Gaussian Maximum Likelihood Estimation For ARMA Models II: Spatial Processes

Qiwei Yao*

Department of Statistics, London School of Economics, London, WC2A 2AE, UK

Guanghua School of Management, Peking University, China

Email: q.yao@lse.ac.uk

Peter J. Brockwell

Department of Statistics, Colorado State University Fort Collins, CO 80532, USA

Email: pjbrock@stat.colostate.edu

Summary

This paper examines the Gaussian maximum likelihood estimator (GMLE) in the context of a general form of spatial autoregressive and moving average (ARMA) processes with finite second moment. The ARMA processes are supposed to be causal and invertible under the half-plane unilateral order (Whittle 1954), but not necessarily Gaussian. We show that the GMLE is consistent. Subject to a modification to confine the edge effect, it is also asymptotically distribution-free in the sense that the limit distribution is normal, unbiased and with a variance depending on the autocorrelation function only. This is an analogue of Hannan's classic result for time series in the context of spatial processes; see Theorem 10.8.2 of Brockwell and Davis (1991).

Keywords: ARMA spatial process, asymptotic normality, consistency, edge effect, Gaussian maximum likelihood estimator, martingale-difference.

Running title: Gaussian MLE for spatial ARMA processes

1. Introduction

Since Whittle's pioneering work (Whittle 1954) on stationary spatial processes, the frequency-domain methods which approximate a Gaussian likelihood by a function of a spectral density became popular, while the Gaussian likelihood function itself was regarded intractable in both theoretical exploration and practical implementation. Guyon (1982) and Dahlhaus and Künsch (1987) established the asymptotic normality for the modified Whittle's maximum likelihood estimators for stationary spatial processes which are not necessarily Gaussian; the modifications were adopted to control the edge effect. On the other hand, the development of time-domain methods was dominated by the seminal work Besag (1974) who put forward an ingenious auto-normal specification based on a conditional probability argument. Besag's proposal effectively specifies the inverse covariance matrix of a Gaussian process, in which the parameters are interpreted in terms of conditional expectations.

In this paper we examine the estimator derived from maximising the Gaussian likelihood function for spatial processes, which we refer to as Gaussian maximum likelihood estimator (GMLE). To study its asymptotic properties, we assume that the data are generated from a spatial autoregressive and moving average (ARMA) model defined on a lattice. Under the condition that the process is causal and invertible according to the half-plane unilateral order (Whittle 1954), the GMLE is consistent (Theorem 1 in §3 below). Subject to a modification to confine the edge effect, it is also asymptotically normal and unbiased with a variance depending on the autocorrelation function only. Thus our modified GMLE is asymptotically distribution-free. The asymptotic normality presented in Theorem 2 below may be viewed as an analogue of Hannan's (1973) classic result for time series in the context of spatial processes, which shows that the limit distribution of the estimator for an ARMA process is determined by two AR models defined by the AR and the MA forms in the original model; see Theorem 2 in §4 below and also §8.8 of Brockwell and Davis (1991). Hannan's proof was based on a frequency-domain argument. He proved the equivalence of a Gaussian MLE and a Whittle's estimator and established the asymptotic normality for the latter; see also §10.8 of Brockwell and Davis (1991). Our proof largely follows the time-domain approach of Yao and Brockwell (2005), although the proposed modified GMLE shares the same asymptotic distribution as the modified Whittle's estimator proposed by Guyon (1982) (see Remark 3 below), which is also the asymptotic distribution of the modified Whittle's estimator proposed by Dahlhaus and Künsch (1986) if the underlying process is Gaussian. For purely autoregressive processes, our asymptotic results are the same as those derived by Tjøstheim (1978, 1983).

For a sample from a spatial model, the number of boundary points typically increases to infinity as the sample size goes to infinity. Therefore the edge effect causes problems. This is the feature which distinguishes high-dimensionally indexed processes from one-dimensional time series. Various modifications to reduce the edge effect have been proposed; see Guyon (1982), Dahlhaus and Künsch (1987) and §2.4 below. Both Guyon (1982) and Dahlhaus and Künsch (1987) adopted a frequency-domain approach, dealing with Whittle's estimators for stationary processes defined on a lattice. Our approach is within the time-domain, dealing with GMLE for the coefficients of ARMA models. Our edge effect modification can be readily performed along with the prewhitening (§2.3 below). By exploring the explicit form of these models, we are able to establish a central limit theorem (Lemma 9 in §4 below) based on an innate martingale structure. Therefore the regularity conditions imposed by Theorem 2 are considerably weaker than those in Guyon (1982) and Dahlhaus and Künsch (1987). For example, we only require the process to have finite second moments, and we do not impose any explicit assumptions on ergodicity and mixing. However it remains as an open problem whether the edge effect modification is essential for the asymptotic normality or not. See §5.1.

Although we only deal with the processes defined in the half-plane order explicitly, the asymptotic results may be derived for any unilaterally-ordered processes in the same manner. For the sake of simplicity, we only present the results for spatial processes with two-dimensional indices. The approach may be readily extended to higher-dimensional cases. In fact, such an extension is particularly appealing in the context of spatio-temporal modelling since a practically meaningful ARMA form can be easily formulated in that context. This is in marked contrast to the case of two-dimensional processes for which a unilateral ordering is often an artifact which limits the potential application. See §5.2 below.

The rest of the paper is organised as follows. In §2 we introduce spatial ARMA models and the conditions for causality and invertibility. The consistency and asymptotic normality will be established respectively in §3 and §4. We conclude with miscellaneous remarks in §5.

We denote by $|\mathbf{A}|$ the determinant of a square matrix $\mathbf{A}$, and by $||\mathbf{a}||$ the Euclidean norm of a vector $\mathbf{a}$.

2. Models and estimators

2.1. Stationary spatial ARMA processes

Let $\mathcal{Z}, \mathcal{R}, \mathcal{C}$ be the integer, the real number and the complex number spaces respectively. We always write $\mathbf{s} = (u, v) \in \mathcal{Z}^2$ and $\mathbf{i} = (j, k) \in \mathcal{Z}^2$. We define $\mathbf{s} > 0$ if either $u > 0$ or $u = 0$ and $v > 0$, and $\mathbf{s} = 0$ if and only if both u and v are 0. A unilateral order on a two-dimensional plane is defined as $\mathbf{s} >$ (or $\geq$) $\mathbf{i}$ if and only if $\mathbf{s} - \mathbf{i} >$ (or $\geq$) 0; see Whittle (1954). This order is often refereed as half plane order, or lexicographic order. Another popular unilateral ordering on a two-dimensional plane is the quarter plane order. Under the quarter plane order, $\mathbf{s} \geq 0$ if and only if both u and v are non-negative; see Guyon (1995). Although we do not discuss *explicitly* the models defined in terms of the quarter plane order in this paper, we will comment on its properties when appropriate.

We define a spatial ARMA model as

$$X(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_1} b_{\mathbf{i}} X(\mathbf{s} - \mathbf{i}) + \varepsilon(\mathbf{s}) + \sum_{\mathbf{i} \in \mathcal{I}_2} a_{\mathbf{i}} \varepsilon(\mathbf{s} - \mathbf{i}), \tag{2.1}$$

where $\{\varepsilon(\mathbf{s})\}$ is a white noise process in the sense that they are uncorrelated with constant first two moments 0 and σ^2 respectively, $\{b_{\mathbf{i}}\}$ and $\{a_{\mathbf{i}}\}$ are AR and MA coefficients, and both index sets $\mathcal{I}_1$ and $\mathcal{I}_2$ contain finite number of elements in the set $\{\mathbf{s} > 0\}$. In this paper, we consider real-valued processes only. Since we only require index sets $\mathcal{I}_1$ and $\mathcal{I}_2$ to be subsets of $\{\mathbf{s} > 0\}$, specification (2.1) includes both half-plane and quarter-plane ARMA models (Tjøstheim, 1978, 1983) as its special cases.

We introduce the back shift operator $\mathbf{B} \equiv (B_1, B_2)$ as follows:

$$\mathbf{B}^{\mathbf{i}} X(\mathbf{s}) \equiv B_1^j B_2^k X(u, v) = X(u - j, v - k) = X(\mathbf{s} - \mathbf{i}), \quad \mathbf{i} = (j, k) \in \mathcal{Z}^2.$$

For $\mathbf{z} \equiv (z_1, z_2)$, write $\mathbf{z}^{\mathbf{i}} = z_1^j z_2^k$. We define

$$b(\mathbf{z}) = 1 - \sum_{\mathbf{i} \in \mathcal{I}_1} b_{\mathbf{i}} \mathbf{z}^{\mathbf{i}} \quad \text{and} \quad a(\mathbf{z}) = 1 + \sum_{\mathbf{i} \in \mathcal{I}_2} a_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}. \tag{2.2}$$

Then model (2.1) can be written as

$$b(\mathbf{B}) X(\mathbf{s}) = a(\mathbf{B}) \varepsilon(\mathbf{s}). \tag{2.3}$$

It is well known that a bivariable polynomial can be factored into irreducible factors which are themselves bivariable polynomials but which cannot be further factored, and these irreducible polynomials are unique up to multiplicative constants. To avoid the ambiguity on the form of the model, we always assume that $b(\mathbf{z})$ and $a(\mathbf{z})$ are *mutually prime* in the sense that they do not have common irreducible factors although they may still have common roots (Goodman 1977, Huang and Anh 1992).

The process $\{X(\mathbf{s})\}$ defined in (2.1) is *causal* if it admits a purely MA representation

$$X(\mathbf{s}) = \varepsilon(\mathbf{s}) + \sum_{\mathbf{i}>0} \psi_{\mathbf{i}}\varepsilon(\mathbf{s}-\mathbf{i}) = \varepsilon(\mathbf{s}) + \sum_{k=1}^{\infty} \psi_{0k}\varepsilon(u, v-k) + \sum_{j=1}^{\infty}\sum_{k=-\infty}^{\infty} \psi_{jk}\varepsilon(u-j, v-k), \qquad (2.4)$$

where $\sum_{\mathbf{i}>0} |\psi_{\mathbf{i}}| < \infty$. It is easy to see that a causal $\{X_t\}$ is always weakly stationary with mean 0 and the autocovariance function

$$\begin{aligned}\gamma(\mathbf{i}) &= E\{X(\mathbf{s}+\mathbf{i})X(\mathbf{s})\} = \sigma^2 \sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty} \psi_{lm}\psi_{l+j,m+k} \\ &= \sigma^2\{\psi_{jk} + \sum_{m=1}^{\infty} \psi_{0m}\psi_{j,m+k} + \sum_{l=1}^{\infty}\sum_{m=-\infty}^{\infty} \psi_{lm}\psi_{l+j,m+k}\} \qquad (2.5)\end{aligned}$$

for $\mathbf{i} = (j,k)$ with $j \geq 1$, and $\gamma(-\mathbf{i}) = \gamma(\mathbf{i})$. In the above expression, $\psi_{00} = 1$ and $\psi_{0m} = 0$ for all $m < 0$. Furthermore, a causal process $\{X(\mathbf{s})\}$ is strictly stationary if $\{\varepsilon(\mathbf{s})\}$ are independent and identically distributed; see (2.4). The lemma below presents a sufficient condition for the causality.

Lemma 1. The process $\{X_t\}$ is causal if

$$b(\mathbf{z}) \neq 0 \text{ for all } |z_1| \leq 1 \text{ and } |z_2| = 1, \quad \text{and} \quad 1 - \sum_{(0,k)\in\mathcal{I}_1} b_{0k} z_2^k \neq 0 \text{ for all } |z_2| \leq 1, \qquad (2.6)$$

where $z_1, z_2 \in \mathcal{C}$. Furthermore, condition (2.6) implies that the coefficients $\{\psi_{jk}\}$ defined in (2.4) decay at an exponential rate, and in particular

$$|\psi_{jk}| \leq C\alpha^{j+|k|} \quad \text{for all } j \geq 0 \text{ and } k, \qquad (2.7)$$

for some constants $\alpha \in (0,1)$ and $C > 0$.

Note (2.7) improves Goodman (1977) which showed $\psi_{jk} = O(\alpha^j)$. Condition (2.6) is not symmetric in (z_1, z_2). This is due to the asymmetric nature of the half-plane order under which the causality is defined; see (2.4). The proof for the validity of (2.4) under condition (2.6) was given in Huang and Anh (1992); see also Justice and Shanks (1973), Strintzis(1977) and the references within. The inequality (2.7) follows from the simple argument as follows. Let $\psi(\mathbf{z}) = 1 + \sum_{\mathbf{i}>0} \psi_{\mathbf{i}}\mathbf{z}^{\mathbf{i}}$, where $\psi_{\mathbf{i}}'$s are given in (2.4). Then $\psi(\mathbf{z}) = a(\mathbf{z})/b(\mathbf{z})$. Due to the continuity of $b(\cdot)$, $b(\mathbf{z}) \neq 0$ for all $\mathbf{z} \in A_\epsilon \equiv \{(z_1, z_2) : 1-\epsilon < |z_j| < 1+\epsilon, j = 1, 2\}$ under condition (2.6), where $\epsilon > 0$ is a constant. Thus $\psi(\cdot)$ is bounded on A_ϵ, *i.e.* $|\sum_{\mathbf{i}>0} \psi_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}| < \infty$ for any $\mathbf{z} \in A_\epsilon$. Thus $\psi_{jk}\alpha^{-j}\alpha^{-|k|} \to 0$ as at least one of j and $|k| \to \infty$.

Remark 1. (i) Under condition (2.6), inequality (2.7) also holds if we replace ψ_{jk} by the derivative of ψ_{jk} with respect to $b_{\mathbf{i}}$ or $a_{\mathbf{i}}$. This can be justified by taking derivatives on both sides of equation $\psi(\mathbf{z}) = a(\mathbf{z})/b(\mathbf{z})$, followed by the same argument as above.

