Asymptotic Properties of Maximum Likelihood Estimates in a Class of Space–Time Regression Models

XUFENG NIU*

Florida State University

For statistical analyses of satellite ozone data, Niu and Tiao introduced a class of space–time regression models which took into account temporal and spatial dependence of the observations. In this paper, asymptotic properties of maximum likelihood estimates of parameters in the models are considered. The noise terms in the space-time regression models are in fact structural periodic vector autoregressive processes. Some properties of the spectral density matrix of the processes are discussed. Under mild conditions, the strong law of large numbers and the central limit theorem for the parameter estimates are proven. © 1995 Academic Press, Inc.

1. INTRODUCTION

Environmental data analysis is a field that today is of critical importance in providing the basis of scientific understanding for setting wise public policy on environmental management. With the huge amount of ozone, temperature, aerosol, and gas data collected by satellite, sophisticated statistical techniques need to be developed to describe the structure of these data. Reinsel *et al* [14, 15] used regression-time series models to analysis ground-based and satellite ozone data. Niu and Tiao [12] introduced a class of space–time regression models for the statistical analyses of satellite ozone and temperature data, in which both spatial and temporal interactions of the observations were considered. Specifically, letting $\{y_j(t)\}$ be the monthly average ozone or temperature observations at longitude j on a fixed latitude, Niu and Tiao [12] considered the regression time-series model

$$y_j(t) = \sum_{k=1}^{m} \beta_{kj} x_k(t) + \xi_j(t), \tag{1.1}$$

82

where the $x_k(t)$'s were some regressors including a linear trend term and $\xi_j(t)$ was a noise component. The noise term $\xi_j(t)$ was further modeled in the space–time form

$$\xi_j(t) = \sum_{k=1}^{q} \left[\alpha_k \xi_{j-k}(t) + \theta_k \xi_{j+k}(t) \right] + \sum_{l=1}^{p} \phi_l \xi_j(t-l) + \varepsilon_j(t),$$

$$j = 1, 2, ..., n; \quad t = 1, 2, ..., T, \tag{1.2}$$

where the $\varepsilon_j(t)$'s were assumed to be independent and normally distributed random variables with mean zero and variance $\sigma^2(t) = \sigma^2(t - 12)$. The model in (1.2) is called a space–time autoregressive model with order q and p (STAR(q, p)). The spatial order q satisfies that $q \leqslant [n/2]$ and $\theta_q = 0$ for $q = n/2$, where $[n/2]$ is the integer part of $n/2$. From now on, we denote the unknown parameters in (1.1) and (1.2) by the following notation:

$$\beta = [\beta_{11}, ..., \beta_{m1}, ..., \beta_{1n}, ..., \beta_{mn}]', \quad \alpha = [\alpha_1, ..., \alpha_q]', \quad \theta = [\theta_1, ..., \theta_q]',$$

$$\phi = [\phi_1, ..., \phi_p]', \quad \sigma = [\sigma^2(1), ..., \sigma^2(12)]', \quad \mu = [\alpha', \theta', \phi', \sigma']'.$$

We call β the regression coefficients, α and θ the spatial coefficients, and ϕ the temporal coefficients.

The space–time model in (1.2) has several special features. In the temporal direction, the value of current observation depends upon the past p values, which is described by the well-knwon AR(p) model. Since the variations of ozone and temperature observations usually depend on month of the year, the $\varepsilon_j(t)$'s are assumed to have different variances for different months. In the spatial direction, the value of current observation depends on its neighboring values. This spatial bilateral dependence scheme is similar to the bilateral autoregression model introduced by Whittle [20] for spatial lattice systems. In practice, the spatially bilateral dependence makes sense physically, but the estimation of parameters in this type of model becomes considerably difficult. In particular, for an infinite two or higher dimensional lattice system, it is usually impossible to find a simple formula for the likelihood function of the parameters, and standard methods of statistical inference are very hard to apply to such models. These difficulties motivated the introduction of unilateral lattice process models. For example, Tjøstheim [16–18] studied the asymptotic theory of causal (quadrant-type) models and half-space models for high dimensional lattice systems. See Korezlioglu and Loubation [10] for further results in the half space case. However, for estimation of the parameters in the space–time model (1.2), there is no difficulty created by the bilateral dependence. One of basic features of satellite ozone and temperature data is their circular property, i.e., for a fixed latitude and at a fixed time t, the data

