aligned}$$

where $J = \{1, \dots, n\} - \{i_1, \dots, i_p\}$. Note that because of assumption (A1), $n \notin J$. It then holds that the REML estimator of (σ^2, ϕ) exists and is unique if and only if $l_p^r(\phi; \mathbf{y})$ has a unique maximum on $[0, \infty)$, say $\tilde{\phi}$, in which case the REML estimator of (σ^2, ϕ) is given by $(\tilde{\sigma}^2(\tilde{\phi}), \tilde{\phi})$. By convention the REML estimator of $\boldsymbol{\beta}$ is $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\tilde{\phi})$, so in this case we have $\tilde{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$. From a similar argument and calculation as in (2) we have that $\tilde{\phi}$, provided it exists, must be either zero or a root of the profile restricted likelihood equation

$$\begin{aligned} 0 &= \frac{\partial}{\partial\phi} l_p^r(\phi; \mathbf{y}) \\ &= \frac{1}{2}\left(\sum_{k\in J}\Big(\frac{\lambda_k}{1+\phi\lambda_k}\Big) - (n-p)\Big(\frac{Q(\mathbf{y})}{1+\phi Q(\mathbf{y})}\Big)\right). \end{aligned} \tag{5}$$

In what follows we assume that $\{\lambda_k\}_{k\in J}$ are not all equal.

Theorem 4.1. *Consider the GMRF model (1) where it holds that $H(C(X)) \subseteq C(X)$, and let $\bar{\lambda}_J = \frac{1}{n-p}\sum_{k\in J}\lambda_k$ and $\bar{\zeta}_J = \big(\frac{1}{n-p}\sum_{k\in J}\lambda_k^{-1}\big)^{-1}$ [the arithmetic and harmonic means of $\{\lambda_k\}_{k\in J}$]. Then:*

(i) If $Q(\mathbf{y}) \le \bar{\zeta}_J$, then $\tilde{\phi}$ does not exist.

(ii) If $Q(\mathbf{y}) \ge \bar{\lambda}_J$, then $\tilde{\phi} = 0$.

(iii) If $\bar{\zeta}_J < Q(\mathbf{y}) < \bar{\lambda}_J$, then $\tilde{\phi} > 0$ is the unique solution to equation (5).

Proof. As stated before, the REML estimate of (σ^2, ϕ) exists and is unique if and only if $l_p^r(\phi; \mathbf{y})$ has a unique maximum on $[0, \infty)$.

(i) From (5) we have that for any $\phi \ge 0$

$$\begin{aligned} \frac{\partial}{\partial\phi} l_p^r(\phi; \mathbf{y}) &\ge \frac{1}{2}\left(\sum_{k\in J}\Big(\frac{\lambda_k}{1+\phi\lambda_k}\Big) - (n-p)\frac{\bar{\zeta}_J}{1+\phi\bar{\zeta}_J}\right) \\ &= \frac{n-p}{2}\left(\frac{1}{n-p}\sum_{k\in J}\Big(\phi + \frac{1}{\lambda_k}\Big)^{-1} - \Big(\phi + \frac{1}{n-p}\sum_{k\in J}\lambda_k^{-1}\Big)^{-1}\right) \\ &> 0, \end{aligned}$$

where the last relation follows from Jensen's inequality applied to the uniform distribution supported at $\{\lambda_k^{-1} : k \in J\}$ and the strictly convex function $(\phi + x)^{-1}, x \geq 0$ (ϕ fixed). Hence $l_p^r(\phi; \mathbf{y})$ is strictly increasing on $[0, \infty)$, so $\tilde{\phi}$ does not exist.

(ii) From (5) we have that for any $\phi > 0$

$$\begin{aligned} \frac{\partial}{\partial \phi} l_p^r(\phi; \mathbf{y}) &\leq \frac{n-p}{2} \left(\frac{1}{n-p} \sum_{k \in J} \left(\frac{\lambda_k}{1 + \phi \lambda_k} \right) - \frac{\bar{\lambda}_J}{1 + \phi \bar{\lambda}_J} \right) \\ &< 0, \end{aligned}$$

where the last relation follows from Jensen's inequality applied to the uniform distribution supported at $\{\lambda_k : k \in J\}$ and the strictly concave function $x(1 + \phi x)^{-1}$, $x > 0$ (ϕ fixed). Since $l_p^r(\phi; \mathbf{y})$ is continuous at $\phi = 0$, the profile log–restricted likelihood of ϕ is strictly decreasing on $[0, \infty)$ and $\tilde{\phi} = 0$.