(ii) Condition (2.6) also ensures that the autocovariance function $\gamma(\cdot)$ decays at an exponential rate, *i.e.* $\gamma(j,k) = O(\alpha^{|j|+|k|})$ as at least one of $|j|$ and $|k| \to \infty$, where $\alpha \in (0,1)$ is a constant. To show this, note that for (j,k) with both j and k non-negative (other cases are similar), (2.5) can be written as

$$\begin{aligned}\gamma(j,k)/\sigma^2 &= \psi_{jk} + \sum_{m=1}^{\infty} \psi_{0m}\psi_{j,m+k} \\ &+ \sum_{l=1}^{\infty}\{\sum_{m=0}^{\infty} \psi_{lm}\psi_{l+j,m+k} + \sum_{m=0}^{\infty} \psi_{l,-k-m}\psi_{l+j,-m} + \sum_{m=1}^{k-1} \psi_{l,-m}\psi_{l+j,k-m}\}.\end{aligned}$$

By (2.7), all the sums on the RHS of the above expression are of the order α^{j+k}.

(iii) A partial derivative of $\gamma(\cdot)$ with respect to $b_{\mathbf{i}}$ or $a_{\mathbf{i}}$ also decays at an exponentially rate. This may be seen through combining (i) and the argument in (ii) together.

(iv) Condition (2.6) is not necessary for the causality, which is characteristically different from the case for one-dimensional time series; see Goodman (1977). On the other hand, a spatial ARMA process defined in term of the quarter plane order is causal if $b(\mathbf{z}) \neq 0$ for all $|z_1| \leq 1$ and $|z_2| \leq 1$ (Justice and Shanks 1973). Under this condition, the autocovariance function, the coefficients in an MA(∞) representation, and their derivatives decay exponentially fast.

(v) The process $\{X_t\}$ is *invertible* if it admits a purely AR representation

$$X(\mathbf{s}) = \varepsilon(\mathbf{s}) + \sum_{\mathbf{i}>0} \varphi_{\mathbf{i}} X(\mathbf{s}-\mathbf{i}) = \varepsilon(\mathbf{s}) + \sum_{k=1}^{\infty} \varphi_{0k} X(u, v-k) + \sum_{j=1}^{\infty}\sum_{k=-\infty}^{\infty} \varphi_{jk} X(u-j, v-k), \quad (2.8)$$

where $\sum_{\mathbf{i}>0} |\varphi_{\mathbf{i}}| < \infty$. It is easy to see from Lemma 1 that the invertibility is implied by the condition

$$a(\mathbf{z}) \neq 0 \text{ for all } |z_1| \leq 1 \text{ and } |z_2| = 1, \quad \text{and} \quad 1 + \sum_{(0,k)\in\mathcal{I}_2} a_{0k} z_2^k \neq 0 \text{ for all } |z_2| \leq 1. \qquad (2.9)$$

Furthermore, under this condition the coefficients $\{\varphi_{jk}\}$ and their partial derivatives (with respect to $b_{\mathbf{i}}$ or $a_{\mathbf{i}}$) decay at an exponential rate.

(vi) The spectral density function of $\{X(\mathbf{s})\}$ is of the form

$$g(\boldsymbol{\omega}) = \frac{\sigma^2}{4\pi^2}\left|\frac{a(e^{i\boldsymbol{\omega}})}{b(e^{i\boldsymbol{\omega}})}\right|^2, \quad \boldsymbol{\omega} \in [-\pi, \pi]^2, \tag{2.10}$$

where $i = \sqrt{-1}$, $\boldsymbol{\omega} = (\omega_1, \omega_2)$ and $e^{i\boldsymbol{\omega}} = (e^{i\omega_1}, e^{i\omega_2})$. Under conditions (2.6) and (2.9), $g(\boldsymbol{\omega})$ is bounded away from both 0 and ∞, which is the condition used in Guyon (1982). Note the condition that $g(\boldsymbol{\omega})$ is bounded away from both 0 and ∞ is equivalent to the condition that $a(\mathbf{z})b(\mathbf{z}) \neq 0$ for all $|z_1| = |z_2| = 1$ and $(z_1, z_2) \in \mathcal{C}^2$. Under this condition equation (2.1) defines a weakly stationary process which, however, is not necessarily causal or invertible (Justice and Shanks 1973). Helson and Lowdenslager (1958) shows that the necessary and sufficient condition for a weakly stationary (but not necessarily ARMA) process $\{X(\mathbf{s})\}$ admitting the MA representation (2.4) with squared-summable coefficients φ_{jk} is that its spectral density $g(\cdot)$ fulfils the condition

$$\int_{[-\pi,\pi]^2} \log g(\boldsymbol{\omega}) d\boldsymbol{\omega} > -\infty. \tag{2.11}$$

Note that for ARMA processes, (2.11) is implied by (2.6).

2.2. Gaussian MLEs

We denote the elements of $\mathcal{I}_1$ and $\mathcal{I}_2$ in the ascending order respectively as

$$\mathbf{j}_1 < \mathbf{j}_2 < \cdots < \mathbf{j}_p \quad \text{and} \quad \mathbf{i}_1 < \mathbf{i}_2 < \cdots < \mathbf{i}_q.$$

Let $\boldsymbol{\theta} \equiv (\theta_1, \cdots, \theta_{p+q})^\tau = (b_{\mathbf{j}_1}, \cdots, b_{\mathbf{j}_p}, a_{\mathbf{i}_1}, \cdots, a_{\mathbf{i}_q})^\tau$. We assume $\boldsymbol{\theta} \in \Theta$, where $\Theta \subset \mathcal{R}^{p+q}$ is the parameter space. To avoid some delicate technical arguments, we assume the condition below holds.

(C1) The parameter space Θ is a compact set containing the true value $\boldsymbol{\theta}_0$ as an interior point. Further, for any $\boldsymbol{\theta} \in \Theta$, conditions (2.6) and (2.9) holds.

Given observations $\{X(u, v), u = 1, \cdots, N_1, v = 1, \cdots, N_2\}$ from model (2.1), the Gaussian likelihood function is of the form

$$L(\boldsymbol{\theta}, \sigma^2) \propto \sigma^{-N} |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} \exp\{-\frac{1}{2\sigma^2}\mathbf{X}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}\}, \tag{2.12}$$

where $N = N_1 N_2$, $\mathbf{X}$ is an $N \times 1$ vector consisting of the N observations in ascending order, and

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\text{Var}(\mathbf{X}),$$

which is independent of σ^2. The estimators which maximise (2.12) can be expressed as

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \left[\log\{\mathbf{X}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}/N\} + N^{-1}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})|\right], \quad \widehat{\sigma}^2 = \mathbf{X}^\tau \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}/N. \tag{2.13}$$

Since we do not assume a special form for the distribution of $\varepsilon(\mathbf{s})$ and the Gaussian likelihood is used only as a contrast function, the derived estimators could be referred to as quasi-MLEs.

2.3. Prewhitening and the innovation algorithm

Gaussian maximum likelihood estimation has been hampered by the computational burden in calculating both the inverse and the determinant of $N\times N$ matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. To overcome the burden, some approximation methods have been developed by, for example, Besag (1975), and Wood and Chan (1994). See also §7.2 of Cressie (1993). The computational difficulty has been gradually eased by the increase of computer power. It is now feasible to compute the genuine Gaussian likelihood functions with N in the order of thousands. As an example, we state below how the idea of prewhitening via the innovation algorithm can be used to facilitate the computation for Gaussian likelihood regardless whether the underlying process is stationary or not, or whether the data are collected on a regular grid or not. Prewhitening is an old and very useful idea in time series analysis. Effectively it is a version of the Cholesky decomposition, and it computes the quadratic form $\mathbf{X}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}$ and the determinant $|\boldsymbol{\Sigma}(\boldsymbol{\theta})|$ simultaneously. Our edge-effect correction method, presented in §2.4 below, is based on a representation of the likelihood in terms of prewhitening.

Denote by $X(\mathbf{s}_1), \cdots, X(\mathbf{s}_N)$ the N observations with the indices $\mathbf{s}_j$ in ascending order. (The order is not important as far as the algorithm presented below is concerned.) Let $\widehat{X}(\mathbf{s}_1) \equiv 0$. For $1 \le k < N$, let

$$\widehat{X}(\mathbf{s}_{k+1}) = \varphi_1^{(k)} X(\mathbf{s}_k) + \cdots + \varphi_k^{(k)} X(\mathbf{s}_1) \tag{2.14}$$

be the best linear predictor for $X(\mathbf{s}_{k+1})$ based on $X(\mathbf{s}_k), \cdots, X(\mathbf{s}_1)$ in the sense that

$$E\{X(\mathbf{s}_{k+1}) - \widehat{X}(\mathbf{s}_{k+1})\}^2 = \min_{\{\psi_j\}} E\{X(\mathbf{s}_{k+1}) - \sum_{j=1}^{k} \psi_j X(\mathbf{s}_{k-j+1})\}^2. \tag{2.15}$$

It can be shown that the coefficients $\varphi_j^{(k)}$ are the solutions of equations

$$\gamma(\mathbf{s}_l) = \sum_{j=1}^{k} \varphi_j^{(k)} \gamma(\mathbf{s}_l - \mathbf{s}_j), \quad l = 1, \cdots, k,$$

and further

$$r(\mathbf{s}_{k+1}) \equiv r(\mathbf{s}_{k+1}, \boldsymbol{\theta}) \equiv \frac{1}{\sigma^2} E\{X(\mathbf{s}_{k+1}) - \widehat{X}(\mathbf{s}_{k+1})\}^2 = \frac{1}{\sigma^2}\{\gamma(\mathbf{0}) - \sum_{j=1}^{k} \varphi_j^{(k)} \gamma(\mathbf{s}_j)\}. \tag{2.16}$$

In the above expressions, $\gamma(\mathbf{i}) \equiv \gamma(\mathbf{i}; \boldsymbol{\theta}) = E\{X(\mathbf{s}+\mathbf{i})X(\mathbf{s})\}$, and $\mathbf{0} = (0, 0)$. It can also be shown that the least square property (2.15) implies that

$$\mathrm{Cov}[\{X(\mathbf{s}_{k+1}) - \widehat{X}(\mathbf{s}_{k+1})\}X(\mathbf{s}_j)] = 0, \quad 1 \leq j \leq k.$$

Note that $X(\mathbf{s}_{k+1}) - \widehat{X}(\mathbf{s}_{k+1})$ is a linear combination of $X(\mathbf{s}_k), \cdots, X(\mathbf{s}_1)$. Thus $X(\mathbf{s}_1) - \widehat{X}(\mathbf{s}_1)$, $\cdots, X(\mathbf{s}_N) - \widehat{X}(\mathbf{s}_N)$ are N uncorrelated random variables. Further it is easy to see from (2.14) that $\widehat{X}(\mathbf{s}_k)$ can be written as a linear combination of $X(\mathbf{s}_k) - \widehat{X}(\mathbf{s}_k), \cdots, X(\mathbf{s}_1) - \widehat{X}(\mathbf{s}_1)$. We write

$$\widehat{X}(\mathbf{s}_{k+1}) = \sum_{j=1}^{k} \beta_{kj}\{X(\mathbf{s}_{k+1-j}) - \widehat{X}(\mathbf{s}_{k+1-j})\}, \quad k = 1, \cdots, N-1. \tag{2.17}$$

Let $\widehat{\mathbf{X}} = (\widehat{X}(\mathbf{s}_1), \cdots, \widehat{X}(\mathbf{s}_N))^\tau$. Then $\widehat{\mathbf{X}} = \mathbf{A}(\mathbf{X} - \widehat{\mathbf{X}})$, where

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \beta_{11} & 0 & 0 & \cdots & 0 & 0 \\ \beta_{22} & \beta_{21} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 & 0 \\ \beta_{N-1,N-1} & \beta_{N-1,N-2} & \beta_{N-1,N-3} & \cdots & \beta_{N-1,1} & 0 \end{pmatrix}.$$

Put $\mathbf{X} = \mathbf{C}(\mathbf{X} - \widehat{\mathbf{X}})$, where $\mathbf{C} = \mathbf{A} + \mathbf{I}_N$ is a lower-triangular matrix with all main diagonal elements 1, and $\mathbf{I}_N$ is the $N \times N$ identity matrix. Let $\mathbf{D} = \mathrm{diag}\{r(\mathbf{s}_1), \cdots, r(\mathbf{s}_N)\}$. Then

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\mathrm{Var}(\mathbf{X}) = \mathbf{C}\mathbf{D}\mathbf{C}^\tau, \quad \text{and} \quad |\boldsymbol{\Sigma}(\boldsymbol{\theta})| = |\mathbf{D}| = \prod_{j=1}^{N} r(\mathbf{s}_j). \tag{2.18}$$

Hence the likelihood function defined in (2.12) can be written as

$$L(\boldsymbol{\theta}, \sigma^2) \propto \sigma^{-N}\{r(\mathbf{s}_1)\cdots r(\mathbf{s}_N)\}^{-1/2} \exp[-\frac{1}{2\sigma^2}\sum_{j=1}^{N}\{X(\mathbf{s}_j) - \widehat{X}(\mathbf{s}_j)\}^2 / r(\mathbf{s}_j)]. \tag{2.19}$$

The calculation of the inverse and the determinant of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is reduced to the calculation of the coefficients β_{kj} and $r(\mathbf{s}_{k+1})$ defined in (2.17) and (2.16) respectively, which can be easily done recursively using the innovation algorithm below; see Proposition 5.2.2 of Brockwell and Davis (1991). We present the algorithm in the form applicable to any (non-stationary) series $\{X(\mathbf{s}_j)\}$ with common mean 0 and auto-covariance $\gamma(\mathbf{s}_k, \mathbf{s}_j) = E\{X(\mathbf{s}_k)X(\mathbf{s}_j)\}$, which reduces to $\gamma(\mathbf{s}_k - \mathbf{s}_j; \boldsymbol{\theta})$ for the stationary spatial ARMA process concerned in this paper. Note that the algorithm is a version of Cholesky decomposition.