$\{y_j(t), j = 1, 2, ..., n\}$ were observed along a circle. Therefore the covariance structure of the space–time model in (1.2) is closely related to circular matrices. Using properties of circular matrices, such as explicit expressions of their eigenvalues and inverses, Niu and Tiao [12] gave an explicit formula for the conditional likelihood function of the parameters in the space–time regression models and discussed stationary conditions of the noise term $\{\xi_j(t)\}$. They showed that the space–time regression models could be written in transformed forms which may be used to test the uniformity of the long-term trends in different longitudinal ozone series, and they applied the space–time regression models to the total ozone mapping spectrometer (TOMS) ozone data for trend assessment.

In this paper, we investigate asymptotic poperties of the maximum likelihood estimates of the parameters in the space–time regression models. Define

$$\psi(z) = 1 - \sum_{k=1}^{q} (\alpha_k z^{-k} + \theta_k z^k), \qquad \xi(t) = [\xi_1(t), ..., \xi_n(t)]',$$

$$\varepsilon(t) = [\varepsilon_1(t), ..., \varepsilon_n(t)]', \qquad y(t) = [y_1(t), ..., y_n(t)]',$$

$$x(t) = [x_1(t), ..., x_m(t)]', \qquad X(t) = I_n \otimes x'(t)$$

and

$$W_n = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & & & & \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}_{n \times n},$$

where W_n is a circular and orthogonal matrix. Using the above notation, the models in (1.1) and (1.2) can be written in the form

$$y(t) = X(t)\,\beta + \xi(t), \tag{1.3}$$

$$\psi(W_n)\,\xi(t) = \sum_{l=1}^{p} \phi_l \xi(t-l) + \varepsilon(t), \tag{1.4}$$

where $\varepsilon(t)$ has periodic covariance matrix $\sigma^2(t)\,I_n$. Thus, the noise model in (1.2) is in fact a structural periodic vector autoregressive time series model with few natural parameters. Therefore some available asymptotic results and related techniques for maximum likelihood estimates of parameters in vector time series models will be used in this study.

Consistency and convergence properties of maximum likelihood estimators for vector time series models have been proved under various assumption. Dunsmuir and Hannan [4] and Deistler et al. [2] presented

consistency and central limit results for parameter estimates in general linear time series models and in ARMA models. Hannen *et al.* [9] presented the asymptotic results for the ARMAX case (auto-regressive-moving average with exogenous variables). The settings of general linear models in Dunsmuir and Hannan [4] were of the form

$$\xi(t) = \sum_{j=0}^{\infty} C(j, \mathbf{v}) \, \varepsilon(t - j), \qquad \sum_{j=1}^{\infty} \| C(j, \mathbf{v}) \|^2 < \infty, \qquad C(0, \mathbf{v}) = I_n, \qquad (1.5)$$

where $\|C\|$ denotes the norm of the matrix C, for example, the square root of the maximum eigenvalues of C^*C where "*" denotes the complex conjugate transpose; the unknown parameter $\mathbf{v} \in \bar{\Theta}$ and $\bar{\Theta}$ was assumed to be compact: $\{\varepsilon(t)\}$ was an uncorrelated vector series with common variance matrix $K(\mathbf{v})$. The spectral density function of $\{\xi(t)\}$ was assumed to have the form

