Available online at www.sciencedirect.com

ScienceDirect

Journal of Multivariate Analysis 98 (2007) 1417–1440

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models

Petruţa C. Caragea[a,*,1], Richard L. Smith[b,2]

[a] *Department of Statistics, Iowa State University, 314 Snedecor Hall, Ames, IA 50011-1210, USA*

[b] *Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599-3260, USA*

Received 1 November 2005

Available online 18 October 2006

Abstract

Parameters of Gaussian multivariate models are often estimated using the maximum likelihood approach. In spite of its merits, this methodology is not practical when the sample size is very large, as, for example, in the case of massive georeferenced data sets. In this paper, we study the asymptotic properties of the estimators that minimize three alternatives to the likelihood function, designed to increase the computational efficiency. This is achieved by applying the information sandwich technique to expansions of the pseudo-likelihood functions as quadratic forms of independent normal random variables. Theoretical calculations are given for a first-order autoregressive time series and then extended to a two-dimensional autoregressive process on a lattice. We compare the efficiency of the three estimators to that of the maximum likelihood estimator as well as among themselves, using numerical calculations of the theoretical results and simulations.

AMS 1991 subject classification: 62M30; 62M10; 62F12; 62H11

Keywords: Approximate likelihood; Massive data sets; Computational efficiency; Statistical efficiency analysis; Spatial statistics; Autoregressive processes on a lattice

* Corresponding author. Fax: +1 515 294 4040.

E-mail address: pcaragea@iastate.edu (P.C. Caragea).

[1] Supported in part by a SAMSI graduate studentship (NSF Grant DMS-0112069) and by NSF Grant DMS-0084375 while a graduate student at the University of North Carolina at Chapel Hill.

[2] Supported in part by NSF Grant DMS-0084375.

0047-259X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.08.010

1. Introduction

Often times, it is assumed that modeling "geostatistical" processes within the realm of multivariate normal models is appropriate. There is by now a substantial literature on the selection and estimation of models for spatial data (see e.g. [10,22,9,3]). Many of the common geostatistical models may be expressed in the form

$$X \sim \mathcal{N}(\mu, \Sigma(\theta)), \tag{1}$$

with $\{X_i, i = 1, \ldots, N\}$ the vector of observations with mean μ and $N \times N$ covariance matrix $\Sigma(\theta)$ (assumed of a known format), expressed in terms of a finite-dimensional parameter vector θ. A simple example is the exponential covariance model, $\sigma_{ij} = \theta_1 e^{-d_{ij}/\theta_2}$ where σ_{ij} represents the covariance between observations X_i and X_j, d_{ij} is the physical distance between the two locations, and θ_1 and θ_2 are model parameters.

In such models, it is widely recognized that either the method of maximum likelihood or the closely related technique of restricted maximum likelihood [10,22] are the best general methods of estimation. Maximum likelihood is based on minimizing the negative log-likelihood function which, modulo some constants, is of the form

$$\ell(\mu, \theta) = \tfrac{1}{2} \log |\Sigma(\theta)| + \tfrac{1}{2} (Y - \mu)^T \Sigma(\theta)^{-1} (Y - \mu). \tag{2}$$

A disadvantage of likelihood-related techniques, however, is the computational efficiency of the likelihood function. Evaluation of (2) requires calculating the inverse and determinant of a $N \times N$ matrix, which according to most commonly used algorithms, requires $O(N^3)$ steps. Many modern data sets contain thousands of observations, for which such computation is prohibitively time consuming. Therefore, there is some interest in finding approximations to the likelihood function that are more efficient to compute, while trying not to sacrifice much statistical efficiency.

Several computationally efficient alternatives to maximum likelihood estimation have been explored in the literature. One such strategy was proposed by Vecchia [27] who computed approximate conditional densities along some sequence of arbitrarily ordered sampling points and ignoring long-range correlations. Recently, Stein et al. [23] generalized Vecchia's idea in a number of ways. They developed a variant of the method to approximate the restricted likelihood function in place of the likelihood function itself. They argued that, rather than evaluate conditional densities one observation at a time, it might be more efficient to do it in blocks.

Another approach to efficiently estimate semivariograms is explored by Curriero and Lele [11] who exploit the composite likelihood introduced by Lindsay [18] for the estimation of spatial hierarchical model parameters. The same principles are also applied to modeling binary data in a subsequent paper by Heagerty and Lele [14]. The main idea proposed is to construct an approximate log-likelihood function by adding marginal log-likelihoods even when these components do not necessarily represent independent replicates.

For massive data sets, several efforts have been put into developing efficient interpolation techniques that do not rely on using arbitrarily chosen local kriging neighborhoods (see [10]). Recent advancements in the field propose approximations to the kriging equations by tapering the covariance matrix, as in Furrer et al. [12] which may have the disadvantage of ignoring long-range correlations. Other suggestions include iterative methods to solve linear systems such as conjugate gradients as in Billings et al. [5] or using a multiresolution (wavelet) basis function as advocated by Nychka et al. [20].

An alternative approach introduced by Huang et al. [15] is to construct classes of covariance matrices that allow kriging to be performed exactly even for very large data sets. This idea is

based on a multiresolution tree-structured model that preserves mass balance across resolutions (i.e. it is resolution consistent) and uses a change of resolution Kalman filter. Building on these ideas, Johannesson and Cressie [16] and Tzeng et al. [25] propose several extensions of these models to improve the "blocky" structure of the initially proposed covariances and to extend it to spatio-temporal modeling. We have not taken this approach in the present work.

For stationary processes on a lattice, there are several approaches to approximating the likelihood efficiently as proposed, for example, by Whittle [29] or Guyon [13] or simplifying the exact likelihood computations as proposed by Zimmerman [30]. For autoregressive models on a lattice, Smirnov et al. [21] introduce an approach based on characteristic polynomials for calculating the determinant of a very large matrix, while Barry et al. [4] employ a method based on Monte Carlo estimation for sparse matrices.

In this paper we examine three other approximations to the log-likelihood, all aimed at improving computational efficiency. All three methods rely on some initial grouping of observations into blocks, as follows:

(1) The "big blocks" method reduces each block to its block mean; thus if there are B blocks, $\bar{X}_b$ is the mean of observations in the bth block ($1 \leqslant b \leqslant B$), μ_b is the expected value of $\bar{X}_b$ and Σ_B is the covariance matrix of $\bar{X}_1, \ldots, \bar{X}_B$, we evaluate the likelihood corresponding to (2) based on just the block means.
(2) The "small blocks" method replaces (2) by $\sum_b \ell_b(\mu, \theta)$ where $\ell_b(\mu, \theta)$ is the likelihood based on just the observations within block b; equivalently, this method treats the blocks as if they were independent of each other.
(3) The "hybrid" method combines the two concepts, by first computing the big-blocks likelihood, then the small-blocks likelihood *conditional* on the respective block means; these two log-likelihoods are then added to produce the hybrid log-likelihood. In effect, the hybrid method assumes that, conditional on the block means, the within-block deviations from the block mean are independent from block to block.

    Although in the present paper we have chosen to concentrate on the statistical rather than computational efficiency, the trade-off between the two could be a factor in practical applications. One may consider other conditioning choices for the hybrid estimator, that could lead to an even greater reduction in the computational effort. Such alternatives include conditioning on an arbitrarily chosen single observation or on the average of a smaller subset of observations within each block. Although these are interesting possibilities, evaluating their asymptotic efficiencies, along the lines given later in the paper, would essentially require a completely new set of calculations. For this reason, they are not considered further in the present work.

Provided B is chosen appropriately, all three approximate likelihoods are computable in at most $O(N^2)$ steps, a significant computational saving when N is of the order of several hundred to a few thousand, which is the range within which we envision such approximation being useful.

In a parallel paper, Caragea and Smith [8] adopt a more intuitive approach based on the "information sandwich" formula for assessing asymptotic variances of these estimators defined by estimating equations. The theoretical investigation is followed by numerical approximations and simulations to examine the properties of these methods in situations that are more plausible for modeling spatial data. This work is developed in the classical context of spatial regression with known (and constant) mean, and lays out the details of extending the present method to REML estimation, estimation of regression parameters and kriging. Two separate properties of these

estimators are examined in detail: (a) comparing the asymptotic variance of the proposed estimator with that of MLE; (b) assessing how well standard errors computed from the observed information approach (treating the approximate likelihood as if it were an exact likelihood) correspond to the true standard deviations of the estimators. Theoretical developments and numerical results are presented for stationary processes on a lattice with an exponential or Matérn covariance matrix. Comparisons of efficiencies suggest that the big blocks estimator is poor except when the range of the spatial covariance function is comparable with the range of the sampling locations. The efficiencies of the small blocks and hybrid estimators appear comparable in most circumstances, except when the Matérn model with small shape parameter, in which case the hybrid method appears clearly superior. In terms of the second criterion, the hybrid method is superior in the sense that estimated standard errors from inverting the approximate observed information matrix are closer to the true standard errors than those derived by the small blocks method. A real-data example based on rainfall trends suggests that the hybrid estimator is very often, but not invariably, closer to the true MLE than the small-blocks estimator, while the three sets of standard errors (using the direct method) are comparable. The quality of predictions produced by the three methods, assessed by a cross-validated mean squared prediction error, was almost identical for this example. In conclusion, the hybrid method is recommended as a good all-round alternative to the exact maximum likelihood estimation.

The present paper concentrates on more theoretical aspects, in particular deriving rigorous asymptotic efficiencies of the proposed estimators when the number of blocks, B tends to ∞ keeping the block sizes fixed.

The paper is organized as follows. Section 2 describes the theoretical tools we employ to derive the asymptotic variance of the alternative estimators, essentially based on the martingale central limit theorem applied to a white noise expansion of the model. Section 3 then describes a simple example based on the one-dimensional first-order autoregressive process (AR(1)). Although, for this example, the exact likelihood is easy to calculate and the asymptotic efficiency of MLE has also been established under suitable regularity conditions [1], the main purpose of our calculation is to provide a relatively simple illustration of how the general method works, in a situation where it leads to concrete analytic calculations of the asymptotic efficiency of our three approximate estimators. Section 4 extends the calculations for the asymptotic variances of the alternative estimators for a particular class of stationary processes on a lattice (essentially, Kronecker products of AR(1) processes). We conclude by drawing general remarks based on the results obtained for the two special cases considered here and suggest several extensions to the problem.