(iii) Note that the equation $\frac{\partial}{\partial \phi} l_p^r(\phi; \mathbf{y}) = 0$ is equivalent to the equation $m(\phi, \boldsymbol{\lambda}_J) = Q(\mathbf{y})$, where $m(\phi, \boldsymbol{\lambda}_J)$ is the function defined in equation (6) in the Appendix. Using parts (i), (ii) and (iii) of Lemma 6.2 with $\mathbf{v} = \boldsymbol{\lambda}_J$ (recall that $n \notin J$), we have that $\bar{\zeta}_J < Q(\mathbf{y}) < \bar{\lambda}_J$ implies that $\frac{\partial}{\partial \phi} l_p^r(\phi; \mathbf{y}) = 0$ has a unique solution on $(0, \infty)$, which is the unique REML estimator of ϕ. □

Once model (1) has been specified with particular matrices X and H, the application of the above results requires to determine whether or not $H(C(X)) \subseteq C(X)$ holds, and in the case it holds, to determine the set of indexes J. To that purpose, recall that the columns of $T = (\mathbf{t}_1 \cdots \mathbf{t}_n)$ are eigenvectors of H forming an orthonormal basis of $\mathbb{R}^n$. Hence there is an $n \times p$ matrix F such that $X = TF$, and $F = T'X$ since T is orthogonal. If $K = \{k_1, \ldots, k_q\}$ are the indexes of the rows of F that have at least one nonzero entry, and q ($\geq p$) is its cardinality, then the columns of X belong to the subspace of $\mathbb{R}^n$ spanned by $\{\mathbf{t}_{k_1}, \ldots, \mathbf{t}_{k_q}\}$. From this follows that $H(C(X)) \subseteq C(X)$ holds if and only if $q = p$, in which case $J = \{1, \ldots, n\} - K$. Alternative, $H(C(X)) \subseteq C(X)$ holds if and only if $(I_n - A)HX = O_{n \times p}$. Finally, it is worth pointing out that in the case the data arise as the result of a designed experiment where the matrix X is chosen by the researcher, it may be possible to choose X purposely in a way that $H(C(X)) \subseteq C(X)$ holds. This is briefly discussed in Section 6.

5. Comparison

In this section we compare the behavior in small samples of ML and REML estimators of the covariance parameters of model (1) for the case when $H(C(X)) \subseteq C(X)$. Recall that in this case $X = WB$, where $W = (\mathbf{t}_{i_1} \cdots \mathbf{t}_{i_p})$ with columns being eigenvectors of H corresponding to the eigenvalues $\lambda_{i_1}, \ldots, \lambda_{i_p}$ and B is $p \times p$ non-singular. Also recall the eigenvalue summaries

$$\bar{\lambda} = \frac{1}{n} \sum_{k=1}^{n} \lambda_k \quad , \quad \bar{\lambda}_J = \frac{1}{n-p} \sum_{k \in J} \lambda_k \quad , \quad \bar{\zeta}_J = \left(\frac{1}{n-p} \sum_{k \in J} \frac{1}{\lambda_k} \right)^{-1},$$

where $J = \{1, \ldots, n\} - \{i_1, \ldots, i_p\}$; note that $\bar{\zeta}_J < \bar{\lambda}_J$. We point out some immediate consequences of Theorems 3.1 and 4.1. First, the ML estimators of covariance parameters always exist while the REML estimators do not exist with positive probability. We investigate how this probability depends on some features of the model. Second, for models where $X = \mathbf{1}_n$ (so $p = 1$ and $J = \{1, \ldots, n-1\}$), we have that $\bar{\lambda} < \bar{\lambda}_J$ and hence it holds that $P_\eta\{\tilde{\phi} = 0\} < P_\eta\{\hat{\phi} = 0\}$ for any $\boldsymbol{\eta}$. For other models, both $\bar{\lambda} < \bar{\lambda}_J$ and $\bar{\lambda} > \bar{\lambda}_J$ are possible, so the above relation between the probabilities of the events $\{\tilde{\phi} = 0\}$ and $\{\hat{\phi} = 0\}$ may or may not hold.