$$f(\omega, \mathbf{v}) = (2\pi)^{-1} h(e^{i\omega}, \mathbf{v})^{-1} g(e^{i\omega}, \mathbf{v}) K(\mathbf{v}) g^*(e^{i\omega}, \mathbf{v})^{-1} h^*(e^{i\omega}, \mathbf{v})^{-1}, \qquad (1.6)$$

where $h(e^{i\omega}, \mathbf{v})$ and $g(e^{i\omega}, \mathbf{v})$ were continuous functions in $(\omega; \mathbf{v}) \in [-\pi, \pi] \times \bar{\Theta}$ and

$$h(0, \mathbf{v}) \equiv g(0, \mathbf{v}) \equiv I_n.$$

Dunsmuir and Hannan [4] imposed some basic conditions on the spectral density function and the parameter space and proved that estimators of $\mathbf{v}$, derived from a Gaussian likelihood and some certain spectral approximations to this, are strongly consistent and asymptotically normally distributed. For vector ARMA models, Dunsmuir and Hannan [4] claimed the conditions were fulfilled under special choices of parameter spaces. Hence the maximum likelihood estimates of parameters in a vector ARMA model converge almost surely to the true parameters. However, Pötscher [13] pointed out that the proof of one of the conditions, namely condition B6 in Dunsmuir and Hannan [4], was not conclusive and found that if one assumes Gaussianity of the vector processes, then the consistence results of the likelihood estimators are still valid without imposing the condition B6.

For regression time–series models, some asymptotic results of parameter estimates are indeed available in the literature. For example, Hannan [6, Chap. 8] discussed the asymptotic theory of the least squares estimates of parameters in vector linear processes. Hannan [7, 8] proved the strong consistency and the asymptotic normality of maximum likelihood estimates of parameters in univariate regression time series models. See also Whittle [21] and Walker [19]. In practice, conditional likelihood functions are often used as time domain approximations to the Gaussian likelihood for

estimation purpose when the number of observations is moderately large. However, fewer results are available for parameter estimates derived from the exact Gaussian likelihood and its time domain approximations in vector regression time–series models.

The asymptotic theory of the maximum (exact and conditional) likelihood estimates of the parameters β and μ in the space–time regression models will be addressed in this study. In Section 2, we give the estimation procedure of β and μ. In Section 3, we first show that if the $\{\xi_j(t)\}$'s are observable, the maximum conditional likelihood estimates of μ are strongly consistent. The proof draws on results and ideas of Dunsmuir and Hannan [4]. Then under some mild conditions on the regressors, we prove the strong law of large numbers for the estimates of β and μ. In Section 4, the central limit theorem for the maximal likelihood estimates of the parameters in the space–time regression model is studied.

2. The Maximum Likelihood Estimates

In this section, we give the estimation procedure of the parameters in the space–time regression model. Niu and Tiao [12] derived the exact likelihood function and the conditional likelihood function for the parameters. The notation and terminology of that paper will be used here. Define

$$\xi'_* = [\xi(p'), ..., \xi(1)'], \qquad \xi' = [\xi(p+1)', ..., \xi(T)'],$$

$$\varepsilon' = [\varepsilon(p+1)', ..., \varepsilon(T)'],$$

$$U_{T-p} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & & & & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}_{(T-p) \times (T-p)},$$

$$C = \begin{bmatrix} \phi_1 I_n & \phi_2 I_n & \cdots & \phi_{p-1} I_n & \phi_p I_n \\ \cdots & \cdots & & & \\ \phi_p I_n & O & \cdots & O & O \\ \cdots & \cdots & & & \\ O & O & \cdots & O & O \end{bmatrix}_{n(T-p) \times np},$$

and

$$A = I_{T-p} \otimes \psi(W_n), \qquad D = \sum_{k=1}^{p} U^k_{T-p} \otimes \phi_k I_n. \tag{2.1}$$