It should be noted here that although the proposed ideas of approximating the likelihood function were motivated by practical geostatistical applications, the present paper restricts its attention to one- and two-dimensional autoregressive processes. We are confining our theoretical calculations to such regular and simpler processes because their dependence structure provides an environment where mathematical calculations, although still not easy to obtain, are feasible, and not because there is any evidence they provide a good model for spatial data.

2. The "expansion method"

This section outlines the so-called expansion method, which is the main tool we use to prove our asymptotic results. It has three components: (a) the information sandwich formula for the asymptotic covariance matrix of a consistent estimator defined by general estimating equations (Section 2.1); (b) conditions for consistency (Section 2.2) and (c) an adaptation of the martingale

central limit theorem for proving asymptotic normality of quadratic forms of normal random variables (Section 2.3).

2.1. Information sandwich approach

Suppose we have a statistical model indexed by a finite-dimensional parameter θ, whose true value is denoted θ_0, and a consistent estimator $\tilde{\theta}_N$ constructed by minimizing a criterion function $S_N(\theta)$. We assume $S_N(\theta)$ is at least twice continuously differentiable in θ, and that its underlying distribution is sufficiently smooth that the function $H(\theta)$, defined below, is continuous in a neighborhood of θ_0. We write $\nabla f(\theta)$ for the vector of first-order partial derivatives of any function f with respect to the components of θ, and $\nabla^2 f$ for the matrix of second-order partial derivatives. Suppose:

(SA1) $\frac{1}{N}\nabla^2 S_N(\theta) \xrightarrow{\text{p}} H(\theta)$ as $N \to \infty$ uniformly on some neighborhood of θ_0, where $H(\cdot)$ is a matrix-valued function, continuous near θ_0, with $H(\theta_0)$ invertible;

(SA2) $\frac{1}{\sqrt{N}}\nabla S_N(\theta_0) \xrightarrow{\text{d}} \mathcal{N}(0, V(\theta_0))$ for some covariance matrix $V(\theta_0)$.

Then the asymptotic distribution of $\tilde{\theta}_N$ is

$$\sqrt{N}(\tilde{\theta}_N - \theta_0) \xrightarrow{\text{d}} \mathcal{N}(0, H(\theta_0)^{-1} V(\theta_0) H(\theta_0)^{-1}). \tag{3}$$

References for this method include Liang and Zeger [17] and White [28]. In addition, Stein et al. [23] developed a similar "information sandwich" approximation for the spatial covariance matrix of the resulting estimator to the one we propose here.

2.2. Consistency

We demonstrate consistency of our proposed estimators using Theorem 4.1.2 of Amemiya [2], which is as follows:

Theorem 1. *Assume*:

(A) Θ *is an open subset of the Euclidean p-space (the true value* θ_0 *is an interior point of* Θ*)*;

(B) *the criterion function* $S_N(\theta)$ *is a measurable function for all* $\theta \in \Theta$, *and* ∇S_N *exists and is continuous in an open neighborhood of* θ_0;

(C) $\frac{1}{N} S_N(\theta)$ *converges in probability uniformly to a non-stochastic function* $S(\theta)$ *in an open neighborhood of* θ_0, *and* $S(\theta)$ *attains a strict local maximum at* θ_0.

Then there exists a sequence $\varepsilon_N \to 0$ *such that*

$$P\left\{\exists \theta^* s.t. |\theta^* - \theta_0| < \varepsilon_N, \nabla S_N(\theta^*) = 0\right\} \to 1 \quad \text{as } N \to \infty.$$

Assumption (A) is one of the assumptions of our method, while (B) is satisfied by all our criterion functions. Since each of the approximate log-likelihoods considered in this paper is a sum of exact log-likelihoods for some subset of the data, it follows that the first-order derivatives of S_N are bounded on a neighborhood of θ_0, and that $\frac{1}{N} E|\nabla S_N(\theta)| \leqslant K$, on a neighborhood of θ_0. Using a first-order Taylor's expansion, it is clear that for some θ_N^* and θ_N^{**} between θ_0 and θ,

we have that

$$\frac{1}{N}S_N(\theta) - \frac{1}{N}S_N(\theta_0) = \frac{1}{N}\nabla S_N(\theta_N^*)(\theta - \theta_0) \tag{4}$$

and

$$S(\theta) - S(\theta_0) = \nabla S(\theta_N^{**})(\theta - \theta_0). \tag{5}$$

Therefore the difference between (4) and (5) is

$$\left\| \left(\frac{1}{N}S_N(\theta) - S(\theta) \right) - \left(\frac{1}{N}S_N(\theta_0) - S(\theta_0) \right) \right\| \leqslant \Gamma \|\theta - \theta_0\|, \tag{6}$$

where Γ has finite expectation. Note that the right-hand side of Eq. (6) converges to 0 uniformly over a decreasing sequence of neighborhoods of the form $\|\theta - \theta_0\| < \varepsilon_N$, for any sequence of ε_N tending to 0. Also, $N^{-1}S_N(\theta_0) - S(\theta_0) \xrightarrow{\text{p}} 0$ by the law of large numbers. Therefore, $N^{-1}S_N(\theta) - S(\theta)$ converges to 0 uniformly on a neighborhood of θ_0 such that $\|\theta - \theta_0\| < \varepsilon_N$, which proves condition (C) of Theorem 1.

Henceforth, we assume that the conditions in this subsection are satisfied for all of our estimators.

2.3. Properties of quadratic forms of normal random variables

For all our estimators, the approximate log-likelihood is a quadratic form of normal random variables, so the information sandwich approach requires that we prove a weak law of large numbers and a central limit theorem for such functions. We concentrate here on the CLT; the corresponding WLLN is an easy corollary.

Consider the sequence

$$S_N = \sum_{\{i,j:i \leqslant j\}} a_{N,i,j}\xi_i\xi_j, \tag{7}$$

where $\{\xi_i\}$ are independent $\mathcal{N}(0, 1)$, and coefficients $\{a_{N,i,j}\}$ are defined for each N. We are interested in limits as $N \to \infty$. In principle the sum in (7) extends across $1 \leqslant i \leqslant j < \infty$ though in practice the sum is often truncated, with N denoting the length of the sequence. We can then calculate the mean

$$m_N = \mathrm{E}[S_N] = \sum_i a_{N,i,i} \tag{8}$$

and the variance

$$v_N = \mathrm{Var}[S_N] = 2\sum_i a_{N,i,i}^2 + \sum_{\{i,j:i<j\}} a_{N,i,j}^2. \tag{9}$$

This implies the natural conjecture that with m_N and v_N defined by (8) and (9)

$$\frac{S_N - m_N}{\sqrt{v_N}} \xrightarrow{\text{d}} \mathcal{N}(0, 1). \tag{10}$$

Theorem 2. *Suppose*

(A1) $\max_i a_{N,i,i}^2/v_N \to 0 \quad as\ N \to \infty.$

(A2) $\max_k \left(\sum_{i:i<k} a_{N,i,k}^2 \right) / v_N \to 0 \quad as\ N \to \infty.$

Then (10) *holds*.

This is a consequence of the discrete-time martingale central limit theorem [6, Theorem 35.12].

We also note a "symmetric" version of the same result: if $a_{N,i,j} = a_{N,j,i}$ for all N, i, j and S_N is defined by

$$S_N = \sum_i \sum_j a_{N,i,j} \xi_i \xi_j, \tag{11}$$

then with m_N again defined by (8), and v_N by

$$v_N = 2 \sum_i \sum_j a_{N,i,j}^2, \tag{12}$$

Theorem 2 holds with (A1) and (A2) combined into a single condition:

(A3) $\max_k \sum_i a_{N,i,k}^2/v_N \to 0.$

In this formulation, the result is no longer dependent on any ordering of the indices. This is particularly useful when the indices are no longer integers but may be arbitrary points on a lattice, as is the case for our later results based on random fields.

3. Applications of the expansion method to a one-dimensional autoregressive process

In this section, we illustrate the expansion method to derive the asymptotic properties of our three estimators in a relatively easy (but still novel) case: the one-parameter first-order autoregressive process (AR(1)).

We define the process by $X_{i+1} = \phi X_i + \varepsilon_{i+1}$, where $|\phi| < 1$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ independently. Alternatively, we write the process in the form

$$X_i = \sigma_\varepsilon \sum_{r=-\infty}^{i} \phi^{i-r} \xi_r, \tag{13}$$

with ξ_r independent $\mathcal{N}(0, 1)$. If $U = (u_{i,j})_{1 \leqslant i,j \leqslant N}$ denotes the covariance matrix, $U^{-1} = (u_{i,j}^\star)_{1 \leqslant i,j \leqslant N}$ its inverse and $|U|$ its determinant, then

$$u_{i,j} = \frac{\sigma_\varepsilon^2}{1-\phi^2} \phi^{|i-j|}, \tag{14}$$

$$|U| = \frac{\sigma_\varepsilon^{2N}}{1-\phi^2}, \tag{15}$$

$$u_{i,j}^\star = \frac{1}{\sigma_\varepsilon^2} \begin{cases} 1 & \text{if } i = j = 1 \text{ or } i = j = N, \\ 1+\phi^2 & \text{if } 2 \leqslant i = j \leqslant N-1, \\ -\phi & \text{if } |i-j| = 1, \\ 0 & \text{elsewhere.} \end{cases} \tag{16}$$

For simplicity of subsequent calculations, we assume σ_ε^2 is known.