5.1. Simulation Experiment

We use a Monte Carlo simulation experiment to compare some properties of the ML and REML estimators of the covariance parameters and investigate how these properties depend on some features of the model. Specifically, the quantities to be considered and compared are:

- The probability that the REML estimator $\tilde{\phi}$ does not exist;
- The probabilities of the events $\{\hat{\phi} = 0\}$ and $\{\tilde{\phi} = 0\}$;
- The root mean squared error (RMSE) of the ML and REML estimators of ϕ and σ^2.

The model features (factors) to be varied in the simulation experiment are lattice size, strength of spatial association and mean structure. We consider models defined over 12×12, 20×20 and 32×32 regular lattices with 'first order' neighborhood system and $g_{kl} = 1$ if $k \in N_l$. The spatial parameter ϕ would vary over a fine grid in $[0, 10]$ and $\sigma^2 = 4$. For the mean structure we consider models where:

(M0) $p = 1$, $\beta_1 = 1$ and $X = \mathbf{1}_n$;

(M1) $p = 5$, $\boldsymbol{\beta} = \mathbf{1}_5$ and $X = \left(W_1 \vdots \mathbf{1}_n\right)$ where W_1 is the $n \times 4$ matrix whose columns are eigenvectors of H corresponding to the four largest eigenvalues of H;

(M2) $p = 5$, $\boldsymbol{\beta} = \mathbf{1}_5$ and X is the $n \times 5$ matrix whose columns are eigenvectors of H corresponding to the five smallest eigenvalues of H (so the last column of X is $\mathbf{1}_n$).

For each possible combination of lattice size, spatial parameter and mean structure, 10000 datasets were simulated from model (1) and these replicated samples were used to estimate the quantities defined above.

5.2. Results

Results for mean structures M0, M1 and M2 were qualitatively very similar, so we only show figures for mean structure M0. For each of the considered models the probability that the REML estimator $\tilde{\phi}$ does not exist is estimated by the proportion of simulated datasets, $\mathbf{y}_{\text{sim}}$, for which $Q(\mathbf{y}_{\text{sim}}) \leq \bar{\zeta}_J$ holds. Figure 1 shows how this probability varies with ϕ for each of considered lattice sizes. The probability that $\tilde{\phi}$ does not exist increases with the true value of ϕ and decreases with lattice size. As pointed out by a referee, it is straightforward to prove analytically that $P_\phi(\tilde{\phi}$ does not exist) is an increasing function of ϕ. Except for the smallest lattice size, this probability is quite small and displays a similar pattern of variation with ϕ for mean structures M0 and M1. For mean structure M2 this probability is considerably higher, being about 0.3 when $\phi = 10$ for the 12×12 lattice (not shown). The use of REML estimation may then be problematic in small datasets.

For each of the considered models the probabilities $P_\eta\{\hat{\phi} = 0\}$ and $P_\eta\{\tilde{\phi} = 0\}$ are estimated by the proportion of simulated datasets for which, respectively, $Q(\mathbf{y}_{\text{sim}}) \geq \bar{\lambda}$ and $Q(\mathbf{y}_{\text{sim}}) \geq \bar{\lambda}_J$ hold. Both probabilities decrease quite rapidly with the true value of ϕ and with the lattice size (not shown). For example, for mean structure M0 with lattice size 12×12, both $P_\eta\{\hat{\phi} = 0\}$ and $P_\eta\{\tilde{\phi} = 0\}$ are less than 0.01 for all $\phi > 0.5$. Also, $P_\eta\{\tilde{\phi} = 0\}$ is fairly insensitive to the considered mean structures, while $P_\eta\{\hat{\phi} = 0\}$ is not. Overall, both methods have a small chance of estimating ϕ as zero for models with $\phi > 0$.

Finally, for each of the considered models the RMSE of $\hat{\phi}$ is estimated by the square root of the average of $\{(\hat{\phi}_i - \phi)^2 : i = 1, \ldots, 10^4\}$, with similar estimates for the RMSEs of $\hat{\sigma}^2$, $\tilde{\phi}$ and $\tilde{\sigma}^2$, except that the estimation of the RMSEs of REML estimators is made conditional on their existence, so it uses only those simulated datasets for which $Q(\mathbf{y}_{\rm sim}) > \bar{\zeta}_J$ holds. Figure 2 shows, for mean structure M0, how the RMSE of the ML (left panel) and REML (right panel) estimators of ϕ vary with ϕ for each of the considered lattice sizes. The RMSEs of both estimators decrease with lattice size and increase with ϕ. But the RMSE of $\tilde{\phi}$ increases with ϕ much faster than the RMSE of $\hat{\phi}$, specially for the smaller lattice sizes, to the point of being unacceptably large for most values of ϕ. The same behavior holds for mean structures M1 and M2 (not shown).