From now on, we will assume that the matrix $\psi(W_n)$ is invertible. Hence the model in (1.4) can also be written in the "reduced" form

$$\Phi(B)\,\xi(t) = \psi^{-1}(W_n)\,\varepsilon(t), \tag{2.2}$$

where B is the backward shift operator such that $B\xi(t) = \xi(t-1)$ and

$$\Phi(z) = I_n - \sum_{l=1}^{p} (\phi_l\psi^{-1}(W_n))\,z^l.$$

When the determinantal polynomial $|\Phi(z)|$ have zeros outside the unit circle, it is easy to show that the process $\xi(t)$ can be expressed as a causal function of $\{\varepsilon(t)\}$ (see Brockwell and Davis [1]) which has the form

$$\xi(t) = \sum_{j=0}^{\infty} \Psi_j[\psi^{-1}(W_n)\,\varepsilon(t-j)]. \tag{2.3}$$

Therefore $\xi(t)$ is independent of $\{\varepsilon(t+1),\ \varepsilon(t+2),\ ...\}$, which implies that ξ_* and ε are independent.

Denote the covariance matrix of ξ_* by Γ and the log exact likelihood function of the parameters in (1.2) by $l(\mu \mid \xi, \xi_*)$. Then by Niu and Tiao [12], maximizing $l(\mu \mid \xi, \xi_*)$ is equivalent to maximizing the function

$$l_T(\mu \mid \xi, \xi_*) = -\frac{1}{T}\log(|\Gamma|) - \frac{n}{T}\sum_{t=p+1}^{T}\log(\sigma^2(t)) + \frac{T-p}{T}\sum_{s=0}^{n-1}\log|\psi(e^{i2\pi s/n})|^2$$

$$-\frac{1}{T}[\xi_*'\,\Gamma^{-1}\xi_* + S(\xi, \xi_*)], \tag{2.4}$$

where

$$S(\xi, \xi_*) = \sum_{t=p+1}^{T}\left\{\frac{1}{\sigma^2(t)}\sum_{j=1}^{n}\left(\xi_j(t) - \sum_{k=1}^{q}[\alpha_k\xi_{j-k}(t)\right.\right.$$

$$\left.\left. + \theta_k\xi_{j+k}(t)] - \sum_{l=1}^{p}\phi_l\xi_j(t-l)\right)^2\right\}.$$

In practice, for moderately large T, $(T \gg p)$, we may approximate the likelihood function in (2.4) by a conditional likelihood function which ignores $-(1/T)\log(|\Gamma|)$ and $-(1/T)\xi_*'\,\Gamma^{-1}\xi_*$ in (2.4), i.e.,

$$\bar{l}_T(\mu \mid \xi, \xi_*) = -\frac{n}{T}\sum_{t=p+1}^{T}\log(\sigma^2(t)) + \frac{T-p}{T}$$

$$\times \sum_{s=0}^{n-1}\log|\psi(e^{i2\pi s/n})|^2 - \frac{1}{T}S(\xi, \xi_*). \tag{2.5}$$

Furthermore, define

$$\Sigma_{T-p} = \text{Diag}(\sigma^2(p+1), ..., \sigma^2(T)), \qquad \mathbf{Y} = [\,\mathbf{y}(1)', ..., \mathbf{y}(T)'\,]'.$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}(1) \\ \vdots \\ \mathbf{X}(T) \end{bmatrix},$$

$$F = [\,-C \vdots (A-D)\,]'\,(\Sigma_{T-p}^{-1} \otimes I_n)[\,-C \vdots (A-D)\,],$$

and

$$S(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\,F(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Then by Niu and Tiao [12], an equivalent form of the log likelihood function for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ is

$$l_T(\boldsymbol{\beta}, \boldsymbol{\mu} \mid \mathbf{Y}) = -\frac{1}{T}\log(|\Gamma|) - \frac{n}{T}\sum_{t=p+1}^{T} \log \sigma^2(t) + \frac{T-p}{T}\sum_{s=0}^{n-1} \log |\psi(e^{i2\pi s/n})|^2$$

$$-\frac{1}{T}[\,\xi_*'\Gamma^{-1}\xi_* + S(\mathbf{Y} - \mathbf{Y}\boldsymbol{\beta})\,]. \tag{2.6}$$