Innovation algorithm: Set $r(\mathbf{s}_1) = \gamma(\mathbf{s}_1, \mathbf{s}_1)/\sigma^2$. Based on the cross-recursion equations

$$\begin{aligned}\beta_{k,k-j} &= \{\gamma(\mathbf{s}_{k+1}, \mathbf{s}_{j+1})/\sigma^2 - \sum_{i=0}^{j-1} \beta_{j,j-i}\beta_{k,k-i}\, r(\mathbf{s}_{i+1})\}/r(\mathbf{s}_{j+1}),\\ r(\mathbf{s}_{k+1}) &= \gamma(\mathbf{s}_{k+1}, \mathbf{s}_{k+1})/\sigma^2 - \sum_{j=0}^{k-1} \beta_{k,k-j}^2 r(\mathbf{s}_{j+1}),\end{aligned}$$

compute the values of $\{\beta_{ij}\}$ and $\{r(\mathbf{s}_j)\}$ in the order $\beta_{11}, r(\mathbf{s}_2), \beta_{22}, \beta_{21}, r(\mathbf{s}_3), \beta_{33}, \beta_{32}, \beta_{31}, r(\mathbf{s}_4), \cdots$ $\beta_{N-1,N-1}, \beta_{N-1,N-2}, \cdots, \beta_{N-1,1}, r(\mathbf{s}_N)$.

2.4. A modified estimator

In order to establish the asymptotic normality, we propose a modified maximum likelihood estimator which may be viewed as a counterpart of conditional maximum likelihood estimators for (one-dimensional) time series processes. Our edge correction scheme depends on the way in which the sample size tends to infinity. Condition (C2) specifies that $N = N_1 N_2 \to \infty$ in one of three ways.

(C2) One of the following three conditions holds,

(i) $N_1 \to \infty$, and N_1/N_2 has a limit $d \in (0, \infty)$,

(ii) $N_2 \to \infty$ and $N_1/N_2 \to \infty$,

(iii) $N_1 \to \infty$ and $N_1/N_2 \to 0$.

For, $n_1, n_2 \to \infty$ and $n_1/N_1, n_2/N_2 \to 0$, define

$$\mathcal{I}^* = \begin{cases} \{(u,v):\ n_1 \le u \le N_1,\ n_2 \le v \le N_2 - n_2\} & \text{if } N_1/N_2 \to d \in (0,\infty),\\ \{(u,v):\ 1 \le u \le N_1,\ n_2 \le v \le N_2 - n_2\} & \text{if } N_1/N_2 \to \infty,\\ \{(u,v):\ n_1 \le u \le N_1,\ 1 \le v \le N_2\} & \text{if } N_1/N_2 \to 0.\end{cases}$$

Write $\mathcal{I}^* = \{\mathbf{t}_1, \cdots, \mathbf{t}_{N^*}\}$ with $\mathbf{t}_1 < \cdots < \mathbf{t}_{N^*}$. Then $N^*/N \to 1$ under (C2). Based on (2.19), the modified likelihood function is defined as

$$L^*(\boldsymbol{\theta}, \sigma^2) \propto \sigma^{-N^*} \{r(\mathbf{t}_1) \cdots r(\mathbf{t}_{N^*})\}^{-1/2} \exp[-\frac{1}{2\sigma^2} \sum_{j=1}^{N^*} \{X(\mathbf{t}_j) - \widehat{X}(\mathbf{t}_j)\}^2 / r(\mathbf{t}_j)]. \tag{2.20}$$

The modified estimators, obtained from maximising the above, are denoted as $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\sigma}^2$.

3. Consistency

Theorem 1. Let $\{\varepsilon(\mathbf{s})\} \sim \text{IID}(0, \sigma^2)$ and condition (C1) hold. Then as both N_1 and $N_2 \to \infty$, $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ and $\widehat{\sigma}^2 \xrightarrow{P} \sigma^2$. Furthermore, $\widetilde{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ and $\widetilde{\sigma}^2 \xrightarrow{P} \sigma^2$ provided condition (C2) also holds.

Proof. We only prove the consistency for $\widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma}^2$ below. The proof for the consistency of $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\sigma}^2$ is similar and therefore omitted.

Note that $\widehat{\boldsymbol{\theta}}$ does not depend on σ^2; see (2.13). It follows from (2.12) and Lemma 2 below that

$$\frac{1}{N}\mathbf{X}^\tau \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X} \leq \frac{1}{N}\mathbf{X}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}\mathbf{X} + \frac{\sigma^2}{N}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)|.$$

By Lemmas 2 & 3 below, it holds that

$$\limsup_{N\to\infty} \frac{1}{N}\mathbf{X}^\tau \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X} \leq \lim_{N\to\infty} \frac{1}{N}\mathbf{X}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)^{-1}\mathbf{X} = \sigma^2. \tag{3.1}$$

For any $\epsilon > 0$, define $B_{N_1,N_2} = \{|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| > \epsilon\}$ and $B = \cup_{k_1\geq 1, k_2 \geq 1}\{\cap_{N_1 \geq k_1, N_2 \geq k_2} B_{N_1,N_2}\}$. For any $\omega \in B$, there exists a subsequence of $\{N_1, N_2\}$, which we still denote as $\{N_1, N_2\}$, for which $\widehat{\boldsymbol{\theta}}(\omega) \equiv \widehat{\boldsymbol{\theta}}_{N_1,N_2}(\omega) \to \boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. By Lemma 4 below, we have for any $\epsilon > 0$,

$$\frac{1}{N}|\mathbf{X}^\tau \boldsymbol{\Sigma}\{\widehat{\boldsymbol{\theta}}(\omega)\}^{-1}\mathbf{X} - \mathbf{X}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}| \leq \epsilon\widehat{\gamma}(0),$$

where $\widehat{\gamma}(0) = N^{-1}\sum_{j=1}^N X_{s_j}^2$. Thus

$$\limsup_{N\to\infty} \frac{1}{N}\mathbf{X}^\tau \boldsymbol{\Sigma}\{\widehat{\boldsymbol{\theta}}(\omega)\}^{-1}\mathbf{X} = \limsup_{N\to\infty} \frac{1}{N}\mathbf{X}^\tau \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}$$

provided one of the above two limits exist. Now Lemma 3 and (3.1) imply $P(B) = 0$. Thus $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$. By Lemma 4 and (3.1) again, $\widehat{\sigma}^2 = \mathbf{X}^\tau \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}/N \xrightarrow{P} \sigma^2$. ■

In this paper, we assume that the observations are taken from a rectangle. Theorem 1 requires that the two sides of the rectangle increase to infinity. In fact this assumption can be relaxed. Theorem 1 still holds if the observations were taken over a connected region in $\mathcal{Z}^2$, and both minimal length of side of the squares containing the region N_1 and the maximal length of side of the squares contained in the region N_2 converge to ∞. For general discussion on the condition of sampling sets, we refer to Perera (2001).

We denote by $b_0(\cdot)$ and $a_0(\cdot)$ the polynomials defined as in (2.2) with coefficients corresponding to the true parameter vector $\boldsymbol{\theta}_0$, and $b(\cdot)$ and $a(\cdot)$ the polynomials corresponding to $\boldsymbol{\theta}$. For

$\mathbf{s} = (u, v)$ with $u \geq 1$ and $1 \leq v < N_2$, define

$$\begin{aligned} \mathcal{A}_{\mathbf{s}} &= \{(0,k):\ k \geq v\} \cup \{(j,k):\ j \geq u,\ -\infty < k < \infty\} \\ &\cup\ \{(j,k):\ 1 \leq j < u,\ k \geq v \text{ or } k < -(N_2 - v)\}, \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} \widetilde{\boldsymbol{\varphi}}_{\mathbf{s}} &= (\varphi_{01}, \varphi_{02}, \cdots, \varphi_{0,v-1}, \varphi_{1,-(N_2-v)}, \varphi_{1,-(N_2-v)+1}, \cdots, \\ &\quad \varphi_{1,v-1}, \varphi_{2,-(N_2-v)}, \cdots, \varphi_{u-1,v-1})^\tau, \end{aligned} \tag{3.3}$$

$$\begin{aligned} \boldsymbol{\varphi}_{\mathbf{s}} &= (\varphi_{01}^{(u,v)}, \varphi_{02}^{(u,v)}, \cdots, \varphi_{0,v-1}^{(u,v)}, \varphi_{1,-(N_2-v)}^{(u,v)}, \varphi_{1,-(N_2-v)+1}^{(u,v)}, \cdots, \\ &\quad \varphi_{1,v-1}^{(u,v)}, \varphi_{2,-(N_2-v)}^{(u,v)}, \cdots, \varphi_{u-1,v-1}^{(u,v)})^\tau. \end{aligned} \tag{3.4}$$

We use $C, C_1, C_2, \cdots$ to denote positive generic constants, which may be different in different places. In the remainder of this section, we always assume that the condition of Theorem 1 holds, i.e. $\{\varepsilon(\mathbf{s})\} \sim \text{IID}(0, \sigma^2)$ and that condition (C1) holds.

Lemma 2. For any $\boldsymbol{\theta} \in \Theta$, $\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| > 0$ and $\frac{1}{N}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| \to 0$.

For its proof, see Lemma 1 of Yao and Brockwell (2005).

Lemma 3. For any $\boldsymbol{\theta} \in \Theta$,

$$\mathbf{X}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}/N \xrightarrow{P} \text{Var}\{a(\mathbf{B})^{-1}b(\mathbf{B})X(\mathbf{s})\} \geq \text{Var}\{\varepsilon(\mathbf{s})\} = \sigma^2,$$

and the equality holds if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof. Let $\{Y(\mathbf{s})\}$ be the process defined by $b(\mathbf{B})Y(\mathbf{s}) = a(\mathbf{B})e(\mathbf{s})$ with $\{e(\mathbf{s})\} \sim \text{IID}(0,1)$. Let $\mathbf{Y} = \{Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_N)\}^\tau$. Then $\text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. Let $\widehat{Y}(1,1) \equiv 0$, and for $(u,v) > (1,1)$ let

$$\widehat{Y}(u,v) \equiv \sum_{k=1}^{v-1} \varphi_{0k}^{(u,v)} Y(u, v-k) + \sum_{j=1}^{u-1} \sum_{k=-(N_2-v)}^{v-1} \varphi_{jk}^{(u,v)} Y(u-j, v-k) \tag{3.5}$$

be the best linear predictor of $Y(u,v)$ based on its lagged values occurring on the RHS of the above equation. Then it may be shown that the coefficients $\{\varphi_{jk}^{(u,v)}\}$ are determined by the equations

$$\begin{aligned} \gamma(l,m) &= \sum_{k=1}^{v-1} \varphi_{0k}^{(u,v)} \gamma(l, m-k) + \sum_{j=1}^{u-1} \sum_{k=-(N_2-v)}^{v-1} \varphi_{jk}^{(u,v)} \gamma(l-j, m-k), \\ &\qquad l = 0 \text{ and } 1 \leq m < v, \text{ or } 1 \leq l < u \text{ and } -(N_2 - v) \leq m < v. \end{aligned} \tag{3.6}$$

Let $\widehat{\mathbf{Y}} = \{\widehat{Y}(\mathbf{s}_1), \cdots, \widehat{Y}(\mathbf{s}_N)\}^\tau$. It follows from the same argument as in §2.3 that $\mathbf{Y} = \mathbf{C}(\mathbf{Y} - \widehat{\mathbf{Y}})$ where $\mathbf{C}$ is a $N \times N$ lower-triangular matrix with all the main diagonal elements 1 (hence its inverse exists), and $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{C}\mathbf{D}\mathbf{C}^\tau$ and $|\boldsymbol{\Sigma}(\boldsymbol{\theta})| = |\mathbf{D}|$, where $\mathbf{D} = \text{diag}\{r(\mathbf{s}_1), \cdots, r(\mathbf{s}_N)\}$, and

$$\begin{aligned} r(\mathbf{s}) &\equiv r(\mathbf{s}, \boldsymbol{\theta}) \equiv E\{\widehat{Y}(u,v) - Y(u,v)\}^2 \\ &= \gamma(0,0) - \sum_{k=1}^{v-1} \varphi_{0k}^{(u,v)} \gamma(0,k) - \sum_{j=1}^{u-1} \sum_{k=-(N_2-v)}^{v-1} \varphi_{jk}^{(u,v)} \gamma(j,k). \end{aligned} \tag{3.7}$$

Since $\{Y(\mathbf{s})\}$ is invertible, *i.e.*

$$Y(u,v) = e(u,v) + \sum_{k=1}^{\infty} \varphi_{0k} Y(u, v-k) + \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} \varphi_{jk} Y(u-j, v-k), \tag{3.8}$$

it may be shown that

$$1 = \text{Var}\{e(u,v)\} \le r(u,v) \to 1, \quad \text{as } \min\{u, v, N_2 - v\} \to \infty, \tag{3.9}$$

where r is defined in (3.7).

It follows from (3.5) and (3.8) that

$$\begin{aligned} M &\equiv E\{e(\mathbf{s}) + \sum_{\mathbf{i} \in \mathcal{A}_\mathbf{s}} \varphi_\mathbf{i} Y(\mathbf{s} - \mathbf{i}) - Y(\mathbf{s}) + \widehat{Y}(\mathbf{s})\}^2 \\ &= E\{\sum_{k=1}^{v-1} (\varphi_{0k}^{(u,v)} - \varphi_{0k}) Y(u, v-k) + \sum_{j=1}^{u-1} \sum_{k=-(N_2-v)}^{v-1} (\varphi_{jk}^{(u,v)} - \varphi_{jk}) Y(u-j, v-k)\}^2, \end{aligned} \tag{3.10}$$

where $\mathbf{s} = (u,v)$, $\mathbf{i} = (j,k)$. Let $\mathbf{Y}_\mathbf{s}$ be defined as in (3.18) below. It is easy to see from the second equation in (3.10) that

$$M = (\boldsymbol{\varphi}_\mathbf{s} - \widetilde{\boldsymbol{\varphi}}_\mathbf{s})^\tau \boldsymbol{\Sigma}_\mathbf{s}(\boldsymbol{\theta}) (\boldsymbol{\varphi}_\mathbf{s} - \widetilde{\boldsymbol{\varphi}}_\mathbf{s}) \ge \lambda_{\min} ||\boldsymbol{\varphi}_\mathbf{s} - \widetilde{\boldsymbol{\varphi}}_\mathbf{s}||^2, \tag{3.11}$$

where $\lambda_{\min}$ is the minimum eigenvalue of $\boldsymbol{\Sigma}_\mathbf{s}(\boldsymbol{\theta}) \equiv \text{Var}(\mathbf{Y}_\mathbf{s})$. By Lemma 5 below and condition (2.9), $\lambda_{\min}$ is uniformly (in N) bounded away from 0 (see also (2.10)). On the other hand, the first equation in (3.10) implies that

$$\begin{aligned} M &\le 2E\{\sum_{\mathbf{i} \in \mathcal{A}_\mathbf{s}} \varphi_\mathbf{i} Y(\mathbf{s} - \mathbf{i})\}^2 + 2E\{e(\mathbf{s}) - Y(\mathbf{s}) + \widehat{Y}(\mathbf{s})\}^2 \\ &\le 2\gamma(\mathbf{0}) \big(\sum_{\mathbf{i} \in \mathcal{A}_\mathbf{s}} |\varphi_\mathbf{i}| \big)^2 + 2\{r(\mathbf{s}) - 1\} \le 4\gamma(\mathbf{0}) \big(\sum_{\mathbf{i} \in \mathcal{A}_\mathbf{s}} |\varphi_\mathbf{i}| \big)^2. \end{aligned}$$

Recalling that $\mathbf{s} = (u,v)$, it follows from (3.11) and Lemma 1 that

$$||\boldsymbol{\varphi}_\mathbf{s} - \widetilde{\boldsymbol{\varphi}}_\mathbf{s}||^2 \le M/\lambda_{\min} \le \frac{4\gamma(\mathbf{0})}{\lambda_{\min}} \sum_{\mathbf{i} \in \mathcal{A}_\mathbf{s}} \varphi_\mathbf{i}^2 \le C(\alpha^u + \alpha^v + \alpha^{N_2 - v}), \tag{3.12}$$

which converges to 0 as $\min(u, v, N_2 - v) \to \infty$, where $\alpha \in (0,1)$ is a constant.