Figure 3 shows how the RMSE of the ML and REML estimators of σ^2 vary with ϕ for mean structure M0 and the considered lattice sizes. The pattern of variation of the RMSEs of these estimators is similar to those displayed by the estimators of ϕ, where again the RMSE of $\tilde{\sigma}^2$ is much larger than the RMSE of $\hat{\sigma}^2$. We discuss in the next section the (possible) cause for the unexpected behavior of REML estimators in this model.

6. Conclusions and Discussion

This work provides results on existence and uniqueness of ML and REML estimators of covariance parameter for GMRF model (1) when a particular relation holds between the regression and precision matrices. For the case when these estimators exist, the results also provide a simple way to compute them. The findings of this work can be potentially useful for the design and analysis of experiments when the available experimental units are grid-cells of a regular grid. For instance, a common goal in agricultural experiments is to estimate the effects on yield of different treatments, and model (1) could be used for that purpose. If the assignment of treatments and levels to the experimental units is done in a way that any of the conditions in Lemma 3.1 hold, then inference about the treatment effects based on ML or REML would be greatly simplified. This is currently being investigated.

The results of a small simulation experiment suggest that (some of) the frequentist properties of the ML estimators of the covariance parameters are much better that those of the REML estimators. This observation is both interesting and intriguing since in the statistical literature abound examples of Gaussian models where REML estimators of variance and covariance parameters have overall similar or better frequentist properties than ML estimators, but the opposite is the case for model (1).

The poor behavior of REML estimators seems to be due to fact that these estimators do not exist with positive probability, so there is an "abrupt change" in RMSE behavior between datasets for which REML do not exist and those for which they exist. We have empirically found that for datasets with $Q(\mathbf{y})$ just slightly larger than $\bar{\zeta}_J$, $\tilde{\phi}$ exists but is very large. For small samples, this may indicate that the distribution of $\tilde{\phi}$ is heavy-tailed, perhaps to the extent that RMSEs do not exist (i.e becomes infinity). It is also worth noting that for datasets on a 32×32 lattice the RMSEs of REML estimators are not as large and are somewhat comparable to those of ML estimators, at least for models M0 and M1 considered in Section 5. This suggest that the extremely poor behavior of REML estimators may be ameliorated when the sample size grows. This issue requires further investigation.

On related work, Demidenko and Massam (1999) and Birkes and Wulff (2003) provided general results on existence of ML and REML estimators for a large class of Gaussian variance components models. We note that there are qualitative differences between our results and those in Demidenko and Massam (1999). The conditions for existence of ML estimators in Demidenko and Massam (1999) are more stringent than the respective conditions for existence of REML estimators. That is, existence of ML estimators of variance components parameters implies existence of the respective REML estimators, while there are cases when REML estimators exist but ML estimators do not. For the GMRF model (1) the situation is the opposite since ML estimators always exist, while REML estimators do not exist with positive probability.

Finally, as pointed out by a referee, model (1) can be framed within the class of models studied by Birkes and Wulff (2003). To see this, let $\iota_1 > \ldots > \iota_{d-1} > \iota_d = 0$ be the *distinct* eigenvalues of H. From the spectral decomposition we have $H = \sum_{j=1}^d \iota_j E_j$, where $\{E_j\}_{j=1}^d$ are $n \times n$ orthogonal projection matrices which are also mutually orthogonal and $\sum_{j=1}^d E_j = I_n$. This implies that

$$\mathrm{var}\{\mathbf{Y}\} = \sigma^2 (I_n + \phi H)^{-1} = \sum_{j=1}^d \gamma_j E_j,$$

where $\gamma_j = \sigma^2/(1+\phi\iota_j)$, $j = 1, \ldots, d$. Now note that the parameter space of the covariance parameters, say Γ, can be written as