3.1. Classical maximum likelihood estimator

The asymptotic behavior of the maximum likelihood estimator of an AR(1) model is of course well known (see, for example [7]), but is given here to illustrate the expansion method. The negative log-likelihood function is defined by (2) with $\mu = 0$, $\Sigma = U$, so the maximum likelihood estimator for ϕ solves

$$\frac{\phi}{1-\phi^2} - \frac{1}{\sigma_\varepsilon^2}\left[\sum_{s=1}^{N-1} X_s(X_{s+1} - \phi X_s) + \phi X_1^2\right] = 0. \tag{17}$$

It follows at once that $E\left[\partial_\phi \ell(\phi)\right] = 0$. We also need to calculate $\mathrm{Var}\left[\partial_\phi \ell(\phi)\right]$, which is equivalent to calculating

$$\frac{1}{\sigma_\varepsilon^4}\mathrm{Var}\left[\sum_{s=2}^{N}\sum_{r=-\infty}^{s-1} \phi^{s-r-1}\xi_s\xi_r + 2\sum_{s=-\infty}^{1}\sum_{r=-\infty}^{s-1} \phi^{3-s-r}\xi_s\xi_r + \sum_{s=-\infty}^{1} \phi^{3-2s}\xi_s^2\right].$$

The expression we need to calculate the variance of is of the form given in (7), with

$$a_{N,s,r} = \begin{cases} \sigma_\varepsilon^2\phi^{3-2s} & \text{if } r = s \leqslant 1, \\ 2\sigma_\varepsilon^2\phi^{3-s-r} & \text{if } s \leqslant 1 \text{ and } r \leqslant s-1, \\ \sigma_\varepsilon^2\phi^{s-r-1} & \text{if } 2 \leqslant s \leqslant N-2 \text{ and } r \leqslant s-1, \\ 0 & \text{elsewhere.} \end{cases} \tag{18}$$

It follows that

$$\begin{aligned} \mathrm{Var}\left[\partial_\phi \ell(\phi)\right] &= 2\sum_s a_{N,s,s}^2 + \sum_{s<r} a_{N,s,r}^2 \\ &= 2\sum_{s=-\infty}^{1} \phi^{2(3-2s)} + \sum_{s=2}^{N}\sum_{r=-\infty}^{s-1} \phi^{2(s-r-1)} + \sum_{s=-\infty}^{1}\sum_{r=-\infty}^{s-1} 4\phi^{2(3-s-r)} \\ &= \frac{N-1-(N-3)\phi^2}{(1-\phi^2)^2}. \end{aligned} \tag{19}$$

Also, for any fixed r, the sum $\sum_{s<r} a_{N,s,r}^2$ is bounded by a constant, therefore the condition (A2) in Theorem 2 is satisfied, and so is (A1). As a consequence, the asymptotic distribution of the gradient of the negative log-likelihood function is normal, with mean 0 and variance given by expression (19).

A similar calculation, whose details we omit, shows directly that (19) is also the mean of the second derivative of the negative log-likelihood function, and since this is $\sim N/(1-\phi^2)$, it follows that

$$\sqrt{N}(\hat{\phi} - \phi) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, 1-\phi^2\right), \tag{20}$$

in agreement with Brockwell and Davis [7, Example 8.8.1, p. 259].

3.2. Big blocks estimator

Suppose we divide the time series of length N into B blocks of length K, so that $N = BK$. In practice, we might have blocks of slightly unequal length to allow for the possibility that N is

not divisible by K, but for the purpose of the present exposition, we consider only cases where $N = BK$ exactly, and derive asymptotic results as $B \to \infty$ for fixed K.

If X_b^* denotes the mean of the bth block, in other words $X_b^* = K^{-1}\sum_{j=1}^{K} X_{(b-1)K+j}$, and if $\gamma_m^* = \mathrm{Cov}[X_b^*, X_{b+m}^*]$ denotes the autocovariance function of the block means process, then we readily calculate

$$\gamma_m^* = \begin{cases} \left[\dfrac{2\phi^{K+1} - 2\phi - K\phi^2 + K}{K^2(1-\phi)^2(1-\phi^2)}\right]\sigma_\varepsilon^2 & \text{if } m = 0, \\ \left[\dfrac{\phi(1-\phi^K)^2}{K^2(1-\phi)^2(1-\phi^2)}\right]\sigma_\varepsilon^2 & \text{if } m = 1, \\ (\phi^k)^{m-1}\gamma_1^* & \text{if } m \geqslant 2. \end{cases}$$

This autocovariance function is of ARMA(1,1) form (see, e.g. [7, Problem 3.16, p. 112]) and we could in principle use the asymptotic distribution of MLE for an ARMA(1,1) process [7, Example 8.8.3, pp. 259–260] as the basis for calculating the asymptotic variance of the big blocks estimator. However, the details of this calculation are by no means straightforward. We prefer to proceed directly, using the expansion method, since these calculations are also needed for the development of the hybrid estimator.

For notational convenience, we have defined V_{means} to be the covariance matrix of $(X_1^*, \ldots, X_B^*)$, and v_{ij}' to be the (i, j) entry of $\partial_\phi V_{\text{means}}^{-1}(\phi)$. We calculate V_{means}^{-1} analytically, using the algorithm of Trench [24] for inverses of Toeplitz matrices.

If we define $p_{\text{means}}(\phi)$ to be the negative log-likelihood of $(X_1^*, \ldots, X_B^*)$ and $\partial_\phi p_{\text{means}}(\phi)$ its derivative with respect to ϕ, then, adapting the formula (2), we find that, modulo some fixed constants,

$$\partial_\phi p_{\text{means}}(\phi) = \frac{1}{2}\left[\frac{1}{K^2}\sum_{i=1}^{B}\sum_{j=1}^{B} v_{ij}' \sum_{\ell=1}^{K}\sum_{m=1}^{K} X_{(i-1)K+\ell} X_{(j-1)K+m} + \frac{\partial_\phi |V_{\text{means}}|}{|V_{\text{means}}|}\right]. \quad (21)$$

Using the expansion (13) in (21), we find

$$\begin{aligned} &\mathrm{Var}[\partial_\phi p_{\text{means}}(\phi)] \\ &\quad = \frac{\sigma_\varepsilon^4}{4K^2}\,\mathrm{Var}\left[\sum_{i=1}^{B}\sum_{j=1}^{B}\sum_{\ell=1}^{K}\sum_{m=1}^{K}\sum_{r=-\infty}^{(i-1)K+\ell}\sum_{s=-\infty}^{(j-1)K+m} v_{ij}'\phi^{(i+j-2)K+\ell+m-r-s}\xi_r\xi_s\right]. \quad (22) \end{aligned}$$

If we define

$$a_{N,r,s}^{(1)} = \frac{\sigma_\varepsilon^2}{2K}\sum_{i=1}^{B}\sum_{j=1}^{B}\sum_{\ell=\eta}^{K}\sum_{m=v}^{K} v_{ij}'\phi^{(i+j-2)K+\ell+m-r-s}, \quad (23)$$

where the summation lower bounds are defined as $\eta = f_1(i, r, K)$, $v = f_1(j, s, K)$, and f_1 is given by

$$f_1(j, r, K) = \begin{cases} 1 & \text{if } r - (j-1)K \leqslant 1, \\ r - (j-1)K & \text{if } 1 < r - (j-1)K \leqslant K, \\ K+1 & \text{if } r - (j-1)K \geqslant K+1, \end{cases} \quad (24)$$

it follows that the expression within the brackets in Eq. (22) is rewritten as $\sum_{\{r,s:r<s\}} a_{N,r,s}^{(1)}\xi_r\xi_s$.

Table 1
Time series: big blocks asymptotic relative efficiency

ϕ	$B = 5, K = 100$		$B = 10, K = 50$		$B = 50, K = 10$	
	Theory	Sim.	Theory	Sim.	Theory	Sim.
−0.750	0.00214	0.002	0.00330	0.003	0.00549	0.005
−0.250	0.01166	0.013	0.02265	0.020	0.08982	0.080
−0.010	0.01925	0.018	0.03773	0.036	0.15929	0.158
0.010	0.02003	0.019	0.03929	0.039	0.16702	0.165
0.250	0.03280	0.032	0.06434	0.055	0.27280	0.269
0.750	0.13367	0.132	0.25465	0.255	0.73897	0.724

The conditions of Theorem 2 are satisfied and we obtain that

$$E[\partial_\phi p_{\text{means}}(\phi)] = \sum_r a^{(1)}_{N,r,r} = 0$$

and

$$\text{Var}[\partial_\phi p_{\text{means}}(\phi)] = 2\sum_r {a^{(1)}_{N,r,r}}^2 + \sum_{\{r,s:r<s\}} {a^{(1)}_{N,r,s}}^2. \tag{25}$$

Since the summands in (25) include elements of the gradient of the inverse covariance matrix for the block means (V_{means}), we have not attempted to evaluate (25) analytically but instead give numerical results. The numerical calculations involve the use of the Trench [24] algorithm for inverses of Toeplitz matrices mentioned above and truncation of the sums in (25).

Moreover, since the big blocks estimator is the maximum likelihood estimator based on the block means, the expected value of $\partial^2_\phi p_{\text{means}}(\phi)$ is the same as (25) and therefore, in this case, the information sandwich approximation to the asymptotic variance is just the reciprocal of (25).

As a measure of performance for the big blocks estimator $\hat{\phi}_1$, we compute the relative asymptotic efficiency as the ratio between its asymptotic variance and that of the maximum likelihood estimator:

$$e_1(\hat{\phi}, \hat{\phi}_1) = \frac{\text{Var}[\hat{\phi}]}{\text{Var}[\hat{\phi}_1]}. \tag{26}$$

Expression (26) has been evaluated numerically, for various values of ϕ. In an effort to maintain a baseline for comparison provided by the ML estimator, we kept the sample size (N) fixed in all the numerical calculations. Results obtained for an AR(1) time series of length 500 and various number of blocks (5, 10 and 50) are presented in Table 1 (the columns labeled "Theory").

In addition to the numerical calculations of the theoretical results, we also performed a simulation study. These results, based on 1000 replications, are reported in Table 1 under the columns labeled "Sim.". The results obtained from theoretical calculations and simulations agree, taking into consideration the simulation-induced error.