Now define

$$\widetilde{X}(u,v)=\sum_{k=1}^{v-1}\varphi_{0k}^{(u,v)}X(u,v-k)+\sum_{j=1}^{u-1}\sum_{k=-(N_2-v)}^{v-1}\varphi_{jk}^{(u,v)}X(u-j,v-k),$$

where the coefficients φ_{jk} are defined as in (3.5). Let $\widetilde{\mathbf{X}}=\{\widetilde{X}(\mathbf{s}_1),\cdots,\widetilde{X}(\mathbf{s}_N)\}^\tau$, then $\mathbf{X}=\mathbf{C}(\mathbf{X}-\widetilde{\mathbf{X}})$, and

$$\frac{1}{N}\mathbf{X}^\tau\mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}=\frac{1}{N}(\mathbf{X}-\widetilde{\mathbf{X}})^\tau\mathbf{D}^{-1}(\mathbf{X}-\widetilde{\mathbf{X}})=\frac{1}{N}\sum_{m=1}^{N}\{X(\mathbf{s}_m)-\widetilde{X}(\mathbf{s}_m)\}^2/r(\mathbf{s}_m). \tag{3.13}$$

It follows from Lemma 1 that for any $\epsilon>0$, we may choose $K>0$ such that

$$E\left(\sum_{k>K}|\varphi_{0k}X(0,-k)|+\sum_{\substack{j>K,\text{ or}\\ j\le K\ \&\ |k|>K}}|\varphi_{jk}X(-j,-k)|\right)^2<\epsilon. \tag{3.14}$$

For $\mathbf{s}=(u,v)$ with $u>K$ and $K<v<N_2-K$, let $X(\mathbf{s})-\widetilde{X}(\mathbf{s})=\eta_1(\mathbf{s})+\eta_2(\mathbf{s})+\eta_3(\mathbf{s})$, where

$$\eta_1(\mathbf{s})=X(u,v)-\sum_{k=1}^{K}\varphi_{0k}X(u,v-k)-\sum_{j=1}^{K}\sum_{k=-K}^{K}\varphi_{jk}X(u-j,v-k),$$

$$\eta_2(\mathbf{s})=\sum_{k=1}^{v-1}(\varphi_{0k}-\varphi_{0k}^{(u,v)})X(u,v-k)+\sum_{j=1}^{u-1}\sum_{k=-(N_2-v)}^{v-1}(\varphi_{jk}-\varphi_{jk}^{(u,v)})X(u-j,v-k),$$

and $\eta_3(\mathbf{s})=-\sum_{\mathbf{i}\in\mathcal{A}}\varphi_{\mathbf{i}}X(\mathbf{s}-\mathbf{i})$ with

$$\begin{aligned}\mathcal{A}\ &=\ \{(0,k):\ K<k<v\}\cup\{(j,k):\ K<j<u,\ -(N_2-v)\le k<v\}\\ &\quad\cup\{(j,k):\ 1\le j\le K,\ -(N_2-v)\le k<-K\text{ or }K<k<v\}.\end{aligned}$$

By (3.14),

$$E\{\eta_3(u,v)^2\}\le E\{\sum_{\mathbf{i}\in\mathcal{A}}|\varphi_{\mathbf{i}}X(\mathbf{s}-\mathbf{i})|\}^2<\epsilon. \tag{3.15}$$

On the other hand, it is easy to see from (3.10) — (3.12) that

$$E\{\eta_2(\mathbf{s})^2\}=\sigma^2(\boldsymbol{\varphi}_\mathbf{s}-\widetilde{\boldsymbol{\varphi}}_\mathbf{s})^\tau\mathbf{\Sigma}_\mathbf{s}(\boldsymbol{\theta}_0)(\boldsymbol{\varphi}_\mathbf{s}-\widetilde{\boldsymbol{\varphi}}_\mathbf{s})\le\sigma^2\lambda_{\max}||\boldsymbol{\varphi}_\mathbf{s}-\widetilde{\boldsymbol{\varphi}}_\mathbf{s}||^2\to0, \tag{3.16}$$

where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{\Sigma}_\mathbf{s}(\boldsymbol{\theta}_0)$. By Lemma 5 and (2.6), $\lambda_{\max}$ is uniformly (in N) bounded from the above by a finite constant. Based on (2.4) and the fact that $\{\mathbf{e}(\mathbf{s})\}\sim\text{IID}(0,\sigma^2)$, we can show that for any fixed K,

$$\frac{1}{N}\sum_{u=K+1}^{N_1}\sum_{v=K+1}^{N_2-K-1}\eta_1(u,v)^2\xrightarrow{a.s.}E\{\eta_1(\mathbf{s})\}^2. \tag{3.17}$$

Note that for any fixed K, it holds almost surely that

$$\frac{1}{N}\sum_{m=1}^{N}\{X(\mathbf{s}_m)-\widetilde{X}(\mathbf{s}_m)\}^2=\frac{1}{N}\sum_{u=K+1}^{N_1}\sum_{v=K+1}^{N_2-K-1}\{X(u,v)-\widetilde{X}(u,v)\}^2+O(\frac{N_1+N_2}{N}).$$

It follows from (3.15) — (3.17) and the Cauchy-Schwarz inequality that

$$\frac{1}{N}\sum_{m=1}^{N}\{X(\mathbf{s}_m)-\widetilde{X}(\mathbf{s}_m)\}^2\xrightarrow{P}E\{\eta_1(\mathbf{s})\}^2.$$

From (3.14),

$$|E\{\eta_1(\mathbf{s})\}^2-E\{a(\mathbf{B})^{-1}b(\mathbf{B})X(\mathbf{s})\}^2|<\epsilon+2[\epsilon E\{\eta_1(\mathbf{s})\}^2]^{1/2}.$$

Letting $K\to\infty$, we find that

$$E\{\eta_1(\mathbf{s})\}^2\to E\{a(\mathbf{B})^{-1}b(\mathbf{B})X(\mathbf{s})\}^2=E\{a(\mathbf{B})^{-1}b(\mathbf{B})b_0(\mathbf{B})^{-1}a_0(\mathbf{B})\varepsilon(\mathbf{s})\}\geq \mathrm{Var}\{\varepsilon(\mathbf{s})\}.$$

The required result now follows from (3.13) and (3.9). ■

Lemma 4. Let $\boldsymbol{\theta}_k\in\Theta$ and $\boldsymbol{\theta}_k\to\boldsymbol{\theta}\in\Theta$ as $k\to\infty$. Let $\epsilon>0$ be any constant (independent of N). Then there exists $M(\epsilon)>0$ such that for all $N\geq 1$ and $k>M(\epsilon)$,

$$|\mathbf{x}^\tau\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{x}-\mathbf{x}^\tau\boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-1}\mathbf{x}|\leq\epsilon,\quad \mathbf{x}\in\mathcal{R}^N \text{ and } ||\mathbf{x}||=1.$$

Proof. Let $g(\boldsymbol{\omega},\boldsymbol{\theta})$ be the the spectral density function defined in (2.10). Condition (C1) ensures that $g(\boldsymbol{\omega},\cdot)$ is continuous and bounded away from both 0 and ∞ on Θ. Hence for any $\epsilon'>0$, it holds for all sufficiently large k that

$$\sup_{\boldsymbol{\omega}\in[-\pi,\pi)^2}\big|g(\boldsymbol{\omega},\boldsymbol{\theta})-g(\boldsymbol{\omega},\boldsymbol{\theta}_k)\big|<\epsilon'.$$

Note that $\gamma(j,k)=\int_{[-\pi,\pi)^2}e^{i(j\omega_1+k\omega_2)}g(\omega_1,\omega_2,\boldsymbol{\theta})d\omega_1d\omega_2$, where $i=\sqrt{-1}$. Hence

$$\begin{aligned}\mathbf{x}^\tau\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{y} &= \sum_{j,u=1}^{N_1}\sum_{k,v=1}^{N_2}x_{jk}y_{uv}\gamma(j-u,k-v)\\ &= \int_{[-\pi,\pi)^2}g(\boldsymbol{\omega},\boldsymbol{\theta})\sum_{j,u=1}^{N_1}\sum_{k,v=1}^{N_2}x_{jk}e^{i(j\omega_1+k\omega_2)}y_{uv}e^{-i(u\omega_1+v\omega_2)}d\omega_1d\omega_2.\end{aligned}$$

where

$$\mathbf{x}=(x_{11},x_{12},x_{1,N_2},x_{21},\cdots,x_{N_1,N_2})^\tau\quad\text{and}\quad\mathbf{y}=(y_{11},y_{12},y_{1,N_2},y_{21},\cdots,y_{N_1,N_2})^\tau.$$

Under the additional condition $||\mathbf{x}|| = ||\mathbf{y}|| = 1$, it holds now that

$$\begin{aligned}
& \big|\mathbf{x}^\tau\big\{\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)\big\}\mathbf{y}\big| \\
= & \Big|\int_{[-\pi,\pi)^2}\{g(\boldsymbol{\omega},\boldsymbol{\theta}) - g(\boldsymbol{\omega},\boldsymbol{\theta}_k)\}\sum_{j,u=1}^{N_1}\sum_{k,v=1}^{N_2} x_{jk}e^{i(j\omega_1+k\omega_2)}y_{uv}e^{-i(u\omega_1+v\omega_2)}d\omega_1 d\omega_2\Big| \\
\le & \frac{\epsilon'}{2}\int_{[-\pi,\pi)^2}\Big(\Big|\sum_{j=1}^{N_1}\sum_{k=1}^{N_2} x_{jk}e^{i(j\omega_1+k\omega_2)}\Big|^2 + \Big|\sum_{u=1}^{N_1}\sum_{v=1}^{N_2} y_{uv}e^{-i(u\omega_1+v\omega_2)}\Big|^2\Big)d\omega_1 d\omega_2 \\
= & \frac{\epsilon'}{2}\int_{[-\pi,\pi)^2}\Big(\sum_{j=1}^{N_1}\sum_{k=1}^{N_2} x_{jk}^2 + \sum_{u=1}^{N_1}\sum_{v=1}^{N_2} y_{uv}^2\Big)d\omega_1 d\omega_2 \;=\; 4\pi^2\epsilon'.
\end{aligned}$$

Lemma 5 below and condition (C1) ensure that for all $\boldsymbol{\theta} \in \Theta$ and N, the minimum eigenvalue of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is bounded from below by a constant $K^{-1} > 0$, where K is independent of $\boldsymbol{\theta}$ and N. Hence

$$\begin{aligned}
& \big|\mathbf{x}^\tau\{\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-1}\}\mathbf{x}\big| \;=\; \big|\mathbf{x}^\tau\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\{\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_k)\big\}\boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-1}\mathbf{x}\big| \\
\le\; & 4\pi^2\epsilon'\big\{\big|\mathbf{x}^\tau\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\mathbf{x}\big|\,\big|\mathbf{x}^\tau\boldsymbol{\Sigma}(\boldsymbol{\theta}_k)^{-2}\mathbf{x}\big|\big\}^{1/2} \;\le\; 4\pi^2\epsilon' K^2.
\end{aligned}$$

Now the lemma holds by putting $\epsilon' = \epsilon/(4\pi^2 K^2)$. ■

Lemma 5. Let $\{Y(\mathbf{s})\}$ be a weakly stationary spatial process with spectral density $g(\boldsymbol{\omega})$. Let N_2 be a positive integer. For $\mathbf{s} = (u, v)$ with $u \ge 1$ and $1 \le v \le N_2$, define

$$\begin{aligned}
\mathbf{Y_s} \;=\; & \{Y(u, v-1), Y(u, v-2), \cdots, Y(u, 1), Y(u-1, N_2), Y(u-1, N_2-1), \\
& \cdots, Y(u-1, 1), Y(u-2, N_2), \cdots, Y(1, 1)\}^\tau,
\end{aligned} \tag{3.18}$$

and $\boldsymbol{\Sigma}_\mathbf{s} = \text{Var}(\mathbf{Y_s})$. It holds that for any eigenvalue λ of $\boldsymbol{\Sigma}_\mathbf{s}$,

$$\inf_{\boldsymbol{\omega}\in[-\pi,\pi)^2} g(\boldsymbol{\omega}) \le \frac{\lambda}{4\pi^2} \le \sup_{\boldsymbol{\omega}\in[-\pi,\pi)^2} g(\boldsymbol{\omega}). \tag{3.19}$$