$$\begin{aligned}\Gamma &= \Big\{\Big(\frac{\sigma^2}{1+\phi\iota_1}, \ldots, \frac{\sigma^2}{1+\phi\iota_{d-1}}, \sigma^2\Big) \in \mathbb{R}^d : \sigma^2 \geq 0, \phi \geq 0\} \\ &= \{\boldsymbol{\gamma} \in \mathbb{R}^d : \gamma_j > 0 \text{ for all } j\Big\} \cap \mathcal{G},\end{aligned}$$

where

$$\mathcal{G} = \Big\{\Big(\frac{\delta\tau}{1+\phi\iota_1}, \ldots, \frac{\delta\tau}{1+\phi\iota_{d-1}}, \tau\Big) \in \mathbb{R}^d : \phi \geq 0, \tau \geq 0, \delta = 0, 1\Big\},$$

is a closed subset of $\mathbb{R}^d$. Then $\mathrm{var}\{\mathbf{Y}\}$ is of the form given in equation (3.2) of Birkes and Wulff (2003), so their Theorem 3.1(a) guarantees existence of the ML estimator of (σ^2, ϕ) in model (1). Nevertheless, the non-standard parameter space given above and the difficulty of checking the conditions of their result (given in terms of subspace relations) make applying this result cumbersome. In contrast, the approach we use is more natural for models that explicitly parametrize the precision matrix (rather than the covariance matrix) of the data, and our results involve easy-to-check conditions in terms of eigenvalue summaries.

Acknowledgments. We warmly thanks two anonymous referees for their comments and suggestions. Special mention goes to one of the referees for his/her extensive and in depth review that included correcting and sharpening some of the proofs. The first author was partially supported by a summer Grant from the College of Business of the University of Texas at San Antonio and U.S. National Science Foundation Grant DMS-0719508. The second author was partially supported by U.S. National Science Foundation Grant DMS-0907064.

Appendix

LEMMA 6.1. *Let $\Sigma(\phi)$ be an arbitrary $n \times n$ positive definite matrix whose entries are differentiable with respect to ϕ. Then*

$$\frac{\partial}{\partial\phi} S^2(\phi) = \mathbf{y}'\big(I_n - A(\phi)\big)'\big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1}\big)\big(I_n - A(\phi)\big)\mathbf{y},$$

where $\frac{\partial}{\partial\phi}M(\phi)$ denotes element-wise differentiation of the matrix $M(\phi)$.

Proof. To simplify differentiation we rewrite $S^2(\phi)$ as

$$S^2(\phi) = \mathbf{y}'\big(\Sigma(\phi)^{-1} - \Sigma(\phi)^{-1}X(X'\Sigma(\phi)^{-1}X)^{-1}X'\Sigma(\phi)^{-1}\big)\mathbf{y}.$$

Then

$$\begin{aligned}
\frac{\partial}{\partial\phi} S^2(\phi) &= \mathbf{y}'\big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1} - \frac{\partial}{\partial\phi}\big[\Sigma(\phi)^{-1}X(X'\Sigma(\phi)^{-1}X)^{-1}X'\Sigma(\phi)^{-1}\big]\big)\mathbf{y} \\
&= \mathbf{y}'\big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1} - \big[\big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1}X\big)(X'\Sigma(\phi)^{-1}X)^{-1}\big(X'\Sigma(\phi)^{-1}\big) \\
&\quad + \big(\Sigma(\phi)^{-1}X\big)\big(\frac{\partial}{\partial\phi}(X'\Sigma(\phi)^{-1}X)^{-1}\big)\big(X'\Sigma(\phi)^{-1}\big) \\
&\quad + \Sigma(\phi)^{-1}X(X'\Sigma(\phi)^{-1}X)^{-1}\big(\frac{\partial}{\partial\phi}(X'\Sigma(\phi)^{-1})\big)\big]\big)\mathbf{y} \\
&= \mathbf{y}'\big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1} - \big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1}\big)A(\phi) + A'(\phi)\big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1}\big)A(\phi) - A'(\phi)\big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1}\big)\mathbf{y} \\
&= \mathbf{y}'\big(I_n - A(\phi)\big)'\big(\frac{\partial}{\partial\phi}\Sigma(\phi)^{-1}\big)\big(I_n - A(\phi)\big)\mathbf{y},
\end{aligned}$$

where the next to the last identity is obtained by using the definition of $A(\phi)$ and the standard identity

$$\frac{\partial}{\partial\phi}M(\phi)^{-1} = -M(\phi)^{-1}\big(\frac{\partial}{\partial\phi}M(\phi)\big)M(\phi)^{-1}.$$