A scrutiny of the results in Table 1 leads to the conclusion that summarizing block information only through its mean is not statistically efficient (note the poor efficiency of the big blocks estimator with respect to the maximum likelihood estimator). The only cases where its efficiency increases to a satisfactory level is when block sizes are very small (which is to be expected), in which case the method is not attractive from the computational perspective (the number of blocks is very close to the original number of observations).

Based on the results obtained in this subsection, we could not recommend the use of the big blocks estimator as an alternative to the MLE, in spite of its computational efficiency, except in a few specific situations. It is, nevertheless, a very important step in the theoretical development of the asymptotic properties of the proposed alternative estimators. The general methodology used to calculate the relative efficiency is incorporated and extended to the development of the intuitively more interesting cases, small blocks and hybrid pseudo-likelihood functions.

3.3. Small blocks estimator

In this subsection, we assume the same setting as for the big blocks estimator (in particular, we assume a time series of length N is divided into B blocks of length K, where $N = BK$) but instead consider the small blocks estimator.

If we denote by $X_j^K = (X_{(j-1)K+1}, \ldots, X_{jK})$ the vector of K observations in the jth block and by U_K the $K \times K$ covariance matrix given by (14) (identical for all blocks), we can write the negative log-likelihood function for a given block j as

$$\ell_j(\phi) = \frac{K}{2}\log 2\pi + \frac{1}{2}\left(X_j^{K\,T} U_K{}^{-1} X_j^K + \log|U_K|\right). \tag{27}$$

Since the pseudo-likelihood function is the product of the B block likelihoods, it follows that the negative pseudo-log-likelihood function has the form

$$p_{\text{Small Blocks}}(\phi) = \frac{KB}{2}\log 2\pi + \frac{B}{2}\left(\sum_{j=1}^{B} X_j^{K\,T} U_K{}^{-1} X_j^K + \log|U_K|\right), \tag{28}$$

which, modulo fixed constants and using (14) and (16), is equivalent to

$$\begin{aligned} p_{\text{Small Blocks}}(\phi) &\cong B\log\frac{1}{1-\phi^2} + \sum_{j=1}^{B} X_j^{K\,T} U_K^{-1} X_j^K \\ &= -B\log(1-\phi^2) + \frac{1}{\sigma_\varepsilon^2}\sum_{j=1}^{B}\left[(1+\phi^2)\sum_{i=2}^{K-1} X_{(j-1)K+i}^2\right. \\ &\quad \left. - 2\phi\sum_{i=1}^{K-1} X_{(j-1)K+i} X_{(j-1)K+i+1}\right]. \end{aligned} \tag{29}$$

The small blocks estimator, denoted by $\hat{\phi}_2$, minimizes the function in (29).

The first step is to check consistency of the estimator. For this, we could use the general methods of Section 2.2, but in this case, since ϕ is a one-dimensional parameter, it is simpler to use the following result [26, Lemma 5.10]:

Lemma. *Let Φ be a subset of the real line and let ψ_N be random functions and ψ a fixed function of ϕ such that $\psi_N(\phi) \xrightarrow{\text{p}} \psi(\phi)$ for every ϕ. Assume that each map $\psi_N(\phi)$ is non-decreasing with $\psi_N(\hat{\phi}_N) = o_p(1)$ or is continuous and has exactly one zero, $\hat{\phi}_N$. Let ϕ_0 be a point such that $\psi(\phi_0 - \varepsilon) < 0 < \psi(\phi_0 + \varepsilon)$ for every $\varepsilon > 0$. Then $\hat{\phi}_N \xrightarrow{\text{p}} \phi_0$.*

For the case of the small blocks estimator, it is readily checked that the expression $\partial_\phi p_{\text{Small Blocks}}(\phi)$ is an increasing function of ϕ, and the remaining conditions are standard

applications of the WLLN and CLT. Therefore, we conclude that the small blocks estimator $\hat{\phi}_2$ exists (as a function of sample size N) and is consistent as $N \to \infty$.

We use the expansion method to calculate the asymptotic variance of the small blocks estimator. In particular, the first derivative of the negative pseudo-log-likelihood function in (28), modulo fixed constants, equals

$$\partial_\phi p_{\text{Small Blocks}}(\phi) = \sum_{j=1}^{B}\sum_{\ell=1}^{K}\sum_{m=1}^{K} X_{(j-1)K+\ell} u'_{\ell m} X_{(j-1)K+m} + \frac{\partial_\phi |U_K|}{|U_K|}. \tag{30}$$

Here we denote $\partial_\phi U_K^{-1}$ by U' with the (i, j)th entry given by u'_{ij}. Using the expansion (13), the summation in (30) becomes

$$\sigma_\varepsilon^2 \sum_{j=1}^{B}\sum_{\ell=1}^{K}\sum_{m=1}^{K} \sum_{r=-\infty}^{(j-1)K+\ell} \sum_{s=-\infty}^{(j-1)K+m} u'_{\ell m} \phi^{2(j-1)K+\ell+m-r-s} \xi_r \xi_s. \tag{31}$$

Denoting by

$$a_{N,r,s} = \sigma_\varepsilon^2 \sum_{j=1}^{B}\sum_{\ell=\lambda}^{K}\sum_{m=v}^{K} u'_{\ell m} \phi^{2(j-1)K+\ell+m-r-s}, \tag{32}$$

where the lower summation bounds are $\lambda = f_1(j, r, K)$, $v = f_1(j, s, K)$ with f_1 as in (24), we rewrite the expression in (31) as

$$\sum_{\{r,s:r<s\}} a_{N,r,s} \xi_r \xi_s.$$

Since all the conditions set by Theorem 2 are satisfied, we obtain that

$$E[\partial_\phi p_{\text{Small Blocks}}(\phi)] = \sum_r a_{N,r,r} = 0$$

and

$$\text{Var}[\partial_\phi p_{\text{Small Blocks}}(\phi)] = 2\sum_r {a_{N,r,r}}^2 + \sum_{\{r,s:r<s\}} {a_{N,r,s}}^2. \tag{33}$$

The simpler form of the small blocks pseudo-likelihood, which is apparent from Eq. (29), is due to the block independence assumption and the simple structure of the inverse block covariance matrix as given by (16). After lengthy manipulations, it can be shown that the expression (33) reduces to

$$\begin{aligned}\text{Var}[\partial_\phi p_{\text{Small Blocks}}(\phi)] = {} & \frac{1}{B^2(1-\phi^{2K})^2\left(\frac{2\phi^2}{1-\phi^2}+K-1\right)^2} \\ & \times\Big\{-4\phi^2(\phi^{2BK}-1)\Big[-\phi^2+\phi^{2K}(1-\phi^2)^2(K-1)\Big] \\ & +B(\phi^{2K}-1)\Big[1-3\phi^2-4\phi^4-K+K\phi^2+\phi^{2K}(-1+K \\ & +\phi^2(3K-1+4\phi^2(-2+\phi^2)(K-1))\Big)\Big]\Big\}. \end{aligned} \tag{34}$$

Table 2
Time series: small blocks asymptotic relative efficiency

ϕ	$B = 5, K = 100$		$B = 10, K = 50$		$B = 50, K = 10$	
	Theory	Sim.	Theory	Sim.	Theory	Sim.
−0.750	0.98998	0.999	0.97878	0.990	0.92595	0.934
−0.250	0.99292	0.991	0.98407	0.977	0.91329	0.898
−0.010	0.99199	0.990	0.98197	0.980	0.90182	0.891
0.010	0.99199	0.990	0.98197	0.989	0.90182	0.892
0.250	0.99292	0.992	0.98407	0.985	0.91329	0.912
0.750	0.98998	0.993	0.97878	0.992	0.92595	0.942

To complete the derivations of the asymptotic variance of the small blocks estimator we need the expected value of $\partial^2_\phi p_{\text{Small Blocks}}(\phi)$, which is a routine calculation leading to the following expression:

$$E[\partial^2_\phi p_{\text{Small Blocks}}(\phi)] = \frac{2B}{1-\phi^2}\left(\frac{2\phi^2}{1-\phi^2} + K - 1\right). \tag{35}$$

According to the information sandwich method (as in expression (3)), the asymptotic variance of the small blocks estimator is the ratio between (34) and (35), both of which we have calculated explicitly.

Although having the analytical form of the variance above is very appealing, the calculations leading to it are very involved. In addition, these derivations make extensive use of the specific form of the small blocks assumptions and simple structure of the inverse block covariance matrix. Therefore, in the rest of the paper, we will just concentrate on numerical evaluations.

Numerical results for several partitions of the time series (5, 10 and 50 blocks) and various values of ϕ are reported by Table 2, under the columns labeled "Theory". Also, these calculations are accompanied by results from a simulation study with 1000 replications (under the columns labeled "Sim.").

A close analysis of the results presented in Table 2 indicates that the small blocks estimator is asymptotically highly efficient when compared to the maximum likelihood estimator. We notice a slight decrease in efficiency with the decrease of block sizes. This is due to the block independence assumption being most likely violated by configurations consisting of small blocks. Also, we note that simulation based results agree with the theoretical calculations (modulo simulation induced error). These observations lead us to conclude that the small blocks method produces asymptotically highly efficient estimators, while considerably decreasing computation time when compared to the classical maximum likelihood estimation.

3.4. Hybrid estimator

In this subsection, we assume the same setting as in the previous two subsections, but now consider the hybrid estimator. The hybrid negative pseudo-log-likelihood function is

$$p_{\text{Hybrid}}(\phi) = p_{\text{means}}(\phi) + \sum_{j=1}^{B} p_{\text{cond}_j}(\phi), \tag{36}$$

where $p_{\text{means}}(\phi)$, is given in Section 3.2 by (21) and $p_{\text{cond}_j}(\phi)$ is the negative block conditional log-likelihood, whose construction is explained here.

For block j, let us denote by $X_j^{K-1} = (X_{(j-1)K+1}, \ldots, X_{jK-1})$ the vector of all but one observation. The following argument follows identically for any block j, but we illustrate it here for the first block because of its notational simplicity.