Proof. Let

$$\mathbf{x} = (x_{01}, x_{02}, \cdots, x_{0,v-1}, x_{1,-(N_2-v)}, x_{1,-(N_2-v)+1}, \cdots, x_{1,v-1}, x_{2,-(N_2-v)}, \cdots, x_{u-1,v-1})^\tau$$

be an eigenvector of $\boldsymbol{\Sigma}_\mathbf{s}$ corresponding to the eigenvalue λ and $||\mathbf{x}|| = 1$. Let $m = \inf_{\boldsymbol{\omega}} g(\boldsymbol{\omega})$ and $M = \sup_{\boldsymbol{\omega}} g(\boldsymbol{\omega})$. Since $\text{Cov}\{Y(u+j, v+k), Y(u,v)\} = \int_{[-\pi,\pi)^2} e^{i(j\omega_1+k\omega_2)}g(\omega_1,\omega_2)d\omega_1 d\omega_2$, where

$i = \sqrt{-1}$, it holds that

$$\begin{aligned}\lambda &= \mathbf{x}^\tau \boldsymbol{\Sigma}_{\mathbf{s}} \mathbf{x} = \int_{[-\pi,\pi)^2} \left| \sum_{k=1}^{v-1} x_{0k} e^{ik\omega_2} + \sum_{j=1}^{u-1} \sum_{k=-(N_2-v)}^{v-1} x_{jk} e^{i(j\omega_1 + k\omega_2)} \right|^2 g(\omega_1, \omega_2) d\omega_1 d\omega_2 \\ &\in [m,\ M] \times \int_{[-\pi,\pi)^2} \left| \sum_{k=1}^{v-1} x_{0k} e^{ik\omega_2} + \sum_{j=1}^{u-1} \sum_{k=-(N_2-v)}^{v-1} x_{jk} e^{i(j\omega_1 + k\omega_2)} \right|^2 d\omega_1 d\omega_2 \\ &= [m,\ M] \times 4\pi^2 (\sum_{k=1}^{v-1} x_{0k}^2 + \sum_{j=1}^{u-1} \sum_{k=-(N_2-v)}^{v-1} x_{jk}^2) = [4\pi^2 m,\ 4\pi^2 M].\end{aligned}$$

■

Remark 2. (i) Expression (3.19) still holds if we replace (λ, g) by $(\dot\lambda, \dot g)$, where $\dot g$ and $\dot{\boldsymbol{\Sigma}}_{\mathbf{s}}$ are derivatives of g and $\boldsymbol{\Sigma}_{\mathbf{s}}$ with respect to a parameter, and $\dot\lambda$ is an eigenvalue of $\dot{\boldsymbol{\Sigma}}_{\mathbf{s}}$.

(ii) For an ARMA process, condition (2.6) implies $\sup_{\boldsymbol{\omega}} g(\boldsymbol{\omega}) < \infty$ and $\sup_{\boldsymbol{\omega}} \dot g(\boldsymbol{\omega}) < \infty$, and condition (2.9) implies that $\inf_{\boldsymbol{\omega}} g(\boldsymbol{\omega}) > 0$.

4. Asymptotic normality

To state the asymptotic normality of the estimator $\widetilde{\boldsymbol{\theta}}$ obtained from maximising (2.20), we let $\{W(\mathbf{s})\}$ be a spatial white noise process with mean 0 and variance 1. Define

$$b(\mathbf{B})\xi(\mathbf{s}) = W(\mathbf{s}) \quad \text{and} \quad a(\mathbf{B})\eta(\mathbf{s}) = W(\mathbf{s}). \tag{4.1}$$

Let $\boldsymbol{\xi} = \{\xi(-\mathbf{j}_1), \xi(-\mathbf{j}_2), \cdots, \xi(-\mathbf{j}_p), \eta(-\mathbf{i}_1), \eta(-\mathbf{i}_2), \cdots, \eta(-\mathbf{i}_q)\}^\tau$, and put

$$\mathbf{W}(\boldsymbol{\theta}) = \{\mathrm{Var}(\boldsymbol{\xi})\}^{-1}. \tag{4.2}$$

Theorem 2. Let $\{\varepsilon(\mathbf{s})\} \sim \mathrm{IID}(0, \sigma^2)$ and conditions (C1) and (C2) hold. Then $N^{1/2}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N\{\mathbf{0},\ \mathbf{W}(\boldsymbol{\theta}_0)\}$.

Remark 3. In the context of estimating the coefficients of ARMA models, the modified Whittle estimator proposed by Guyon (1982) shares the same asymptotic distribution as the modified GMLE $\widetilde{\boldsymbol{\theta}}$, which may be seen via a similar argument as in p.386-7 of Brockwell and Davis (1991).

In the remainder of this section, we always assume that the condition of Theorem 2 holds. Further, we only consider the case of condition (C2)(i) in the derivation below. The two other cases can be dealt with in a similar manner. We introduce some notation first. For $\mathbf{s} \notin \mathcal{S}_{N_1,N_2} \equiv$

$\{(u,v):\ 1\le u\le N_1, 1\le v\le N_2\}$, let $\check X(\mathbf{s})\equiv Z(\mathbf{s})\equiv 0$. For $\mathbf{s}\in\mathcal{S}_{N_1,N_2}$, let $\check X(\mathbf{s})=X(\mathbf{s})$, and

$$\begin{aligned} Z(\mathbf{s})\equiv Z(\mathbf{s},\boldsymbol\theta) &= \check X(\mathbf{s})-\sum_{\mathbf{i}\in\mathcal{I}_1} b_{\mathbf{i}}\check X(\mathbf{s}-\mathbf{i})-\sum_{\mathbf{i}\in\mathcal{I}_2} a_{\mathbf{i}} Z(\mathbf{s}-\mathbf{i}) \\ &= \check X(\mathbf{s})-\sum_{l=1}^{p} b_{\mathbf{j}_l}\check X(\mathbf{s}-\mathbf{j}_l)-\sum_{m=1}^{q} a_{\mathbf{i}_m} Z(\mathbf{s}-\mathbf{i}_m). \end{aligned} \tag{4.3}$$

Let $\mathcal{Y}=\{X(\mathbf{t}_1),\cdots,X(\mathbf{t}_{N^*})\}^\tau$ and $\mathcal{Z}=\{Z(\mathbf{t}_1,\boldsymbol\theta_0),\cdots,Z(\mathbf{t}_{N^*},\boldsymbol\theta_0)\}^\tau$. We write for $1\le l\le p$ and $1\le m\le q$

$$U_l(\mathbf{s})\equiv U_l(\mathbf{s},\boldsymbol\theta)=-\frac{\partial Z(\mathbf{s})}{\partial b_{\mathbf{j}_l}}\quad\text{and}\quad V_m(\mathbf{s})\equiv V_m(\mathbf{s},\boldsymbol\theta)=-\frac{\partial Z(\mathbf{s})}{\partial a_{\mathbf{i}_m}}. \tag{4.4}$$

It is easy to see from (4.3) that

$$U_l(\mathbf{s})=a(\mathbf{B})^{-1}\check X(\mathbf{s}-\mathbf{j}_l),\quad\text{and}\quad V_m(\mathbf{s})=a(\mathbf{B})^{-1}Z(\mathbf{s}-\mathbf{i}_m)=a(\mathbf{B})^{-2}b(\mathbf{B})\check X(\mathbf{s}-\mathbf{i}_m). \tag{4.5}$$

Let $\mathcal{X}_1$ and $\mathcal{U}_1$ be $N^*\times p$ matrices with, respectively, $X(\mathbf{t}_l-\mathbf{j}_m)$ and $U_m(\mathbf{t}_l,\boldsymbol\theta_0)$ as the (l,m)-th element, and let $\mathcal{X}_2$ and $\mathcal{U}_2$ be $N^*\times q$ matrices with, respectively, $X(\mathbf{t}_l-\mathbf{i}_m)$ and $V_m(\mathbf{t}_l,\boldsymbol\theta_0)$ as the (l,m)-th element. Write $\mathcal{X}=(\mathcal{X}_1,\mathcal{X}_2)$ and $\mathcal{U}=(\mathcal{U}_1,\mathcal{U}_2)$. Let $\mathcal{R}=\text{diag}\{r(\mathbf{t}_1,\boldsymbol\theta_0),\cdots,r(\mathbf{t}_{N^*},\boldsymbol\theta_0)\}^\tau$, where $r(\cdot)$ was defined in (3.7).

Lemma 6. For $\widehat X(\mathbf{s})$ as defined in (2.14) and $k=1,\cdots,p+q$, it holds that

$$N^{-1/2}\left\{\left|\frac{\partial}{\partial\theta_k}\sum_{m=1}^{N^*}\log r(\mathbf{t}_m)\right|+\left|\sum_{m=1}^{N^*}\frac{\{X(\mathbf{t}_m)-\widehat X(\mathbf{t}_m)\}^2}{r(\mathbf{t}_m)^2}\frac{\partial r(\mathbf{t}_m)}{\partial\theta_k}\right|\right\}_{\boldsymbol\theta=\widetilde{\boldsymbol\theta}}\xrightarrow{P}0.$$

Proof. Let $\{Y(\mathbf{s})\}$ be the same process as defined in the proof of Lemma 3. Write for $\mathbf{s}=(u,v)\in\ \mathcal{I}^*$

$$\begin{aligned}\boldsymbol\gamma_{\mathbf{s}} \equiv \boldsymbol\gamma_{\mathbf{s}}(\boldsymbol\theta) = \{&\gamma(0,1),\gamma(0,2),\cdots,\gamma(0,v-1),\gamma(1,-(N_2-v)),\gamma(1,-(N_2-v)+1),\\ &\cdots,\gamma(1,v-1),\gamma(2,-(N_2-v)),\cdots,\gamma(u-1,v-1)\}^\tau.\end{aligned}$$

For $\boldsymbol\varphi_{\mathbf{s}}$ defined as in (3.5) (see also (3.4)), it follows from (3.6) that

$$\boldsymbol\varphi_{\mathbf{s}}=\boldsymbol\Sigma_{\mathbf{s}}^{-1}\boldsymbol\gamma_{\mathbf{s}}, \tag{4.6}$$

where $\boldsymbol\Sigma_{\mathbf{s}}=\text{Var}(\mathbf{Y}_{\mathbf{s}})$ and $\mathbf{Y}_{\mathbf{s}}$ is defined as in (3.18). It follows from (3.8) that

$$\begin{aligned} 1 &= \gamma(0,0)-\sum_{k=1}^{\infty}\varphi_{0k}\gamma(0,k)-\sum_{j=1}^{\infty}\sum_{k=-\infty}^{\infty}\varphi_{jk}\gamma(j,k) \\ &= \gamma(0,0)-\sum_{k=1}^{v-1}\varphi_{0k}\gamma(0,k)-\sum_{j=1}^{u-1}\sum_{k=-(N_2-v)}^{v-1}\varphi_{jk}\gamma(j,k)-\zeta_{\mathbf{s}}(0,0), \end{aligned} \tag{4.7}$$

and for $(l,m) \in \mathcal{B}_{\mathbf{s}} \equiv \{(0,m):\, 1 \le m < v\} \cup \{(l,m):\, 1 \le l < u, -(N_2-v) \le m < v\}$,

$$\begin{aligned}
\gamma(l,m) &= \sum_{k=1}^{\infty} \varphi_{0k}\gamma(l,m-k) + \sum_{j=1}^{\infty}\sum_{k=-\infty}^{\infty} \varphi_{jk}\gamma(l-j,m-k) \\
&= \sum_{k=1}^{v-1} \varphi_{0k}\gamma(l,m-k) + \sum_{j=1}^{u-1}\sum_{k=-(N_2-v)}^{v-1} \varphi_{jk}\gamma(l-j,m-k) + \zeta_{\mathbf{s}}(l,m),
\end{aligned} \tag{4.8}$$

where $\zeta_{\mathbf{s}}(l,m) = \sum_{\mathbf{i}\in\mathcal{A}_{\mathbf{s}}} \varphi_{\mathbf{i}}\gamma(l-j,m-k)$ and $\mathcal{A}_{\mathbf{s}}$ is given in (3.2). By (3.7) and (4.7), $r(\mathbf{s}) = \boldsymbol{\gamma}_{\mathbf{s}}^{\tau}(\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}} - \boldsymbol{\varphi}_{\mathbf{s}}) + 1 + \zeta_{\mathbf{s}}(0,0)$, where $\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}}$ is given in (3.3). Thus

$$\frac{\partial r(\mathbf{s})}{\partial\theta_k} = \frac{\partial\boldsymbol{\gamma}_{\mathbf{s}}^{\tau}}{\partial\theta_k}(\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}} - \boldsymbol{\varphi}_{\mathbf{s}}) + \boldsymbol{\gamma}_{\mathbf{s}}^{\tau}\frac{\partial(\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}} - \boldsymbol{\varphi}_{\mathbf{s}})}{\partial\theta_k} + \frac{\partial\zeta_{\mathbf{s}}(0,0)}{\partial\theta_k}. \tag{4.9}$$

Write

$$\begin{aligned}
\boldsymbol{\zeta}_{\mathbf{s}} \;\equiv\; \boldsymbol{\zeta}_{\mathbf{s}}(\boldsymbol{\theta}) \;=\; &\{\zeta_{\mathbf{s}}(0,1), \zeta_{\mathbf{s}}(0,2), \cdots, \zeta_{\mathbf{s}}(0,v-1), \zeta_{\mathbf{s}}(1,-(N_2-v)), \zeta_{\mathbf{s}}(1,-(N_2-v)+1), \\
&\cdots, \zeta_{\mathbf{s}}(1,v-1), \zeta_{\mathbf{s}}(2,-(N_2-v)), \cdots, \zeta_{\mathbf{s}}(u-1,v-1)\}^{\tau}.
\end{aligned}$$

Then (4.8) implies that $\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}} = \boldsymbol{\Sigma}_{\mathbf{s}}^{-1}(\boldsymbol{\gamma}_{\mathbf{s}} - \boldsymbol{\zeta}_{\mathbf{s}})$. Together with (4.6), we have

$$\frac{\partial(\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}} - \boldsymbol{\varphi}_{\mathbf{s}})}{\partial\theta_k} = -\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\frac{\partial\boldsymbol{\Sigma}_{\mathbf{s}}}{\partial\theta_k}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\boldsymbol{\zeta}_{\mathbf{s}} - \boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\frac{\partial\boldsymbol{\zeta}_{\mathbf{s}}}{\partial\theta_k}. \tag{4.10}$$