Similarly, an equivalent form of the log conditional likelihood function for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ is

$$\tilde{l}_T(\boldsymbol{\beta}, \boldsymbol{\mu} \mid \mathbf{Y}) = -\frac{n}{T}\sum_{t=p+1}^{T} \log \sigma^2(t) + \frac{T-p}{T}$$

$$\times \sum_{s=0}^{n-1} \log |\psi(e^{i2\pi s/n})|^2 - \frac{1}{T}S(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \tag{2.7}$$

Setting $\partial \tilde{l}_T(\boldsymbol{\beta}, \boldsymbol{\mu} \mid \mathbf{Y})/\partial\boldsymbol{\beta} = 0$, we have

$$\tilde{\boldsymbol{\beta}}_T = (X'FX)^{-1}X'FY \tag{2.8}$$

and

$$\tilde{l}_T(\boldsymbol{\mu} \mid \tilde{\boldsymbol{\beta}}_T, \mathbf{Y}) = -\frac{n}{T}\sum_{t=p+1}^{T} \log \sigma^2(t) + \frac{T-p}{T}\sum_{s=0}^{n-1} \log |\psi(e^{i2\pi s/n})|^2$$

$$-\frac{1}{T}S(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_T). \tag{2.9}$$

To find the maximum conditional likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$, we first obtain $\hat{\boldsymbol{\mu}}_T$ by maximizing $\tilde{l}_T(\boldsymbol{\mu}, \boldsymbol{\sigma} \mid \tilde{\boldsymbol{\beta}}_T, \mathbf{Y})$. Then the estimate of $\boldsymbol{\beta}$ is of the form

$$\hat{\boldsymbol{\beta}}_T = (X'\hat{F}X)^{-1}X'\hat{F}Y. \tag{2.10}$$

3. STRONG CONSISTENCY

In this section, we discuss the consistency properties of the parameter estimates. The process $\{\xi(t)\}$ in (1.4) is usually not stationary since the variances $\{\sigma^2(t)\}$ depend upon t. However, since the variances are periodic with $\sigma^2(t) = \sigma^2(t-12)$, instead of the monthly series $\{\xi(t)\}$, we may consider a yearly vector process which is stationary. Specifically, define $\eta(t) = \psi^{-1}(W_n)\,\varepsilon(t)$ and

$$\tilde{\eta}(\tilde{t}) = [\,\eta(12\tilde{t}+1)', \ldots, \eta(12\tilde{t}+12)'\,]',$$

$$\tilde{\xi}(t) = [\,\xi(12\tilde{t}+1)', \ldots, \xi(12\tilde{t}+12)'\,]'.$$

Then $\{\tilde{\eta}(\tilde{t})\}$ and $\{\tilde{\xi}(\tilde{t})\}$ are yearly vector series. The covariance matrix of $\tilde{\eta}(\tilde{t})$ is

$$K(\mu) = \Sigma \otimes [\,\psi(W_n)\,]^{-1}\,[\,\psi(W_n')\,]^{-1}, \tag{3.1}$$

where $\Sigma = \mathrm{Diag}(\sigma^2(1), \sigma^2(2), \ldots, \sigma^2(12))$. Furthermore, the model in (2.2) can be written in the form

$$\tilde{\Phi}(B, \mu)\,\tilde{\xi}(\tilde{t}) = \tilde{\eta}(\tilde{t}), \tag{3.2}$$

where $\tilde{\Phi}(z, \mu) = I_{12} \otimes \Phi(z)$ and note that B operates on $t = 12\tilde{t}+m$ for $m = 1, 2, \ldots, 12$, i.e., $B\xi(12\tilde{t}+m) = \xi(12\tilde{t}+m-1)$. If the roots of $|\Phi(z)| = 0$ lie outside the unit circle, the yearly series $\{\xi(\tilde{t})\}$ is strictly stationary and ergodic since we assume that $\{\varepsilon(t),\ t = 1, \ldots, T\}$ are independent and normally distributed. Define $\Phi_l = \phi_l[\,\psi(W_n)\,]^{-1}$ and

$$G = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ I_{(p-1)n} & & & O \end{bmatrix}.$$