In Section 3.2 we denoted the block average by $X_1^* = \frac{1}{K}\sum_{i=1}^{K} X_i$. It follows that the joint distribution of X_1^{K-1} and the group average X_1^* has the form

$$(X_1^{K-1}, X_1^*) \sim \mathcal{N}\left(0, \begin{pmatrix} U_{K-1} & \tau \\ \tau^T & \eta \end{pmatrix}\right) \tag{37}$$

and that the conditional distribution of X_1^{K-1} given X_1^* is given by

$$(X_1^{K-1} | X_1^*) \sim \mathcal{N}\left(\frac{\tau}{\eta} X_1^*, U_{K-1} - \tau\eta^{-1}\tau^T\right). \tag{38}$$

Here U_{K-1} is the covariance matrix of an AR(1) time series of length $K-1$, (see (14)), whose inverse is given by expression (16) and the determinant is $\frac{\sigma_\varepsilon^{2(K-1)}}{1-\phi^2}$. Also, $\eta = \text{Var}[X_1^*] = \gamma_0^* = \left[\frac{2\phi^{K+1} - 2\phi - K\phi^2 + K}{K^2(1-\phi)^2(1-\phi^2)}\right]\sigma_\varepsilon^2$ as given by (21), and $\tau = \text{Cov}[X_1^{K-1}, X_1^*] = (\tau_i)_{1 \leqslant i \leqslant K-1}$, where

$$\tau_i = \frac{\gamma_{i-1} + \gamma_{i-2} + \cdots + \gamma_0 + \gamma_1 + \cdots + \gamma_{K-i}}{K} = \frac{1 + \phi - \phi^i - \phi^{K-i+1}}{K(\phi^2 - 1)(\phi - 1)}\sigma_\varepsilon^2. \tag{39}$$

If we denote by $V(\phi)_{\text{cond}_1}$ the conditional covariance matrix for the first block, we can calculate its determinant and inverse as

$$|V_{\text{cond}_1}^{-1}(\phi)| = \frac{2\phi^{K+1} - K\phi^2 - 2\phi + K}{\sigma_\varepsilon^{2(K-1)}(1-\phi)^2}$$

and

$$V_{\text{cond}_1}^{-1}(\phi) = \frac{1}{\sigma_\varepsilon^2}\begin{pmatrix} 2 & 1-\phi & 1 & \ldots & 1 & 1 & 1+\phi \\ 1-\phi & 2+\phi^2 & 1-\phi & \ldots & 1 & 1 & 1+\phi \\ 1 & 1-\phi & 2+\phi^2 & \ldots & 1 & 1 & 1+\phi \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \ldots & 2+\phi^2 & 1-\phi & 1+\phi \\ 1 & 1 & 1 & \ldots & 1-\phi & 2+\phi^2 & 1 \\ 1+\phi & 1+\phi & 1+\phi & \ldots & 1+\phi & 1 & \phi^2+2\phi+2 \end{pmatrix}.$$

Since the form of the block conditional covariance matrix is independent of the block number, we omit the index in the following derivations. Also, denote the conditional mean for block j by μ^{cond_j}.

Henceforth, from (21), (36) and (38) the hybrid negative pseudo-log-likelihood function has the form

$$p_{\text{Hybrid}}(\phi) = -\frac{1}{2}\Bigg\{\log|V_{\text{means}}^{-1}| + X^{*T} V_{\text{means}}^{-1} X^* + B\log|V_{\text{cond}}| + \sum_{j=1}^{B}\left(X_j^{K-1} - \mu^{\text{cond}_j}\right)^T V_{\text{cond}}^{-1}\left(X_j^{K-1} - \mu^{\text{cond}_j}\right)\Bigg\}. \tag{40}$$

The hybrid estimator, denoted by $\hat{\phi}_3$, minimizes the function in (40), whose first and second derivatives with respect to ϕ are denoted by $\partial_\phi p_{\text{Hybrid}}(\phi)$ and $\partial^2_\phi p_{\text{Hybrid}}(\phi)$.

To simplify further notation, denote by

$$g(\phi) = \log|V_{\text{means}}| + B\log|V_{\text{cond}}|, \tag{41}$$

by w'_{ij}, w''_{ij}, v'_{ij}, v''_{ij} the (i,j) entry of $\partial_\phi V^{-1}_{\text{cond}}(\phi)$, $\partial^2_\phi V^{-1}_{\text{cond}}(\phi)$, $\partial_\phi V^{-1}_{\text{means}}(\phi)$, $\partial^2_\phi V^{-1}_{\text{means}}(\phi)$, and by ${\mu'}_i^{\text{cond}_j}$, ${\mu''}_i^{\text{cond}_j}$ the ith element of $\partial_\phi \mu^{\text{cond}_j}(\phi)$ and $\partial^2_\phi \mu^{\text{cond}_j}(\phi)$, respectively. Then the first derivative of the pseudo-log-likelihood function, modulo fixed constants, is given by

$$\begin{aligned}
\partial_\phi p_{\text{Hybrid}}(\phi) \;\cong\; & \partial_\phi g(\phi) + \frac{1}{K^2}\sum_{j=1}^{B}\sum_{i=1}^{B}\sum_{\ell=1}^{K}\sum_{m=1}^{K} v'_{ij} X_{(i-1)K+\ell} X_{(j-1)K+m} \\
& + \sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1} w'_{\ell m} X_{(j-1)K+\ell} X_{(j-1)K+m} \\
& - 2\sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1}\left(w'_{\ell m} X_{(j-1)K+\ell}\mu_m^{\text{cond}_j} + w_{\ell m} X_{(j-1)K+\ell}{\mu'}_m^{\text{cond}_j}\right) \\
& + 2\sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1}\left(w_{\ell m}{\mu'}_\ell^{\text{cond}_j}\mu_m^{\text{cond}_j} + w'_{\ell m}\mu_\ell^{\text{cond}_j}\mu_m^{\text{cond}_j}\right)
\end{aligned} \tag{42}$$

while the second derivative is

$$\begin{aligned}
\partial^2_\phi p_{\text{Hybrid}}(\phi) = & \partial^2_\phi g(\phi) + \frac{1}{K^2}\sum_{j=1}^{B}\sum_{i=1}^{B}\sum_{\ell=1}^{K}\sum_{m=1}^{K} v''_{ij} X_{(i-1)K+\ell} X_{(j-1)K+m} \\
& + \sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1} w''_{\ell m} X_{(j-1)K+\ell} X_{(j-1)K+m} \\
& - 2\sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1}(w''_{\ell m} X_{(j-1)K+\ell}\mu_m^{\text{cond}_j} + w_{\ell m} X_{(j-1)K+\ell}{\mu''}_m^{\text{cond}_j}) \\
& + 4\sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1}(w'_{\ell m} X_{(j-1)K+\ell}{\mu'}_m^{\text{cond}_j} + w'_{\ell m}{\mu'}_\ell^{\text{cond}_j}\mu_m^{\text{cond}_j}) \\
& + 2\sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1}(w_{\ell m}{\mu'}_\ell^{\text{cond}_j}{\mu'}_m^{\text{cond}_j} + w_{\ell m}\mu_\ell^{\text{cond}_j}{\mu''}_m^{\text{cond}_j}) \\
& + \sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1} w''_{\ell m}\mu_\ell^{\text{cond}_j}\mu_m^{\text{cond}_j}.
\end{aligned}$$

Further notational simplifications are possible if we recall that

$$\mu_\ell^{\text{cond}_j} = \tau^*_\ell \sum_{p=1}^{K} X_{(j-1)K+p} \quad \text{where } \tau^*_\ell = \frac{\tau_\ell}{K\eta}$$

and we rewrite all the terms containing conditional means in the above sums as, for example, the last one:

$$\sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1} w''_{\ell m}\mu_{\ell}^{\text{cond}_j}\mu_{m}^{\text{cond}_j} = \sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1}\sum_{p=1}^{K}\sum_{q=1}^{K} w''_{\ell m}\tau^*_{\ell}\tau^*_{m}X_{(j-1)K+p}X_{(j-1)K+q}.$$

Since for an AR(1) time series we have that $E[X_tX_{t'}] = \frac{\phi^{|t-t'|}}{1-\phi^2}\sigma_\varepsilon^2$, it follows that the expected value of the second derivative of $p_{\text{Hybrid}}(\phi)$ is now expressed as a function of ϕ and the data only. We shall refer to it as $P_2(\phi) \equiv E[\partial^2_\phi p_{\text{Hybrid}}(\phi)]$.

Further on, using the expansion (13) in (42), we find

$$\partial_\phi p_{\text{Hybrid}}(\phi) = \partial_\phi g(\phi) + \sum_{\{r,s:r\leqslant s\}} a_{N,r,s}\xi_r\xi_s, \tag{43}$$

where

$$a_{N,r,s} = a^{(1)}_{N,r,s} + a^{(2)}_{N,r,s} + a^{(3)}_{N,r,s} + a^{(4)}_{N,r,s} \tag{44}$$

and

$$a^{(1)}_{N,r,s} = \sigma_\varepsilon^2\sum_{j=1}^{B}\sum_{i=1}^{B}\sum_{\ell=\eta}^{K}\sum_{m=v}^{K} v'_{ij}\phi^{(i+j-2)K+l+m-r-s},$$

$$a^{(2)}_{N,r,s} = \sigma_\varepsilon^2\sum_{j=1}^{B}\sum_{\ell=\lambda^*}^{K-1}\sum_{m=v^*}^{K-1} w'_{\ell m}\phi^{2(j-1)K+l+m-r-s},$$

$$a^{(3)}_{N,r,s} = -2\sigma_\varepsilon^2\sum_{j=1}^{B}\sum_{\ell=\lambda^*}^{K-1}\sum_{m=1}^{K-1}\sum_{p=v}^{K}(w'_{\ell m}\tau^*_m + w_{\ell m}\tau^{*\prime}_m)\phi^{2(j-1)K+l+p-r-s}$$

and

$$a^{(4)}_{N,r,s} = \sigma_\varepsilon^2\sum_{j=1}^{B}\sum_{\ell=1}^{K-1}\sum_{m=1}^{K-1}\sum_{p=\lambda}^{K}\sum_{q=v}^{K}(2w_{\ell m}\tau^{*\prime}_\ell\tau^*_m + w'_{\ell m}\tau^*_\ell\tau^*_m)\phi^{2(j-1)K+p+q-r-s}.$$