From (4.9) and (4.6), we find that

$$\frac{\partial r(\mathbf{s})}{\partial\theta_k} = \{\frac{\partial\boldsymbol{\gamma}_{\mathbf{s}}^{\tau}}{\partial\theta_k} + \boldsymbol{\varphi}_{\mathbf{s}}^{\tau}\frac{\partial\boldsymbol{\Sigma}_{\mathbf{s}}}{\partial\theta_k}\}(\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}} - \boldsymbol{\varphi}_{\mathbf{s}}) - \boldsymbol{\varphi}_{\mathbf{s}}^{\tau}\{\frac{\partial\boldsymbol{\Sigma}_{\mathbf{s}}}{\partial\theta_k}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\boldsymbol{\zeta}_{\mathbf{s}} + \frac{\partial\boldsymbol{\zeta}_{\mathbf{s}}}{\partial\theta_k}\} + \frac{\partial\zeta_{\mathbf{s}}(0,0)}{\partial\theta_k}.$$

Now by the Cauchy-Schwarz inequality,

$$|\frac{\partial r(\mathbf{s})}{\partial\theta_k}| \le \left\{||\frac{\partial\boldsymbol{\gamma}_{\mathbf{s}}}{\partial\theta_k}|| + C_1||\boldsymbol{\varphi}_{\mathbf{s}}||\right\}||\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}} - \boldsymbol{\varphi}_{\mathbf{s}}|| + C_2||\boldsymbol{\varphi}_{\mathbf{s}}||\left\{||\boldsymbol{\zeta}_{\mathbf{s}}|| + ||\frac{\partial\boldsymbol{\zeta}_{\mathbf{s}}}{\partial\theta_k}||\right\} + |\frac{\partial\zeta_{\mathbf{s}}(0,0)}{\partial\theta_k}|, \tag{4.11}$$

where $C_1, C_2 \in (0,\infty)$ are some constants. The existence of C_1 and C_2 is guaranteed by Lemma 5 and Remark 2. Note that

$$|\zeta_{\mathbf{s}}(l,m)|^2 \;\le\; \Big(\sum_{\mathbf{i}\in\mathcal{A}_{\mathbf{s}}} \varphi_{\mathbf{i}}^2\Big)\Big\{\sum_{\mathbf{i}\in\mathcal{A}_{\mathbf{s}}} \gamma(l-j,m-k)^2\Big\}.$$

Since $\gamma(\cdot)$ decays at an exponential rate (Remark 1(v)), it may be shown that $\sum_{(l,m)\in\mathcal{B}\mathbf{s},\mathbf{i}\in\mathcal{A}_{\mathbf{s}}} \gamma(l-j,m-k)^2 < \infty$. Hence for some constant $\alpha \in (0,1)$, it holds that

$$||\boldsymbol{\zeta}_{\mathbf{s}}||^2 \le C_3 \sum_{\mathbf{i}\in\mathcal{A}_{\mathbf{s}}} \varphi_{\mathbf{i}}^2 \le C_4(\alpha^u + \alpha^v + \alpha^{N_2-v}). \tag{4.12}$$

Note Remark 1(vi). It also holds that

$$||\frac{\partial \boldsymbol{\zeta}_{\mathbf{s}}}{\partial \theta_k}|| \le C_5(\alpha^u + \alpha^v + \alpha^{N_2-v}). \tag{4.13}$$

By (3.12) and (4.11), we have that

$$|\frac{\partial r(\mathbf{s}, \boldsymbol{\theta})}{\partial \theta_k}| \le C(\boldsymbol{\theta})\{\alpha(\boldsymbol{\theta})^u + \alpha(\boldsymbol{\theta})^v + \alpha(\boldsymbol{\theta})^{N_2-v}\},$$

where $C(\cdot) \in (0, \infty)$ and $\alpha(\cdot) \in (0,1)$ are continuous. By Lemma 1, there exists a subset of the sample space A with $P(A) > 1-\epsilon$ and $||\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}|| < \epsilon$ on A for all sufficiently large N. Therefore there exist constants $C_1 \in (0,\infty)$ and $\alpha_1 \in (0,1)$ for which $|\partial r(\mathbf{s}, \boldsymbol{\theta})/\partial \theta_k|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \le C_1(\alpha_1^u + \alpha_1^v + \alpha_1^{N_2-v})$ on A. Since $r(\mathbf{s}) \ge 1$ for all $\boldsymbol{\theta} \in \Theta$, it holds on the set A that

$$\begin{aligned} \frac{1}{N^{1/2}} \left| \frac{\partial}{\partial \theta_k} \sum_{m=1}^{N^*} \log r(\mathbf{t}_m) \right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} &\le \frac{C_1}{N^{1/2}} \sum_{u=n_1}^{N_1} \sum_{v=n_2}^{N_2-n_2} (\alpha_1^u + \alpha_1^v + \alpha_1^{N_2-v}) \\ &\le \frac{C}{N^{1/2}}(N_2\alpha_1^{n_1} + N_1\alpha_1^{n_2} + N_1\alpha_1^{N_2-n_2}), \end{aligned} \tag{4.14}$$

which converges to 0 under condition (C2). Thus $N^{-1/2}\left|\frac{\partial}{\partial\theta_k}\sum_{m=1}^{N^*}\log r(\mathbf{t}_m)\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \xrightarrow{P} 0.$

On the other hand,

$$\begin{aligned} & N^{-1/2} E\left(\left|\sum_{m=1}^{N^*} \frac{\{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m)\}^2}{t(\mathbf{t}_m)^2} \frac{\partial r(\mathbf{t}_m)}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} I(A)\right) \\ \le\ & C_1 N^{-1/2} E\left(\sum_{u=n_1}^{N_1} \frac{\{X(u,v) - \widehat{X}(u,v)\}^2}{t(u,v)^2}(\alpha_1^u + \alpha_1^v + \alpha_1^{N_2-v})\right) \\ & \le C N^{-1/2} \sum_{u=n_1}^{N_1} \sum_{v=n_2}^{N_2} (\alpha_1^u + \alpha_1^v + \alpha_1^{N_2-v}) \to 0. \end{aligned}$$

Thus the required result holds. ■

Lemma 7. For $k = 1, \cdots, p+q$,

$$\begin{aligned} N^{-1/2} \sum_{m=1}^{N^*} \Bigg\{ & \left| \frac{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)}{r(\mathbf{t}_m)} \frac{\partial\{\widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)\}}{\theta_k} \right| \\ & + \left| \frac{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m) - Z(\mathbf{t}_m)}{r(\mathbf{t}_m)} \frac{\partial\{\widehat{X}(\mathbf{t}_m) - Z(\mathbf{t}_m)\}}{\theta_k} \right| \Bigg\}_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \xrightarrow{P} 0. \end{aligned}$$

Proof. We only prove that $N^{-1/2}\sum_{m=1}^{N^*}\left|\frac{X(\mathbf{t}_m)-\widehat{X}(\mathbf{t}_m)+Z(\mathbf{t}_m)}{r(\mathbf{t}_m)}\frac{\partial\{\widehat{X}(\mathbf{t}_m)+Z(\mathbf{t}_m)\}}{\theta_k}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \xrightarrow{P} 0$, since the other half may be proved in a similar and simpler manner.

It follows from (4.3) that for $\mathbf{s} = (u, v) \in \mathcal{I}^*$,

$$
\begin{aligned}
Z(\mathbf{s}) &= a(\mathbf{B})^{-1} b(\mathbf{B}) X(\mathbf{s}) \\
&= X(\mathbf{s}) - \sum_{k=1}^{v-1} \varphi_{0k} X(u, v-k) - \sum_{j=1}^{u-1} \sum_{k=-(N_2-v)}^{v-1} \varphi_{jk} X(u-j, v-k) = X(\mathbf{s}) - \widetilde{\boldsymbol{\varphi}}_{\mathbf{s}}^{\tau} \mathbf{X}_{\mathbf{s}},
\end{aligned}
$$

where $\widetilde{\boldsymbol{\varphi}}_{\mathbf{s}}$ is given in (3.3) and $\mathbf{X}_{\mathbf{s}}$ is defined in the same way as $\mathbf{Y}_{\mathbf{s}}$ in (3.18). Since $\widehat{X}(\mathbf{s}) = \boldsymbol{\varphi}_{\mathbf{s}}^{\tau} \mathbf{X}_{\mathbf{s}}$,

$$
E\left(\frac{\partial \{\widehat{X}(\mathbf{s}) + Z(\mathbf{s})\}}{\partial \theta_k} \right)^2 = \sigma^2 \frac{\partial (\boldsymbol{\varphi}_{\mathbf{s}} - \widehat{\boldsymbol{\varphi}}_{\mathbf{s}})^{\tau}}{\partial \theta_k} \boldsymbol{\Sigma}_{\mathbf{s}} \frac{\partial (\boldsymbol{\varphi}_{\mathbf{s}} - \widehat{\boldsymbol{\varphi}}_{\mathbf{s}})}{\partial \theta_k} \leq C || \frac{\partial (\boldsymbol{\varphi}_{\mathbf{s}} - \widehat{\boldsymbol{\varphi}}_{\mathbf{s}})}{\partial \theta_k} ||^2.
$$

Note that for any symmetric matrices $\mathbf{A}_1, \mathbf{A}_2$,

$$
\mathbf{x}^{\tau} \mathbf{A}_1 \mathbf{A}_2 \mathbf{A}_1 \mathbf{x} \ \leq \ \lambda_{\max}(\mathbf{A}_2) \, ||\mathbf{A}_1 \mathbf{x}||^2 \ \leq \ \lambda_{\max}(\mathbf{A}_2) \, \{\lambda_{\max}(\mathbf{A}_1)\}^2 \, ||x||^2,
$$

where $\lambda_{\max}(\mathbf{A})$ denotes the maximal eigenvalue of $\mathbf{A}$. It follows from (4.10) that the RHS of the above expression is not greater than

$$
C(||\boldsymbol{\varphi}_{\mathbf{s}} - \widehat{\boldsymbol{\varphi}}_{\mathbf{s}}||^2 + ||\boldsymbol{\zeta}_{\mathbf{s}}||^2 + ||\frac{\partial \boldsymbol{\zeta}_{\mathbf{s}}}{\partial \theta_k}||) \leq C_1(\alpha^u + \alpha^v + \alpha^{N_2 - v}),
$$

see (3.12), (4.12) and (4.13). By the same argument as in the proof of Lemma 6, we may show that

$$
E\left(\frac{\partial \{\widehat{X}(\mathbf{s}) + Z(\mathbf{s})\}}{\partial \theta_k} \Big|_{\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}} I(A) \right)^2 \leq C(\alpha^u + \alpha^v + \alpha^{N_2 - v}),
$$

where A is an event with probability close to 1. Now by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
& N^{-1/2} \sum_{m=1}^{N^*} E\Big(\Big| \frac{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)}{r(\mathbf{t}_m)} \frac{\partial \{\widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)\}}{\theta_k} \Big|_{\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}} I(A) \Big)^2 \\
\leq \ & N^{-1/2} \sum_{m=1}^{N^*} \Big[E\Big\{ X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m) \Big\}^2 E\Big(\frac{\partial \{\widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)\}}{\theta_k} \Big|_{\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}} I(A) \Big)^2 \Big]^{1/2} \\
\leq \ & C_2 \sum_{u=n_1}^{N_1} \sum_{v=n_2}^{N_2} (\alpha^u + \alpha^v + \alpha^{N_2 - v})^{1/2} \to 0.
\end{aligned}
$$

Thus the required limit holds. ■

Lemma 8. $N^{-1} \mathcal{U}^{\tau} \mathcal{R}^{-1} \mathcal{U} \xrightarrow{P} \sigma^2 \mathbf{W}(\theta_0)^{-1}$.

Proof. Within this proof, all $U_l(\mathbf{s}), V_m(\mathbf{s}), Z(\mathbf{s})$ and $r(\mathbf{s})$ are defined at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Let $a_0(\mathbf{z})^{-1} = 1 - \sum_{\mathbf{i}>0} d_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}$, and $b_0(\mathbf{z})/a_0(\mathbf{z})^2 = 1 - \sum_{\mathbf{i}>0} c_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}$. Then the coefficients d_{jk} and c_{jk} decay at an exponential rate (see (2.7)). It follows from (4.5) that

$$U_l(\mathbf{s}) = a_0(\mathbf{B})^{-1}X(\mathbf{s}-\mathbf{j}_l) + \sum_{\mathbf{i}\in\mathcal{A}_{\mathbf{s}-\mathbf{j}_l}} d_{\mathbf{i}}X(\mathbf{s}-\mathbf{j}_l-\mathbf{i}) \tag{4.15}$$
$$= b_0(\mathbf{B})^{-1}\varepsilon(\mathbf{s}-\mathbf{j}_l) + \sum_{\mathbf{i}\in\mathcal{A}_{\mathbf{s}-\mathbf{j}_l}} d_{\mathbf{i}}X(\mathbf{s}-\mathbf{j}_l-\mathbf{i}) \equiv \widetilde{U}_l(\mathbf{s}) + u_l(\mathbf{s}),$$
$$V_m(\mathbf{s}) = a_0(\mathbf{B})b_0(\mathbf{B})X(\mathbf{s}-\mathbf{i}_m) + \sum_{\mathbf{i}\in\mathcal{A}_{\mathbf{s}-\mathbf{i}_m}} c_{\mathbf{i}}X(\mathbf{s}-\mathbf{i}_m-\mathbf{i}) \tag{4.16}$$
$$= a_0(\mathbf{B})^{-1}\varepsilon(\mathbf{s}-\mathbf{i}_m) + \sum_{\mathbf{i}\in\mathcal{A}_{\mathbf{s}-\mathbf{i}_m}} c_{\mathbf{i}}X(\mathbf{s}-\mathbf{i}_m-\mathbf{i}) \equiv \widetilde{V}_m(\mathbf{s}) + v_m(\mathbf{s}),$$

where $\mathcal{A}_{\mathbf{s}}$ is defined in (3.2). By an argument similar to the one used for (3.12) we may show that for $\mathbf{s}-\mathbf{j}_l = (\mu,\nu)$ and $\mathbf{s}-\mathbf{i}_m = (\zeta,\beta)$,

$$E\{u_l(\mathbf{s})^2\} \le C(\alpha^\mu + \alpha^\nu + \alpha^{N_2-\nu}), \qquad E\{v_l(\mathbf{s})^2\} \le C(\alpha^\zeta + \alpha^\beta + \alpha^{N_2-\beta}),$$

$$E\big|\widetilde{U}_l(\mathbf{s})u_l(\mathbf{s})\big| \le \big[E\{\widetilde{U}_l(\mathbf{s})^2\}E\{u_l(\mathbf{s})^2\}\big]^{1/2} \le C\{\alpha^\mu+\alpha^\nu+\alpha^{N_2-\nu}\}^{1/2} \le C\{\alpha^{\mu/2}+\alpha^{\nu/2}+\alpha^{(N_2-\nu)/2}\},$$

and

$$E\big|\widetilde{V}_m(\mathbf{s})v_m(\mathbf{s})\big| \le \big[E\{\widetilde{V}_m(\mathbf{s})^2\}E\{v_m(\mathbf{s})^2\}\big]^{1/2} \le C\{\alpha^\zeta+\alpha^\beta+\alpha^{N_2-\beta}\}^{1/2} \le C\{\alpha^{\zeta/2}+\alpha^{\beta/2}+\alpha^{(N_2-\beta)/2}\},$$

where $\alpha \in (0,1)$ is a constant. Consequently the (l,m)-th element of $\mathcal{U}_1^\tau\mathcal{R}^{-1}\mathcal{U}_1/N$ may be expressed as

$$\frac{1}{N}\sum_{d=1}^{N^*} U_l(\mathbf{t}_d)U_m(\mathbf{t}_d)/r(\mathbf{t}_d) = \frac{1}{N}\sum_{d=1}^{N^*} \widetilde{U}_l(\mathbf{t}_d)\widetilde{U}_m(\mathbf{t}_d)/r(\mathbf{t}_d) + R_N, \tag{4.17}$$

where $E(R_N^2) < \epsilon$ for all sufficiently large N, and $\epsilon > 0$ is any given constant (see (4.14)).