Then the fact that the roots of $\Phi(z) = 0$ lie outside the unit circle is equivalent to the fact that the eigenvalues of G lie inside the unit circle.

Let A be an $n \times n$ matrix and $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Define $\lambda_*(A) = \min_{1 \leqslant k \leqslant n} |\lambda_k|$ and $\lambda^*(A) = \max_{1 \leqslant k \leqslant n} |\lambda_k|$. For the parameters μ defined in Section 1, we assume, throughout the rest of this paper, that $\mu \in \Theta$ such that

(E1) $\quad \lambda_*(\psi(W_n)) = \min_{0 \leqslant s \leqslant n-1} |\psi(e^{i2s\pi/n})| \geqslant \tau_0 > 0,$

(E2) $\quad \lambda^*(G) \leqslant 1 - \delta_0 < 1,$

(E3) $\quad 0 < c_1 \leqslant \sigma^2(s) \leqslant c_2 < \infty;\ s = 1, 2, \ldots, 12,$

(E4) $\quad \Theta$ is compact,

where the constants τ_0 and δ_0 in the conditions E1 and E2 can be chosen as small as we like. We refer to the above conditions as conditions E. Condition E1 simply ensures that the matrix $\psi(W_n)$ is invertible, and condition E2 guarantees that the yearly series $\{\tilde{\xi}(\bar{t})\}$ is strictly stationary and ergodic. Finally, condition E4 requires that the coefficients in model (1.2) are bounded. Conditions E may be weakened such that some non-stationary models could be included, but we do not intend to go further in that direction. From now on, we denote the true parameters in the space-time regression model as β_0 and $\mu_0 \in \bar{\Theta}$.

The spectral density of the yearly vector process $\{\tilde{\xi}(\bar{t})\}$ is of the following form:

$$f(\omega, \mu) = (2\pi)^{-1} \tilde{\Phi}(e^{i\omega}, \mu)^{-1} K(\mu) \tilde{\Phi}^*(e^{i\omega}, \mu)^{-1}. \qquad (3.3)$$

The eigenvalues of $K(\mu)$ are

$$v_{js} = \sigma^2(j)[\psi(e^{i2s\pi/n}) \psi(e^{-i2s\pi/n})]^{-1}, \qquad j = 1, 2, ..., 12; \quad s = 0, 1, ..., n-1.$$

By conditions E1 and E3, we have

$$\lambda^*(K(\mu)) \leqslant \frac{c_2}{\tau_0^2} < \infty. \qquad (3.4)$$

Since $\psi(e^{i2s\pi/n}) \psi(e^{i2s\pi/n})$ are continuous functions of α and θ, there exists a constant N such that

$$\sup_{\alpha, \theta \in \Theta} |\psi(e^{i2s\pi/n}) \psi(e^{-i2s\pi/n})| \leqslant N < \infty, \qquad s = 0, 1, ..., n-1.$$

Hence

$$\lambda_*(K(\mu)) \geqslant \frac{c_1}{N} > 0. \qquad (3.5)$$

Furthermore, the eigenvalues of the matrix $\tilde{\Phi}(e^{i\omega}, \mu)$ are

$$v_s(\omega, \mu) = 1 - \sum_{l=1}^{p} \phi_l e^{il\omega} \psi^{-1}(e^{i2s\pi/n}), \qquad s = 0, 1, ..., n-1.$$