The summation lower bounds are defined as $\eta = f_1(i,r,K)$, $\lambda = f_1(j,r,K)$, $\lambda^* = f_2(j,r,K)$, $v = f_1(j,s,K)$, and $v^* = f_2(j,s,K)$ where f_1 is defined in (24) and f_2 is given by

$$f_2(j,r,K) = \begin{cases} 1 & \text{if } r-(j-1)K \leqslant 1,\\ r-(j-1)K & \text{if } 1 < r-(j-1)K \leqslant K-1,\\ K+1 & \text{if } r-(j-1)K \geqslant K. \end{cases} \tag{45}$$

The conditions stated by Theorem 2 are satisfied, and we obtain that

$$P_1(\phi) \equiv \text{Var}[\partial_\phi p_{\text{Hybrid}}(\phi)] = 2\sum_r a^2_{N,r,r} + \sum_{\{r,s:r<s\}} a^2_{N,r,s}.$$

Therefore, according to the information sandwich technique, we compute the asymptotic variance of the hybrid estimator as

$$\text{Var}[\hat{\phi}_3] = P_2^{-1}(\phi)P_1(\phi)P_2^{-1}(\phi). \tag{46}$$

Table 3
Time series: hybrid asymptotic relative efficiency

ϕ	$B = 5, K = 100$		$B = 10, K = 50$		$B = 50, K = 10$	
	Theory	Sim.	Theory	Sim.	Theory	Sim.
−0.750	0.99953	0.999	0.99665	0.996	0.92267	0.943
−0.250	0.99802	0.998	0.97725	0.977	0.91373	0.921
−0.010	0.99495	0.995	0.97028	0.971	0.89989	0.898
0.010	0.99457	0.995	0.97033	0.970	0.89739	0.897
0.250	0.99203	0.992	0.97385	0.972	0.91409	0.903
0.750	0.99134	0.991	0.98952	0.990	0.91800	0.922

Table 3 presents the numerical results obtained for various values of ϕ, under the columns labeled "Theory''. Calculations are performed for an AR(1) time series of length 500, divided into various number of blocks (5, 10 and 50), and several values of the autoregressive parameter ϕ. Under the columns labeled "Sim.'', we present results obtained through a simulation study based on 1000 iterations. We note immediately that the simulated results concur with the theoretical calculations, modulo the simulation induced error.

It is remarkable that, as illustrated by Table 3, the hybrid estimator is very efficient asymptotically, when compared to the maximum likelihood estimator. There is a slight decrease in efficiency when block sizes are small. Nevertheless, the relative efficiency remains above 90% even in the least favorable cases, and this seems to be consistently true across the full range of ϕ.

3.5. Concluding remarks regarding the three proposed estimators

As depicted in Tables 1–3, among the three estimators suggested by Caragea and Smith [8] the big blocks estimator has the lowest asymptotic relative efficiency, in spite of achieving the greatest reduction in computation time. Its inferiority is a consequence of the fact that this method ignores the inter-block correlations.

The surprising element of this analysis is that the small blocks estimator seems to perform as well as the hybrid estimator. Both theoretical and simulation based results indicate that, for example, the hybrid estimator is more efficient for larger absolute values of the true autoregressive parameter (which is due to the fact that this method takes into account the inter-block correlation, which is stronger for large values of ϕ). The small blocks estimator is more efficient for values of ϕ closer to 0, but the difference between the asymptotic performance of the two estimators is not significant.

Although results are not explicitly shown in the paper, we have also taken a close look at the bias of the three estimators for the situations illustrated by the simulation study. We calculated the bias as the difference between the true value ϕ which was used to generate the Monte Carlo realizations and the average of the estimated parameter values (over all simulations). It appears that the bias for the big blocks estimator is much larger than that of the small blocks and hybrid estimators. The bias for the former seems to increase as the size of the blocks increases, which is intuitive (the larger the block size, the more local information is lost by aggregation) while the bias for the latter remains very close to zero throughout the various block sizes considered here (and basically indistinguishable from the bias associated with the MLE).

The application of the expansion method to the one-dimensional context of the first-order autoregressive process provides valuable insight. It confirms that the big blocks approach is not

recommended on its own except for certain situations, but it is one of the main components of the hybrid method. The structure of the coefficients in the expansions of the derivatives of the pseudo-likelihood functions as quadratic sums of independent normal random variables allows generalization to higher-dimensional setups, for which we provide an illustration in the following section.

4. Application of the expansion method to two-dimensional autoregressive processes on a lattice

The purpose of this section is to illustrate how the methods of the paper can be extended to a higher-dimensional process. Once again, the calculations that the method involves are highly intricate, and for this reason, we restrict our detailed calculations to a single simple model. Our motivation and justification for doing this is that by calculating asymptotic efficiencies for this example, we can suggest some general guidelines for comparisons among the three methods, that should be applicable to more general classes of spatial processes.

4.1. General description of the two-dimensional AR(1) process and the maximum likelihood estimator

Consider a two-dimensional process X_{ij} on a $N_1 \times N_2$ lattice which is assumed to be the composition of two AR(1) time series (one on each of the two directions defining the lattice) with the same autoregressive parameter, ϕ. It follows that the covariance structure is given by the Kronecker product of the one-dimensional covariances:

$$\mathrm{Cov}[X_{ij}, X_{t\ell}] = \gamma_{it}^{(1)}\gamma_{j\ell}^{(2)} = \sigma_X^2\phi^{|i-t|+|j-\ell|}, \tag{47}$$

where $|\phi| < 1$ to ensure stationarity. Thus, we can represent the spatial process X_{ij} as

$$X_{ij} - \phi(X_{i+1,j} + X_{i,j+1}) + \phi^2 X_{i+1,j+1} = \varepsilon_{ij}, \tag{48}$$

where $\varepsilon_{ij} = \sigma_X(1-\phi^2)\xi_{ij}$ are independent $\mathcal{N}(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 = \sigma_X^2(1-\phi^2)^2$. Alternatively, we write the process in the form

$$X_{ij} = \sigma_\varepsilon \sum_{r=-\infty}^{i}\sum_{s=-\infty}^{j} \phi^{i+j-r-s}\xi_{i,j}. \tag{49}$$

Note that the processes we have defined here lie within the general class of spatial processes on lattices first defined by Whittle [29].

The calculations leading to the maximum likelihood estimator in this case rely on the joint normal distribution of the observations $\{X_{ij}, 1 \leqslant i \leqslant N_1, 1 \leqslant j \leqslant N_2\}$, with mean 0 and covariance given by (47) (assume σ_ε^2 is known). Using the Kronecker product notation, the covariance matrix of this process is $U_{N_1} \otimes U_{N_2}$, where U_{N_i} for any N_i is given by (14). It follows that the inverse covariance matrix has the form $U_{N_1}^{-1} \otimes U_{N_2}^{-1}$, with $U_{N_i}^{-1}$ (for any N_i) with entries u_{ij}^* as given by (16). Then the negative log-likelihood function, modulo fixed constants, is given by

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{N_1}\sum_{\ell=1}^{N_2} X_{ij}X_{t\ell}u_{ij}^*u_{t\ell}^* - \log|U_{N_1}^{-1} \otimes U_{N_2}^{-1}|. \tag{50}$$

4.2. Alternative estimators

In this section we consider the spatial counterpart of the one-dimensional grouping introduced before: assume that we divide the $N_1 \times N_2$ locations on the lattice into $B_1 \times B_2$ disjoint subregions, each consisting of $K_1 \times K_2$ locations. That is, $N_1 = B_1 \times K_1$ and $N_2 = B_2 \times K_2$.

In each of the three cases examined in this section, derivations of the asymptotic variances use the information sandwich technique. To calculate the expected value of the second derivative of the pseudo-log-likelihood function, we exploit the properties of the underlying AR(1) covariance structure, in particular (47), as we have for the one-dimensional setup.

To calculate the variance of the first derivative, we first expand it, using (49), as a sum of quadratic forms of independent normal random variables:

$$\sum_{\{r_1,s_1,r_2,s_2:(r_1,r_2)<(s_1,s_2)\}} a_{N,r_1,r_2,s_1,s_2}\xi_{r_1r_2}\xi_{s_1s_2}. \tag{51}$$

Once the coefficients for the quadratic forms are identified, we apply Theorem 2 to calculate the mean and the variance of the gradient as

$$m_N = \sum_{\{r_1,r_2\}} a_{N,r_1,r_1,r_2,r_2}$$

and

$$v_N = 2\sum_{\{r_1,r_2\}} a^2_{N,r_1,r_1,r_2,r_2} + \sum_{\{r_1,s_1,r_2,s_2:(r_1,r_2)<(s_1,s_2)\}} a^2_{N,r_1,r_2,s_1,s_2}. \tag{52}$$

Since the general methodology of deriving the asymptotic properties of the alternative estimators is similar to the one-dimensional case, in the subsequent sections we give only the form of the coefficients of the quadratic forms, a_{N,r_1,r_2,s_1,s_2}, for each of the three situations considered in this paper, in the spatial context.