Let $b_0(\mathbf{z}) = 1 + \sum_{\mathbf{i}>0} h_{\mathbf{i}}\mathbf{z}^{\mathbf{i}}$, $\mathbf{t}_d = (\alpha_d, \beta_d)$ and $\mathbf{j}_l = (u_l, v_l)$. Then

$$\widetilde{U}_l(\mathbf{t}_d) = \varepsilon(\alpha_d-u_l, \beta_d-v_l) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} h_{jk}\varepsilon(\alpha_d-u_l-j, \beta_d-v_l-k) \tag{4.18}$$
$$= \varepsilon(\alpha_d-u_l, \beta_d-v_l) + \sum_{j=0}^{n_1}\sum_{k=-n_1}^{n_1} h_{jk}\varepsilon(\alpha_d-u_l-j, \beta_d-v_l-k) + \widetilde{u}_l(\mathbf{t}_k)$$
$$= U_l^*(\mathbf{s}) + \widetilde{u}_l(\mathbf{t}_k), \quad \text{say}.$$

In the above expressions, we assume that $h_{00} = 1$ and $h_{0,-k} = 0$ for all $k > 0$. Similar to (4.17), we may choose n_1 sufficiently large such that

$$\frac{1}{N}\sum_{d=1}^{N^*} \widetilde{U}_l(\mathbf{t}_d)\widetilde{U}_m(\mathbf{t}_d)/r(\mathbf{t}_d) = \frac{1}{N}\sum_{d=1}^{N^*} U_l^*(\mathbf{t}_d)\widetilde{U}_m^*(\mathbf{t}_d)/r(\mathbf{t}_d) + R_N^* \tag{4.19}$$

with $E(R_N^*)^2 < \epsilon$ for any given $\epsilon > 0$. Now since $\{\varepsilon(\mathbf{s})\}$ are independent and identically distributed, it holds that

$$\begin{aligned}
&\frac{1}{N}\sum_{d=1}^{N^*} U_l^*(\mathbf{t}_d)\widetilde{U}_m^*(\mathbf{t}_d) \\
= &\sum_{j_1,j_2=0}^{n_1}\ \sum_{k_1,k_2=-n_1}^{n_2} h_{j_1,k_1}h_{j_2,k_2}\frac{1}{N}\sum_{d=1}^{N^*}\varepsilon(\alpha_d-u_l-j_1,\beta_d-v_l-v_1)\varepsilon(\alpha_d-u_m-j_2,\beta_d-v_m-v_2) \\
\xrightarrow{a.s.} &\ \sigma^2\sum_{j=0}^{n_1}\ \sum_{k=-n_1}^{n_1} h_{jk}h_{u_l-u_m+j,v_l-v_m+k},
\end{aligned}$$

which converges to $\sigma^2\text{Cov}(\xi_{\mathbf{j}_l},\xi_{\mathbf{j}_m})$ as $n_1 \to \infty$. Now combining this with (4.17), (4.19) and the fact that $r(\mathbf{s}) \to 1$, we have

$$\frac{1}{N}\sum_{d=1}^{N^*} U_l(\mathbf{t}_d)U_m(\mathbf{t}_d)/r(\mathbf{t}_d) \xrightarrow{P} \sigma^2\text{Cov}(\xi_{\mathbf{j}_l},\xi_{\mathbf{j}_m}).$$

Similar results hold for other elements in $\mathcal{U}^\tau\mathcal{R}^{-1}\mathcal{U}/N$. Thus the lemma holds. ■

Lemma 9. $N^{-1/2}\mathcal{U}^\tau\mathcal{R}^{-1}\mathcal{Z} \xrightarrow{D} N(0,\sigma^4\mathbf{W}(\boldsymbol{\theta}_0)^{-1})$.

Proof. Within this proof, all $U_l(\mathbf{s}), V_m(\mathbf{s}), Z(\mathbf{s})$ and $r(\mathbf{s})$ are defined at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. It follows from (4.3) and (2.8) that $Z(\mathbf{s}) = \varepsilon(\mathbf{s}) + z(\mathbf{s})$, where $z(\mathbf{s}) = \sum_{\mathbf{i}\in\mathcal{A}_\mathbf{s}}\varphi_\mathbf{i}X(\mathbf{s}-\mathbf{i})$ and $\mathcal{A}_\mathbf{s}$ is defined in (3.2). For $k = 1,\cdots,N^*$, let

$$\mathbf{U}_k = \{\widetilde{U}_1(\mathbf{t}_k),\cdots,\widetilde{U}_p(\mathbf{t}_k),\widetilde{V}_1(\mathbf{t}_k),\cdots,\widetilde{V}_q(\mathbf{t}_k)\}^\tau,$$

$$\mathbf{u}_k = \{u_1(\mathbf{t}_k),\cdots,u_p(\mathbf{t}_k),v_1(\mathbf{t}_k),\cdots,v_q(\mathbf{t}_k)\}^\tau,$$

where $\widetilde{U}_l, \widetilde{V}_m, u_l$ and v_m are defined in (4.15) and (4.16). Now

$$\frac{1}{N^{1/2}}\mathcal{U}^\tau\mathcal{R}^{-1}\mathcal{Z} = \frac{1}{N^{1/2}}\sum_{k=1}^{N^*}(\mathbf{U}_k+\mathbf{u}_k)\frac{\varepsilon(\mathbf{t}_k)+z(\mathbf{t}_k)}{r(\mathbf{t}_k)} = \frac{1}{N^{1/2}}\sum_{k=1}^{N^*}\mathbf{U}_k\varepsilon(\mathbf{t}_k)/r(\mathbf{t}_k) + o_p(1).$$

The last equality may be justified using the same argument as in the proof of Lemma 8.

Define $\mathcal{F}_k$ to be the σ-algebra generated by $\{\varepsilon(\mathbf{s}):\ \mathbf{s} < \mathbf{t}_{k+1}\}$ for $k = 1,\cdots,N^*-1$, and $\mathcal{F}_{N^*}$ generated by $\{\varepsilon(\mathbf{s}):\ \mathbf{s} \le \mathbf{t}_{N^*}\}$. Then $\mathcal{F}_{k-1} \subset \mathcal{F}_k$, $\mathbf{U}_k\varepsilon(\mathbf{t}_k)$ is $\mathcal{F}_k$-measurable, and

$$E\{\mathbf{U}_k\varepsilon(\mathbf{t}_k)|\mathcal{F}_{k-1}\} = \mathbf{U}_kE\{\varepsilon(\mathbf{t}_k)\} = 0.$$

Therefore $\{\mathbf{U}_k\varepsilon(\mathbf{t}_k)\}$ are martingale differences with respect to $\{\mathcal{F}_k\}$. Note that $r(\mathbf{s}) \ge 1$. For any $\epsilon > 0$ and $\boldsymbol{\alpha} \in \mathcal{R}^{p+q}$,

$$\begin{aligned}
&\frac{1}{N}\sum_{k=1}^{N^*} E\left\{ \left(\frac{\boldsymbol{\alpha}^\tau \mathbf{U}_k \varepsilon(\mathbf{t}_k)}{r(\mathbf{t}_k)}\right)^2 I\{|\boldsymbol{\alpha}^\tau \mathbf{U}_k \varepsilon(\mathbf{t}_k)/r(\mathbf{t}_k)| > N^{1/2}\epsilon\} \middle| \mathcal{F}_{k-1} \right\} \\
\leq\ & \frac{1}{N}\sum_{k=1}^{N^*} E[\{\boldsymbol{\alpha}^\tau \mathbf{U}_k \varepsilon(\mathbf{t}_k)\}^2 I\{|\boldsymbol{\alpha}^\tau \mathbf{U}_k \varepsilon(\mathbf{t}_k)| > N^{1/2}\epsilon\}\{I(|\boldsymbol{\alpha}^\tau \mathbf{U}_k| > \log N) + I(|\boldsymbol{\alpha}^\tau \mathbf{U}_k| \leq \log N)\}|\mathcal{F}_{k-1}] \\
\leq\ & \frac{\sigma^2}{N}\sum_{k=1}^{N^*} (\boldsymbol{\alpha}^\tau \mathbf{U}_k)^2 I(|\boldsymbol{\alpha}^\tau \mathbf{U}_k| > \log N) + \frac{1}{N}\sum_{k=1}^{N^*} (\boldsymbol{\alpha}^\tau \mathbf{U}_k)^2 E[\varepsilon(\mathbf{t}_k)^2 I\{|\varepsilon(\mathbf{t}_k)| > N^{1/2}\epsilon/\log N\}].
\end{aligned}$$

The first sum on the RHS of the above expression is, for all sufficiently large N, smaller than

$$\frac{\sigma^2}{N}\sum_{k=1}^{N^*} (\boldsymbol{\alpha}^\tau \mathbf{U}_k)^2 I(|\boldsymbol{\alpha}^\tau \mathbf{U}_k| > K)$$

which converges in probability, via an argument as in the proof of Lemma 8, to an arbitrarily small constant (by choosing K large enough but fixed). Therefore it converges to 0. In the same vein, the second sum also converges to 0 in probability. Note that

$$\frac{1}{N}\sum_{k=1}^{N^*} (\frac{\boldsymbol{\alpha}^\tau \mathbf{U}_k \varepsilon(\mathbf{t}_k)}{r(\mathbf{t}_t)})^2 \sim \frac{1}{N}\sum_{k=1}^{N^*} \{\boldsymbol{\alpha}^\tau \mathbf{U}_k \varepsilon(\mathbf{t}_k)\}^2 \xrightarrow{P} E\{\boldsymbol{\alpha}^\tau \mathbf{U}_1 \varepsilon(\mathbf{t}_1)\}^2 = \sigma^4 \boldsymbol{\alpha}^\tau \mathbf{W}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\alpha}.$$

It follows from Theorem 4 on p.511 of Shiryayev (1984) that

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{N^*} \boldsymbol{\alpha}^\tau \mathbf{U}_k \varepsilon(\mathbf{t}_k)/r(\mathbf{t}_t) \xrightarrow{D} N(0, \sigma^4 \boldsymbol{\alpha}^\tau \mathbf{W}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\alpha}), \quad \text{for any } \boldsymbol{\alpha} \in \mathcal{R}^{p+q}.$$

This leads to the required CLT. ■

Proof of Theorem 2. It follows from (2.20) that

$$\begin{aligned}
M(\boldsymbol{\theta}) \ &\equiv\ -2\sigma^2 \log L(\boldsymbol{\theta}, \sigma^2) = N^*\sigma^2 \log \sigma^2 + \sigma^2 \sum_{j=1}^{N^*} \log r(\mathbf{t}_j) + \sum_{j=1}^{N^*} \{X(\mathbf{t}_j) - \widehat{X}(\mathbf{t}_j)\}^2 / r(\mathbf{t}_j) \\
&=\ N^*\sigma^2 \log \sigma^2 + \sigma^2 \sum_{j=1}^{N^*} \log r(\mathbf{t}_j) + \sum_{j=1}^{N^*} \frac{Z(\mathbf{t}_j)^2}{r(\mathbf{s}_j)} + \sum_{j=1}^{N^*} \frac{\{X(\mathbf{t}_j) - \widehat{X}(\mathbf{t}_j)\}^2 - Z(\mathbf{t}_j)^2}{r(\mathbf{s}_j)},
\end{aligned}$$

where $Z(\cdot)$ is defined in (4.4). Note that $\widehat{\boldsymbol{\theta}}$ is the solution of the equation $\frac{\partial}{\partial \boldsymbol{\theta}} M(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0$. For $1 \leq k \leq p$, the equality $\frac{\partial}{\partial b_{\mathbf{j}_k}} M(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0$ leads to

$$0 = \sum_{m=1}^{N^*} Z(\mathbf{t}_m, \widehat{\boldsymbol{\theta}}) U_k(\mathbf{t}_m, \widehat{\boldsymbol{\theta}}) / r(\mathbf{t}_m, \widehat{\boldsymbol{\theta}}) + \delta_k \tag{4.20}$$

$$= \sum_{m=1}^{N^*} \{X(\mathbf{t}_m) - \sum_{\ell=1}^{p} \widehat{b}_{\mathbf{j}_\ell} X(\mathbf{t}_m - \mathbf{j}_\ell) - \sum_{\ell=1}^{q} \widehat{a}_{\mathbf{i}_\ell} Z(\mathbf{t}_m - \mathbf{i}_\ell, \boldsymbol{\theta}_0)\} \frac{U_k(\mathbf{t}_m, \boldsymbol{\theta}_0)}{r(\mathbf{t}_m, \boldsymbol{\theta}_0)} + \boldsymbol{\eta}_k^\tau (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \delta_k,$$