Under condition E2, we have $|\tilde{\Phi}(z, \mu)| \neq 0$ for $|z| \leqslant 1/(1-\delta_0)$, which implies that $|v_s(\omega, \mu)| > 0$ for $(\omega, \mu) \in [-\pi, \pi] \times \bar{\Theta}$. Since $\{v_s(\omega, \mu)\}$ are continuous functions of $(\omega, \mu) \in [-\pi, \pi] \times \bar{\Theta}$, there exist constants $\tau_1 > 0$ and $\tau_2 > 0$ such that

$$0 < \tau_1 \leqslant \lambda_*(\tilde{\Phi}(e^{i\omega}, \mu) \tilde{\Phi}^*(e^{i\omega}, \mu))$$

$$\leqslant \lambda^*(\tilde{\Phi}(e^{i\omega}, \mu) \tilde{\Phi}^*(e^{i\omega}, \mu)) \leqslant \tau_2 < \infty. \qquad (3.6)$$

Hence we have

$$\lambda^*(f(\omega, \mu)) = \sup_{\mathbf{a} \neq 0} \frac{\mathbf{a}' f(\omega, \mu)\, \mathbf{a}}{\mathbf{a}'\mathbf{a}} = \sup_{\mathbf{a} \neq 0} \frac{\mathbf{a}' K(\mu)\, \mathbf{a}}{\mathbf{a}'(\tilde{\Phi}(e^{i\omega}, \mu)\, \tilde{\Phi}^*(e^{i\omega}, \mu))\, \mathbf{a}}$$

$$\leqslant \frac{\lambda^*(K(\mu))}{\lambda_*((\tilde{\Phi}(e^{i\omega}, \mu)\, \tilde{\Phi}^*(e^{i\omega}, \mu)))} \leqslant \frac{c_2}{\tau_1 \tau_0^2} < \infty \tag{3.7}$$

and

$$\lambda_*(f(\omega, \mu)) = \inf_{\mathbf{a} \neq 0} \frac{\mathbf{a}' f(\omega, \mu)\, \mathbf{a}}{\mathbf{a}'\mathbf{a}} = \inf_{\mathbf{a} \neq 0} \frac{\mathbf{a}' K(\mu)\, \mathbf{a}}{\mathbf{a}'(\tilde{\Phi}(e^{i\omega}, \mu)\, \tilde{\Phi}^*(e^{i\omega}, \mu))\, \mathbf{a}}$$

$$\geqslant \frac{\lambda_*(K(\mu))}{\lambda^*((\tilde{\Phi}(e^{i\omega}, \mu)\, \tilde{\Phi}^*(e^{i\omega}, \mu)))} \geqslant \frac{c_1}{N\tau_2} > 0. \tag{3.8}$$

Letting $\{\xi(t)\}$ be the vector process in (1.5) with true parameters $\mathbf{v}_0$ and $E\varepsilon(s)\, \varepsilon'(t) = \delta_{st} K(\mathbf{v})$, Dunsmuir and Hannan [4, Corollary 1] showed that the maximum likelihood estimates of the parameters are strongly consistent under some basic conditions. One of their conditions is $0 < \tau \leqslant \lambda_*(K(\mathbf{v}_0))$. But Deistler *et al.* [2] found that the their conditions were not enough. For the result of Corollary 1 of Dunsmuir and Hannan [4] to be valid, two extra conditions on the spectral density are required:

$$\lambda_*(K(\mathbf{v})) \geqslant \tau > 0 \qquad \text{for all} \quad \mathbf{v} \in \bar{\Theta}, \tag{3.9}$$

$$|\det(h(e^{i\omega}, \mathbf{v}))| \geqslant \delta > 0. \tag{3.10}$$

For the space–time model in (1.4), the condition in (3.10) is equivalent to condition E2. From (3.5), the matrix $K(\mu)$ satisfies the condition in (3.9) too. Since we assume that the vector AR(p) process $\{\tilde{\xi}(\tilde{t})\}$ is Gaussian, the maximum likelihood estimate of μ_0 is strongly consistent if the process $\{\xi(t)\}$ is observable.