4.2.1. Big blocks estimator

This estimator is the spatial analog of the one described in Section 3.2. The big blocks pseudo-likelihood function is defined as the likelihood of the subregional means:

$$\begin{aligned} p_{\text{means}}(\phi) = {} & \frac{1}{K_1^2K_2^2}\sum_{i_1=1}^{B_1}\sum_{i_2=1}^{B_2}\sum_{j_1=1}^{B_1}\sum_{j_2=1}^{B_2} v_{(i_1-1)B_2+i_2,(j_1-1)B_2+j_2} \\ & \times\sum_{\ell_1=1}^{K_1}\sum_{\ell_2=1}^{K_2}\sum_{m_1=1}^{K_1}\sum_{m_2=1}^{K_2} X_{(i_1-1)K_1+\ell_1,(i_2-1)K_2+\ell_2}X_{(j_1-1)K_1+m_1,(j_2-1)K_2+m_2} \\ & + \log(|V_{\text{means}}|), \end{aligned} \tag{53}$$

where V_{means} with entries v_{ij} denotes the $B_1B_2 \times B_1B_2$ covariance matrix of the regional means, the derivative of its inverse is denoted by $\partial_\phi V^{-1}_{\text{means}}$ and has entries v'_{ij}. Using (48) we expand the first derivative with respect to ϕ of the function in (53) as a sum of quadratic forms of i.i.d. normal random variables. The coefficients of this expansion are given by

$$\begin{aligned} a^{(1)}_{N,r_1,s_1,r_2,s_2} = {} & \sigma_\varepsilon^4\sum_{i_1=1}^{B_1}\sum_{i_2=1}^{B_2}\sum_{j_1=1}^{B_1}\sum_{j_2=1}^{B_2}\sum_{\ell_1=\eta_1}^{K_1}\sum_{\ell_2=\eta_2}^{K_2}\sum_{m_1=\nu_1}^{K_1}\sum_{m_2=\nu_2}^{K_2} v'_{(i_1-1)B_2+i_2,(j_1-1)B_2+j_2} \\ & \times\phi^{(i_1+j_1-2)K_1+(i_2+j_2-2)K_2+\ell_1+\ell_2+m_1+m_2-r_1-s_1-r_2-s_2}, \end{aligned} \tag{54}$$

and we apply Theorem 2 to obtain the variance of the first derivative of the pseudo-likelihood function. The calculation of the expected value of the second derivative of (53) is using extensively the covariance structure of the AR(1) process.

The expression for the asymptotic variance was calculated numerically for several values of ϕ and two lattice configurations: a 32×32 lattice with ($B_1 = B_2 = 8$, $K_1 = K_2 = 4$) and a 27×27 lattice, with ($B_1 = B_2 = 9$, $K_1 = K_2 = 3$). These results are presented in Tables 4 and 5. We note that the big blocks estimator is relatively inefficient compared with the MLE, but its efficiency seems to increase while block sizes decrease. These observations are consistent with what was noted in the one-dimensional setup.

4.2.2. Small blocks estimator

This subsection assumes the same setting as for the big blocks, with the additional assumption that the $B_1 \times B_2$ blocks are independent. As in Section 3.3, the relatively simpler structure of the pseudo-likelihood function enables us to study more closely some of the theoretical aspects regarding the asymptotic distribution of the small blocks estimator. If we denote by $U_{K_1K_2}$ the $K_1K_2 \times K_1K_2$ covariance matrix corresponding to any block, whose inverse has entries u^*_{ij} given by (16), it follows from Eq. (50) that the negative pseudo-log-likelihood function, modulo fixed constants, is given by

$$\begin{aligned} p_{\text{Small Blocks}}(\phi) = &\sum_{j_1=1}^{B_1}\sum_{j_2=1}^{B_2}\sum_{\ell_1=1}^{K_1}\sum_{\ell_2=1}^{K_2}\sum_{m_1=1}^{K_1}\sum_{m_2=1}^{K_2} u^*_{(m_1-1)K_2+\ell_1,(m_2-1)K_2+\ell_2} \\ &\times X_{(j_1-1)K_1+\ell_1,(j_2-1)K_2+\ell_2} X_{(j_1-1)K_1+m_1,(j_2-1)K_2+m_2} \\ &+(B_1+B_2)\log|U_{K_1K_2}|. \end{aligned} \tag{55}$$

Note that the function in Eq. (55) is a degree four polynomial in the unknown parameter ϕ (this is a direct consequence of the correlation structure for the two-dimensional AR(1) process). It is straightforward to check that the first two conditions stated by Amemiya [2] to ensure consistency of a local maximum are satisfied here (regarding the parameter space, measurability on the entire parameter space and continuity in an open neighborhood of the true value of the parameter). Also, as stated in Section 2.2, for the last condition to be satisfied, we need to have a bounded expectation of the first-order derivative in a neighborhood of the true parameter value. This condition is satisfied by the function in Eq. (55), therefore we conclude that the two-dimensional small blocks estimator is consistent.

The rest of the calculations follow the expansion technique ideas. Using (48) we expand (55) as a sum of quadratic forms of independent normal random variables of the form (51). If we denote by $u^{*\prime}_{ij}$ the (i, j)th entry of $\partial_\phi U^{-1}_{K_1K_2}$ the corresponding coefficients of the expansion are given by

$$\begin{aligned} a_{N,r_1,s_1,r_2,s_2} = &\sigma_\varepsilon^4 \sum_{j_1=1}^{B_1}\sum_{j_2=1}^{B_2}\sum_{\ell_1=\lambda_1}^{K_1}\sum_{\ell_2=\lambda_2}^{K_2}\sum_{m_1=\nu_1}^{K_1}\sum_{m_2=\nu_2}^{K_2} u^{*\prime}_{(\ell_1-1)K_2+\ell_2,(m_1-1)K_2+m_2} \\ &\times \phi^{2(j_1-1)K_1+2(j_2-1)K_2+\ell_1+\ell_2+m_1+m_2-r_1-s_1-r_2-s_2}. \end{aligned} \tag{56}$$

Here the lower bounds for summation are given by $\lambda_1 = f_1(j_1, r_1, K_1)$, $\lambda_2 = f_1(j_2, r_2, K_2)$, $\nu_1 = f_1(j_1, s_1, K_1)$, and $\nu_2 = f_1(j_2, s_2, K_2)$, where f_1 and f_2 are defined by (24) and (45).

The asymptotic variance was calculated numerically for several values of ϕ, and two lattice configurations, as shown in Tables 4 and 5. It appears that the small blocks estimator performs

very well when compared to the classical maximum likelihood estimator. The loss in efficiency is remarkably low, even for configurations with small block sizes, which is the most unfavorable situation. Given the gain in the computational time achieved when using this estimation method, we conclude that the small blocks estimator is a good alternative to the MLE for large data sets.

4.2.3. Hybrid estimator

In this subsection we consider the same setting described in the previous two, but concentrate on the hybrid estimator, the two-dimensional analog of the one described in detail in Section 3.4. The covariance structure under the conditional independence assumption is much more complicated than in Section 4.2.2. As a consequence, the identification of the coefficients is more involved for this case, which is why we omit the technical details and only list them here in the form that they are used by the information sandwich technique. Notation is just a generalization of what we defined in the one-dimensional setting (Section 3.4):

$$a_{N,r_1,s_1,r_2,s_2} = a^{(1)}_{N,r_1,s_1,r_2,s_2} + a^{(2)}_{N,r_1,s_1,r_2,s_2} + a^{(3)}_{N,r_1,s_1,r_2,s_2} + a^{(4)}_{N,r_1,s_1,r_2,s_2},$$

where $a^{(1)}_{N,r_1,s_1,r_2,s_2}$ is given in expression (54) and

$$a^{(2)}_{N,r_1,s_1,r_2,s_2} = \sigma_\varepsilon^4 \sum_{j_1=1}^{B_1}\sum_{j_2=1}^{B_2}\sum_{\ell_1=\lambda_1}^{K_1}\sum_{\ell_2=\lambda_2^*}^{K_2^{\ell_1}}\sum_{m_1=v_1}^{K_1}\sum_{m_2=v_2^*}^{K_2^{m_1}} w'_{(\ell_1-1)K_2+\ell_2,(m_1-1)K_2+m_2} \times \phi^{2(j_1-1)K_1+2(j_2-1)K_2+\ell_1+m_1+\ell_2+m_2-r_1-r_2-s_1-s_2},$$

$$\begin{aligned} a^{(3)}_{N,r_1,s_1,r_2,s_2} &= \sigma_\varepsilon^4 \sum_{j_1=1}^{B_1}\sum_{j_2=1}^{B_2}\sum_{\ell_1=\lambda_1}^{K_1}\sum_{\ell_2=\lambda_2^*}^{K_2^{\ell_1}}\sum_{m_1=1}^{K_1}\sum_{m_2=1}^{K_2^{m_1}}\sum_{p_1=v_1}^{K_1}\sum_{p_2=v_2}^{K_2} \Big[w'_{(\ell_1-1)K_2+\ell_2,(m_1-1)K_2+m_2}\tau^*_{(m_1-1)K_2+m_2} \\ &\quad + w_{(\ell_1-1)K_2+\ell_2,(m_1-1)K_2+m_2}\tau^{*\prime}_{(m_1-1)K_2+m_2}\Big] \\ &\quad \times \phi^{2(j_1-1)K_1+2(j_2-1)K_2+\ell_1+p_1+\ell_2+p_2-r_1-r_2-s_1-s_2}, \end{aligned}$$

and

$$\begin{aligned} a^{(4)}_{N,r_1,s_1,r_2,s_2} = \sigma_\varepsilon^4 &\sum_{j_1=1}^{B_1}\sum_{j_2=1}^{B_2}\sum_{\ell_1=\lambda_1}^{K_1}\sum_{\ell_2=\lambda_2^*}^{K_2^{\ell_1}}\sum_{m_1=v_1}^{K_1}\sum_{m_2=1}^{K_2^{m_1}}\sum_{p_1=\lambda_1}^{K_1}\sum_{p_2=\lambda_2}^{K_2}\sum_{q_1=v_1}^{K_1}\sum_{q_2=v_2}^{K_2} \\ &\Big[2w_{(\ell_1-1)K_2+\ell_2,(m_1-1)K_2+m_2}\tau^{*\prime}_{(\ell_1-1)K_2+\ell_2}\tau^*_{(m_1-1)K_2+m_2} \\ &+ w'_{(\ell_1-1)K_2+\ell_2,(m_1-1)K_2+m_2}\tau^*_{(\ell_1-1)K_2+\ell_2}\tau^*_{(m_1-1)K_2+m_2}\Big] \\ &\phi^{2(j_1-1)K_1+2(j_2-1)K_2+p_1+q_1+p_2+q_2-r_1-r_2-s_1-s_2}. \end{aligned}$$