where

$$\delta_k = \left(\frac{\sigma^2}{2} \frac{\partial}{\partial b_{\mathbf{j}_k}} \sum_{m=1}^{N^*} \log r(\mathbf{t}_m) - \frac{1}{2} \sum_{m=1}^{N^*} \frac{\{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m)\}^2}{r(\mathbf{t}_m)^2} \frac{\partial r(\mathbf{t}_m)}{\partial b_{\mathbf{j}_k}} \right.$$
$$- \frac{1}{2} \sum_{m=1}^{N^*} \left\{ \frac{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)}{r(\mathbf{t}_m)} \frac{\partial \{\widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)\}}{\partial b_{\mathbf{j}_k}} \right.$$
$$\left.\left. + \frac{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m) - Z(\mathbf{t}_m)}{r(\mathbf{t}_m)} \frac{\partial \{\widehat{X}(\mathbf{t}_m) - Z(\mathbf{t}_m)\}}{\partial b_{\mathbf{j}_k}} \right\} \right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},$$

$$\boldsymbol{\eta}_k = \sum_{m=1}^{N^*} \frac{U_k(\mathbf{t}_m, \boldsymbol{\theta}_0)}{r(\mathbf{t}_m, \boldsymbol{\theta}_0)} \sum_{\ell=1}^{q} a_{\mathbf{i}_\ell,0} \mathbf{U}(\mathbf{t}_m - \mathbf{i}_\ell, \boldsymbol{\theta}_0) + \sum_{m=1}^{N^*} Z(\mathbf{t}_m, \boldsymbol{\theta}_0) \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{U_k(\mathbf{t}_m)}{r(\mathbf{t}_m)} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + O_p(N||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||), \tag{4.21}$$

where $a_{\mathbf{i}_\ell,0}$ denotes the true value of $a_{\mathbf{i}_\ell}$, and $\mathbf{U}(\mathbf{s}) = \{U_1(\mathbf{s}), \cdots, U_p(\mathbf{s}), V_1(\mathbf{s}), \cdots, V_q(\mathbf{s})\}^\tau$. Similarly the equation $\frac{\partial}{\partial a_{\mathbf{i}_k}} M(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0$ $(1 \le k \le q)$ leads to

$$0 = \sum_{m=1}^{N^*} \{X(\mathbf{t}_m) - \sum_{\ell=1}^{p} \widehat{b}_{\mathbf{j}_\ell} X(\mathbf{t}_m - \mathbf{j}_\ell) - \sum_{\ell=1}^{q} \widehat{a}_{\mathbf{i}_\ell} Z(\mathbf{t}_m - \mathbf{i}_\ell, \boldsymbol{\theta}_0)\} \frac{V_k(\mathbf{t}_m, \boldsymbol{\theta}_0)}{r(\mathbf{t}_m, \boldsymbol{\theta}_0)} + \boldsymbol{\eta}_{p+k}^\tau (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \delta_{p+k}, \tag{4.22}$$

where

$$\delta_{p+k} = \left(\frac{\sigma^2}{2} \frac{\partial}{\partial a_{\mathbf{i}_k}} \sum_{m=1}^{N^*} \log r(\mathbf{t}_m) - \frac{1}{2} \sum_{m=1}^{N^*} \frac{\{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m)\}^2}{r(\mathbf{t}_m)^2} \frac{\partial r(\mathbf{t}_m)}{\partial a_{\mathbf{i}_k}} \right.$$
$$- \frac{1}{2} \sum_{m=1}^{N^*} \left\{ \frac{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)}{r(\mathbf{t}_m)} \frac{\partial \{\widehat{X}(\mathbf{t}_m) + Z(\mathbf{t}_m)\}}{\partial a_{\mathbf{i}_k}} \right.$$
$$\left.\left. + \frac{X(\mathbf{t}_m) - \widehat{X}(\mathbf{t}_m) - Z(\mathbf{t}_m)}{r(\mathbf{t}_m)} \frac{\partial \{\widehat{X}(\mathbf{t}_m) - Z(\mathbf{t}_m)\}}{\partial a_{\mathbf{i}_k}} \right\} \right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},$$

$$\boldsymbol{\eta}_{p+k} = \sum_{m=1}^{N^*} \frac{U_k(\mathbf{t}_m, \boldsymbol{\theta}_0)}{r(\mathbf{t}_m, \boldsymbol{\theta}_0)} \sum_{\ell=1}^{q} a_{\mathbf{i}_\ell,0} \mathbf{U}(\mathbf{t}_m - \mathbf{i}_\ell, \boldsymbol{\theta}_0) + \sum_{m=1}^{N^*} Z(\mathbf{t}_m, \boldsymbol{\theta}_0) \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{V_k(\mathbf{t}_m)}{r(\mathbf{t}_m)} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + O_p(N||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||). \tag{4.23}$$

Now it follows from (4.20) and (4.21) that

$$\mathcal{U}^\tau \mathcal{R}^{-1} \mathcal{X} \widehat{\boldsymbol{\theta}} = \mathcal{U}^\tau \mathcal{R}^{-1} \mathcal{Y} + \mathbf{A}^\tau (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \boldsymbol{\delta}, \tag{4.24}$$

where $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_{p+q})^\tau$, and $\mathbf{A}$ is a $(p+q) \times (p+q)$ matrix with $\boldsymbol{\eta}_k$ as its k-th column. Note that $\mathcal{Y} - \mathcal{X}\boldsymbol{\theta}_0 = \mathcal{Z}$ and

$$\mathcal{U} = \mathcal{X} - \sum_{\ell=1}^{q} a_{\mathbf{i}_\ell,0} \begin{pmatrix} \mathbf{U}(\mathbf{t}_1 - \mathbf{i}_\ell, \boldsymbol{\theta}_0)^\tau \\ \vdots \\ \mathbf{U}(\mathbf{t}_{N^*} - \mathbf{i}_\ell, \boldsymbol{\theta}_0)^\tau \end{pmatrix}.$$

By (4.24), (4.20) and (4.22), we have

$$\mathcal{U}\mathcal{R}^{-1}\mathcal{U}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\mathcal{U}\mathcal{R}^{-1}\mathcal{Z}+\mathbf{A}_1^{\tau}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+\boldsymbol{\delta},$$

where $\mathbf{A}_1$ is a $(p+q)\times(p+q)$ matrix with the sum of the last two terms on the RHS of (4.21) as its k-th column for $k=1,\cdots,p$, and the sum of the last two terms on the RHS of (4.23) as its $(p+k)$-th column for $k=1,\cdots,q$. Hence

$$N^{1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\{\mathcal{U}\mathcal{R}^{-1}\mathcal{U}/N-\mathbf{A}_1^{\tau}/N\}^{-1}N^{-1/2}(\mathcal{U}\mathcal{R}^{-1}\mathcal{Z}-\boldsymbol{\delta})=\{\mathcal{U}\mathcal{R}^{-1}\mathcal{U}/N\}^{-1}N^{-1/2}\mathcal{U}\mathcal{R}^{-1}\mathcal{Z}+o_p(1).$$

The last equality follows from the fact that $N^{-1/2}\boldsymbol{\delta}\xrightarrow{P}0$ and $\mathbf{A}_1/N\xrightarrow{P}0$. The former is guaranteed by Lemmas 6 and 7, and the latter follows from Theorem 1 and a similar argument as in the proof of Lemma 8. Now the theorem follows from Lemmas 8 and 9 immediately. ■

5. Final remarks

5.1. Edge effect correction

So far the asymptotic normality of the estimators for stationary spatial processes has been established via different edge-effect corrections; see Guyon (1982), Dahlhaus and Künsch (1987) and also Theorem 2 above. Whether such a correction is essential or not for the asymptotic normality of the GMLE remains as an open problem, although we would think that the answer should be negative. However it seems to us that an edge-effect correction would be necessary to ensure that the GMLE has the simple asymptotic distribution stated in Theorem 2 which is distribution-free.

For Gaussian processes, Guyon showed that his estimator is asymptotically efficient in a certain sense; see p.101 of Guyon (1982). Note that in the context of estimating coefficients of a spatial Gaussian ARMA process, Guyon's estimator, Dahlhaus and Künsch's estimator and our modified GMLE share the same asymptotic distribution as stated in Theorem 2. However, as far as we can see, Guyon's efficiency does not imply that these estimators will share the same asymptotic distribution with the *genuine* (Gaussian) MLE. This requires, in addition to what has been proved in Guyon (1982), the necessary condition

$$(N_1N_2)^{-1/2}\left|\left|\frac{\partial}{\partial\boldsymbol{\theta}}\{l(\boldsymbol{\theta})-l^*(\boldsymbol{\theta})\}\right|\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\xrightarrow{P}0,$$

where $l(\cdot)$ denotes the log Gaussian likelihood function and $l^*(\cdot)$ denotes the approximation from which the estimator is derived. From the derivation in §4, the above limit seems *unlikely* to hold for, at least, our edge-effect-corrected likelihood. (Note that the Whittle likelihood for spatial processes automatically suppresses the inhomogeneity at the boundary points up to a non-negligible order.) It will be interesting to see the form of the asymptotic distribution of the MLE without any edge-effect-correction, which, to our knowledge, is unknown at present.

5.2. Spatio-temporal ARMA models

A serious drawback of spatial ARMA models is the artifact due to the enforced unilateral order, which rules out some simple and practically meaningful models from the class. (See, e.g. Besag 1974.) In fact, the half-plane ordering is only appropriate for a few applications such as line-by-line image-scanning. Such a drawback may disappear naturally in the context of spatio-temporal modelling. To this end, let $X_t(\mathbf{s})$ denote the variable at time t and location $\mathbf{s}$. Now the index $(t, \mathbf{s})$ is three-dimensional. Under Whittle's half-plane unilateral order, the model

$$X_t(\mathbf{s}) = \sum_{\ell=1}^{p} \sum_{\mathbf{i} \in \mathcal{I}_\ell} b_{\ell,\mathbf{i}} X_{t-\ell}(\mathbf{s}-\mathbf{i}) + \varepsilon_t(\mathbf{s}) + \sum_{\ell=1}^{q} \sum_{\mathbf{i} \in \mathcal{J}_\ell} a_{\ell,\mathbf{i}} \varepsilon_{t-\ell}(\mathbf{s}-\mathbf{i})$$

is legitimate for any subsets $\mathcal{I}_\ell$ and $\mathcal{J}_\ell$ of $\mathcal{Z}^2$, since $X_t(\mathbf{s})$ depends only on its 'lagged' values, $\varepsilon_t(\mathbf{s})$ and the 'lagged' values of $\varepsilon_t(\mathbf{s})$. By letting $\mathcal{I}_\ell$ and $\mathcal{J}_\ell$ contain, for example, $(0,0)$ and its four nearest neighbours, the model is practically meaningful and can be used to model real data over space and time. This is in marked contrast to the spatial models (2.1) in which $\mathcal{I}_1$ and $\mathcal{I}_2$ must be some subsets of $\{\mathbf{s} > 0\}$. The asymptotic theory developed in this paper may be readily extended to deal with the above model.

Acknowledgements

QY's research was partially supported by a Leverhulme Trust grant and PJB's by NSF Grant DMS0308109. The authors thank Professors Rainer Dahlhaus, Peter Robinson and Howell Tong for helpful discussions, Professor Dag Tjøstheim for pointing out an error, and Miss Chrysoula Dimitriou-Fakalou for correcting many typos in an early version of the paper. The critical reading and helpful comments from an anonymous referee is gratefully acknowledged.

References

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Royal Statist. Soc.* **B**, **36**, 192-236.

Besag, J. (1975) Statistical analysis of non-lattice data. *The Statistician*, **24**, 179-195.

Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, New York.

Cressie, N.A.C. (1993) *Statistics for Spatial Data*. Wiley, New York.

Dahlhaus, R. and Künsch, H. (1987) Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, **74**, 877-882.

Goodman, D. (1977) Some stability properties of two-dimensional linear shift-invariant digital filters. *IEEE Transactions on Circuits and Systems*. **CAS-24**, 201-208.

Guyon, X. (1982) Parameter estimation for a stationary process on a d-dimensional Lattice. *Biometrika*, **69**, 95-105.

Guyon, X. (1995) *Random Fields on a Network*. Springer-Verlag, New York.

Hannan, E. J. (1973) The asymptotic theory of linear time-series models. *J. Appl. Prob.* **10**, 130-145.

Helson, H. and Lowdenslager, D. (1958) Prediction theory and Fourier series in several variables. *Acta Math.* **99**, 165-201.

Huang, D. and Anh, V.V. (1992) Estimation of spatial ARMA models. *Austral. J. Statist.* **34**, 513-530.

Justice, J.H. and Shanks, J.L. (1973) Stability criterion for N-dimensional digital filters. *IEEE Transactions on Automatic Control*. **AC-18**, 284-286.

Perera, G. (2001) Random fields on Z^d. limit theorems and irregular sets. In *Spatial Statistics: Methodological Aspects and Applications*, M. Moore (edit.), p.57-82. Springer-Verlag, New York.

Shiryayev, A.N. (1984) *Probability*. Springer-Verlag, New York.

Strintzis, M.G. (1977) Tests of stability of multidimensional filters. *IEEE Transactions on Circuits and Systems*. **CAS-24**, 432-437.

Tjøstheim, D. (1978) Statistical spatial series modelling. *Advances in Applied Probability*, **10**, 130-154.

Tjøstheim, D. (1983) Statistical spatial series modelling II: some further results in unilateral processes. *Advances in Applied Probability*, **15**, 562-584.

Whittle, P. (1954) On stationary processes in the plane. *Biometrika*, **41**, 434-449.

Whittle, P. (1962) Gaussian estimation in stationary time series. *Bull. Inst. Internat. Statist. 39*, 105-129.

Wood, A.T.A. and Chan, G. (1994) Simulation of Stationary Gaussian Processes in $[0,1]^d$. *Journal of Computational and Graphical Statistics*, **3**, 409-432.

Yao, Q. and Brockwell, P.J. (2005) Gaussian maximum likelihood estimation for ARMA models I: time series. Revised for *Journal of Time Series Analysis* .