□

LEMMA 6.2. *For $\phi \geq 0$ and $\mathbf{v} \in \mathbb{R}^r$, with $\mathbf{v} \geq \mathbf{0}$, consider the function $m(\cdot,\cdot)$ defined as*

$$m(\phi, \mathbf{v}) = \frac{\frac{1}{r}\sum_{k=1}^{r}\left(\frac{v_k}{1+\phi v_k}\right)}{1 - \frac{\phi}{r}\sum_{k=1}^{r}\left(\frac{v_k}{1+\phi v_k}\right)}, \tag{6}$$

and assume that $\mathbf{v} \notin C(\mathbf{1}_r)$. Then:

(i) For every fixed $\mathbf{v}$, $m(\cdot, \mathbf{v})$ is continuous and strictly decreasing on $[0, \infty)$.

(ii) $m(0, \mathbf{v}) = \frac{1}{r}\sum_{k=1}^{r} v_k$.

(iii) If $v_k > 0$ for all k, then $\lim_{\phi\to\infty} m(\phi, \mathbf{v}) = \Big(\frac{1}{r}\sum_{k=1}^{r}\frac{1}{v_k}\Big)^{-1}$.

(iv) If $v_k = 0$ for some k, then $\lim_{\phi\to\infty} m(\phi, \mathbf{v}) = 0$.

Proof. (i) The continuity of $m(\cdot, \mathbf{v})$ is clear. To show strict monotonicity we rewrite $m(\phi, \mathbf{v})$ as

$$m(\phi, \mathbf{v}) = -\left(\phi - \Big(\frac{1}{r}\sum_{k=1}^{r}\Big(\frac{v_k}{1+\phi v_k}\Big)\Big)^{-1}\right)^{-1},$$

from which we have

$$\begin{aligned}\frac{\partial}{\partial\phi} m(\phi, \mathbf{v}) &= \left(\phi - \Big(\frac{1}{r}\sum_{k=1}^{r}\Big(\frac{v_k}{1+\phi v_k}\Big)\Big)^{-1}\right)^{-2}\left(1 - \frac{\frac{1}{r}\sum_{k=1}^{r}\left(\frac{v_k}{1+\phi v_k}\right)^2}{\left(\frac{1}{r}\sum_{k=1}^{r}\left(\frac{v_k}{1+\phi v_k}\right)\right)^2}\right) \\ &< 0 \quad \text{for all } \phi \geq 0,\end{aligned}$$

where the last inequality follows from Jensen's inequality and $\mathbf{v} \notin C(\mathbf{1}_r)$.

(ii) Straightforward.

(iii) For any $\phi > 0$, $m(\phi, \mathbf{v})$ can be rewritten as

$$m(\phi, \mathbf{v}) = \frac{\frac{1}{r\phi}\sum_{k=1}^{r}\frac{v_k}{v_k+\phi^{-1}}}{1 - \frac{1}{r}\sum_{k=1}^{r}\frac{v_k}{v_k+\phi^{-1}}} = \frac{\sum_{k=1}^{r}\frac{v_k}{v_k+\phi^{-1}}}{\sum_{k=1}^{r}\frac{1}{v_k+\phi^{-1}}}. \tag{7}$$

From this follows that when $v_k > 0$ for all k, $m(\phi, \mathbf{v}) \to \left(\frac{1}{r}\sum_{k=1}^{r}\frac{1}{v_k}\right)^{-1}$ as $\phi \to \infty$.

(iv) Without lost of generality let $1 \leq s \leq r$ such that $v_k > 0$ for $k = 1, \ldots, r-s$ and $v_k = 0$ for $k = r-s+1, \ldots, r$. Using (7) we have that for any $\phi > 0$

$$m(\phi, \mathbf{v}) = \frac{\sum_{k=1}^{r-s}\frac{v_k}{v_k+\phi^{-1}}}{\left(\sum_{k=1}^{r-s}\frac{1}{v_k+\phi^{-1}}\right) + s\phi},$$

from which follows that $m(\phi, \mathbf{v}) \to 0$ as $\phi \to \infty$. $\square$