Here the summation lower bounds are given by $\eta_1 = f_1(i_1, r_1, K_1)$, $\eta_2 = f_1(i_2, r_2, K_2)$, $\lambda_1 = f_1(j_1, r_1, K_1)$, $\lambda_2 = f_1(j_2, r_2, K_2)$, $v_1 = f_1(j_1, s_1, K_1)$, $v_2 = f_1(j_2, s_2, K_2)$, $\lambda_2^* = f_3(j_2, r_2, K_2^{\ell_2})$ and $v_2^* = f_3(j_2, s_2, K_2^{m_2})$; functions f_1 and f_2 are defined by (24) and (45), f_3 is defined as

$$f_3(r, j, k) = \begin{cases} f_1(r, j, k) & \text{if } k = K_2, \\ f_2(r, j, k) & \text{if } k = K_2 - 1, \end{cases} \tag{57}$$

Table 4
Relative efficiency of estimators for the AR(1) × AR(1) model with $B_1 = B_2 = 8$, $K_1 = K_2 = 4$

θ	Big blocks efficiency	Small blocks efficiency	Hybrid efficiency
−0.750	0.00483	0.80386	0.80357
−0.500	0.03378	0.81132	0.81286
−0.250	0.10160	0.77615	0.78069
−0.010	0.19665	0.75005	0.75482
0.010	0.20543	0.75005	0.75468
0.250	0.31412	0.77615	0.77780
0.500	0.42135	0.81132	0.80970
0.750	0.49732	0.80386	0.80803

Table 5
Relative efficiency of estimators for the AR(1) × $AR(1)$ model with $B_1 = B_2 = 9$, $K_1 = K_2 = 3$

θ	Big blocks efficiency	Small blocks efficiency	Hybrid efficiency
−0.750	0.25675	0.76861	0.79230
−0.500	0.04912	0.75964	0.76768
−0.250	0.12744	0.70510	0.71818
−0.010	0.26449	0.66674	0.68148
0.010	0.27570	0.66674	0.68129
0.250	0.39986	0.70510	0.71302
0.500	0.50212	0.75964	0.75779
0.750	0.57861	0.76861	0.77155

while the summation upper bounds are defined as

$$K_2^{\ell_1} = \begin{cases} K_2 & \text{if } \ell_1 < K_1, \\ K_2 - 1 & \text{if } \ell_1 = K_1 \end{cases} \quad \text{and} \quad K_2^{m_1} = \begin{cases} K_2 & \text{if } m_1 < K_1, \\ K_2 - 1 & \text{if } m_1 = K_1. \end{cases} \tag{58}$$

The asymptotic relative efficiency was calculated numerically for various values of ϕ and two lattice configurations. The results are presented in Tables 4 and 5. We conclude that, as for the one-dimensional setup, the hybrid estimator is asymptotically highly efficient relative to the maximum likelihood estimator. This, together with the significant reduction of the computational time, recommends it as a good estimation alternative for high-dimensional data sets.

4.3. Conclusions

From the numerical results displayed in Tables 4 and 5 we can conclude that the performance of the simplest of the estimators, the big blocks, is inferior to that of the other two estimators. Its asymptotic efficiency relative to the MLE increases as ϕ gets larger. This is caused by the dependence between farther observations being stronger in these situations, which is a feature that the big blocks method is designed to capture. The other two estimators, the small blocks and hybrid are performing very well when compared to the maximum likelihood estimator (efficiencies ranging between 68% and 80%), with no clear choice between the two. Numerical studies by Caragea and Smith [8] have suggested that there are some situations (spatial models where the range parameter is very large) in which the big blocks method is the best of the three. They have also provided further numerical comparisons of the small blocks and hybrid methods that suggest that the qualitative conclusions of the present theoretical study hold for a much wider range of

spatial models. Caragea and Smith [8] have found the hybrid estimator to be clearly superior for some settings (for example, the case of the Matérn model with small shape parameter). Another advantage of the hybrid estimator, as illustrated by Caragea and Smith [8], is that the estimated standard errors from inverting the approximate observed information matrix are closer to the true standard errors than those derived by the small blocks method.

5. Further discussion and concluding remarks

This paper was intended to take a closer look at the asymptotic properties of the alternative estimators proposed by Caragea and Smith [8]. Since calculations for the general multivariate normal processes are too complicated to permit the derivation of analytical formulae for the asymptotic variances, we considered two particular cases: the autoregressive process of first order in one and two dimensions. Although the covariance structure was much simpler in these two instances than for the general spatial processes that motivated the development of the alternative estimators, derivation of closed form, easy to manipulate formulae for the asymptotic variances was not possible. Instead, we produced and analyzed numerical results for our theoretical calculations for several arbitrarily chosen situations. From these considerations (Tables 1–5) we conclude that the big blocks estimator lacks in asymptotic efficiency, in spite of the reduction in the computational effort, in almost all circumstances. The other two estimators perform very well asymptotically.

We emphasize here that the examples illustrated by Tables 4 and 5 consider only moderate size lattices. Even though for the specific model under consideration there are already several known ways of approximating the likelihood efficiently (see [29,13]) or of simplifying calculations of the exact likelihood (see [30]), for the most general models without an exploitable structure, using the Cholesky decomposition is the most efficient method available to calculate the determinant and inverse of the covariance matrix. This bears the implication that using the maximum likelihood estimation method for a grid of 32×32 locations is on the edge of what is feasible on a single processor desktop computer. Different considerations might apply in totally different computing environments (like parallel computing). However, these calculations could be, in principle, performed for a much larger number of locations.

The greatest shortcoming associated with these theoretical calculations is that they rely heavily on the assumption that the first derivative of the pseudo-likelihood function can be expanded as a sum of quadratic forms of independent normal variables, which is not immediate for much more general spatial processes. However, the conclusions drawn here should remain valid for more complex processes, at least in the sense of recommending the hybrid estimator as a reasonable competitor to the maximum likelihood estimator, for large spatial data sets.

References

[1] M. Akahira, K. Takeuchi, Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency, Lecture Notes in Statistics, Springer, New York, 1981.
[2] T. Amemiya, Advanced Econometrics, Harvard University Press, Cambridge, MA, 1985.
[3] S. Banerjee, B.P. Carlin, A.E. Gelfand, Hierarchichal Modeling and Analysis for Spatial Data, CRC Press, Chapman & Hall, Boca Raton, FL, 2003.
[4] R.P. Barry, R.K. Pace, Monte Carlo estimates of the log determinant of large sparse matrices, Linear Algebra Statist. Linear Algebra Appl. 289 (1999) 41–54.
[5] S.D. Billings, R.K. Beatson, G.N. Newsam, Interpolation of geophysical data by continuous global surfaces, Geophysics 67 (2002) 1810–1822.
[6] P. Billingsley, Probability and Measure, third ed., Wiley, New York, 1995.
[7] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, second ed., Springer, New York, 1991.

[8] P.C. Caragea, R.L. Smith, Approximate Likelihoods for Spatial Processes, 2006, Preprint Series #06–18, Department of Statistics, Iowa State University.
[9] J.-P. Chilès, P. Delfiner, Geostatistics: Modeling Spatial Uncertainty, Wiley, New York, 1999.
[10] N. Cressie, Statistics for Spatial Data, Wiley, New York, 1993.
[11] F.C. Curriero, S.R. Lele, A composite likelihood approach to semivariogram estimation, J. Agricultural Biol. Environ. Statist. 4 (1) (1999) 9–28.
[12] R.M. Furrer, M. Genton, D. Nychka, Covariance tapering for interpolation of large spatial datasets, J. Comput. Graphical Statist. 15 (3) (2006) 502–523.
[13] X. Guyon, Parameter estimation for a stationary process on a d-dimensional lattice, Biometrika 69 (1982) 95–105.
[14] P.J. Heagerty, S.R. Lele, A composite likelihood approach to binary data in space, J. Amer. Statist. Assoc. 92 (1998) 846–854.
[15] H.-C. Huang, N. Cressie, J. Gabrosek, Fast, resolution-consistent spatial prediction of global processes from satellite data, J. Comput. Graphical Statist. 11 (2002) 63–88.
[16] G. Johannesson, N. Cressie, Finding large-scale spatial trends in massive, global, environmental datasets, Environmetrics 15 (2004) 1–44.
[17] K.Y. Liang, S.L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.
[18] B.G. Lindsay, Composite likelihood methods, Contemp. Math. 80 (1988) 221–239.
[19] D. Nychka, Spatial process estimates as smoothers, smoothing and regression, in: M.G. Schimek (Ed.), Approaches, Computation and Application, Wiley, New York, 2000, pp. 393–424.
[20] D. Nychka, C. Wikle, J.A. Royle, Multiresolution models for nonstationary spatial covariance functions, Statist. Modeling 2 (2002) 315–331.
[21] O. Smirnov, L. Anselin, Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach, Comput. Statist. Data Anal. 35 (3) (2001) 301–319.
[22] M.L. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Springer, New York, 1999.
[23] M.L. Stein, Z. Chi, L.J. Welty, Approximating likelihoods for large spatial data sets, J. Roy. Statist. Soc. B 66 (2004) 275–296.
[24] W.F. Trench, An algorithm for the inversion of finite Toeplitz matrices, J. Soc. Indust. Appl. Math. 12 (1964) 515–522.
[25] S. Tzeng, H.C. Huang, N. Cressie, A fast, optimal spatial-prediction method for massive datasets, J. Amer. Statist. Assoc. 100 (2005) 1343–1357.
[26] A.W. Van der Vaart, Asymptotic Statistics, Cambridge University Press, Cambridge, 1998.
[27] A.V. Vecchia, Estimation and identification for continuous spatial processes, J. Roy. Statist. B 50 (1988) 297–312.
[28] H. White, Maximum likelihood estimation of misspecified models, Econometrica 50 (1982) 1–26.
[29] P. Whittle, On stationary processes in the plane, Biometrika 41 (1954) 434–449.
[30] D.L. Zimmerman, Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models, J. Statist. Comput. Simulation 32 (1989) 1–15.