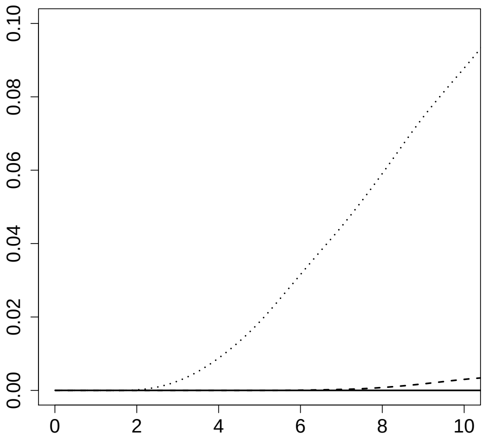

Fig. 1. Probability that the REML estimator $\tilde{\phi}$ does not exist as a function of ϕ for mean structure M0 and lattice sizes 12×12 (dotted), 20×20 (dashed), and 32×32 (solid).

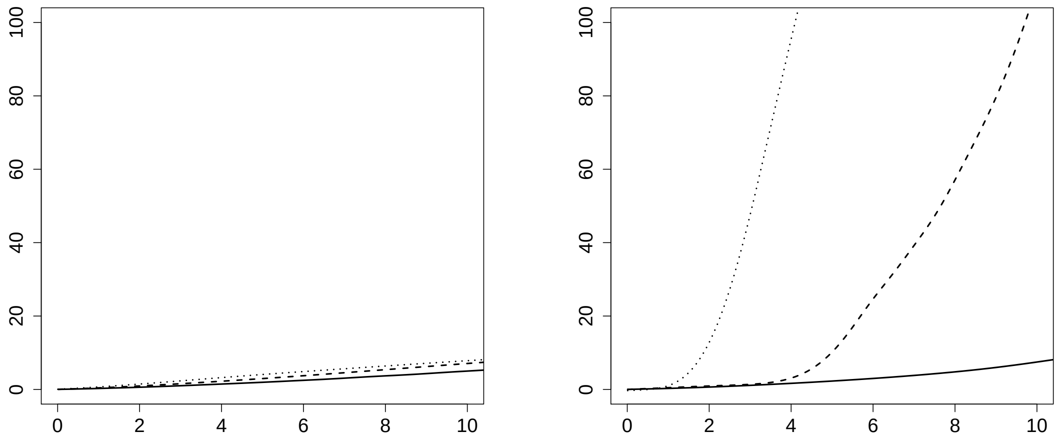

Fig. 2. Root mean squared error of the ML estimator $\hat{\phi}$ (left panel) and the REML estimator $\tilde{\phi}$ (right panel) for mean structure M0 and lattice sizes 12×12 (dotted), 20×20 (dashed) and 32×32 (solid).

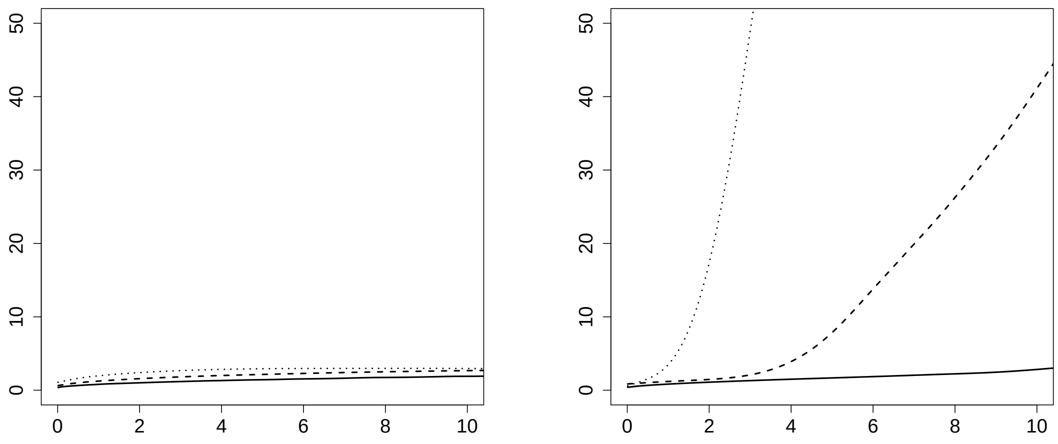

Fig. 3. Root mean squared error of the ML estimator $\hat{\sigma}^2$ (left panel) and the REML estimator $\tilde{\sigma}^2$ (right panel) for mean structure M0 and lattice sizes 12×12 (dotted), 20×20 (dashed) and 32×32 (solid).