To show that the maximum conditional likelihood estimates of μ_0 is strongly consistent, we first state the following general result. The proof of this result is essentially the same as that of Corollary 1 of Dunsmuir and Hannan [4] and theorefore omitted.

LEMMA 3.1. *Suppose that the vector process* $\xi(t)$ *in model* (1.4) *is observable and that the parameters* μ *satisfy conditions* E. *Let* $l_T^*(\mu \mid \xi, \xi_*)$ *be an approximation to* $l_T(\mu \mid \xi, \xi_*)$ *defined in* (2.4) *and let* μ_T^* *be the estimate of* μ_0 *obtained by maximizing* $l_T^*(\mu \mid \xi, \xi_*)$. *If*

$$\lim_{T \to \infty} (l_T^*(\mu \mid \xi, \xi_*) - l_T(\mu \mid \xi, \xi_*)) = 0 \qquad a.s.$$

uniformly in $\mu \in \bar{\Theta}$, *then*

$$\lim_{T \to \infty} \mu_T^* = \mu_0, \qquad a.s.$$

Based on Lemma 3.1, we have the following result for the maximum conditional likelihood estimate of μ_0.

THEOREM 3.1. *Suppose that the vector process* $\xi(t)$ *in model* (1.4) *is observable and* $\bar{\mu}_T$ *is the estimate of* μ_0 *obtained by maximizing* $\bar{l}_T(\mu \mid \xi, \xi_*)$ *defined in* (2.5), *then*

$$\lim_{T \to \infty} \bar{\mu}_T = \mu_0, \qquad a.s.$$

Proof. By Lemma 3.1, it suffices to prove that

$$\lim_{T \to \infty} (l_T(\mu \mid \xi, \xi_*) - \bar{l}_T(\mu \mid \xi, \xi_*)) = \frac{1}{T} \log |\Gamma| - \frac{1}{T} \xi_*' \Gamma^{-1} \xi_* = 0 \qquad a.s.$$

$$(3.11)$$

uniformly in $\mu \in \bar{\Theta}$, where the elements of Γ are functions of μ.

Let $\Gamma(t, t+k) = \mathrm{Cov}(\xi(t), \xi(t+k))$. Then by (2.3), we have

$$\Gamma(t, t+k) = \sum_{j=0}^{\infty} \sigma_{t-j}^2 \Psi_j [\psi^{-1}(W_n) \psi^{-1}(W_n')] \Psi_{j+k}'$$

and

$$\Gamma = \begin{bmatrix} \Gamma(1, 1) & \cdots & \Gamma(1, p) \\ \cdots & \cdots \\ \Gamma(p, 1) & \cdots & \Gamma(p, p) \end{bmatrix}.$$

By condition E3, $\{\varepsilon(t)\}$ is a periodic process with $0 < c_1 \leqslant \sigma^2(t) \leqslant c_2 < \infty$. Let $\{\bar{\varepsilon}(t)\}$ be independent random vectors with mean zero and

$$\mathrm{Cov}(\bar{\varepsilon}(t)) = (\sigma^2(t) - c_1/2) I_n,$$

and suppose that $\{\bar{\xi}(t)\}$ is a vector process which satisfies the equation

$$\Phi(B) \bar{\xi}(t) = \psi^{-1}(W_n) \bar{\varepsilon}(t).$$

Then the covariance matrices of $\{\bar{\xi}(t)\}$ are

$$\bar{\Gamma}(t, t+k) = \sum_{j=0}^{\infty} (\sigma_{t-j}^2 - c_1/2) \Psi_j [\psi^{-1}(W_n) \psi^{-1}(W_n')] \Psi_{j+k}'$$

$$= \Gamma(t, t+k) - \tilde{\Gamma}(k),$$

where

$$\tilde{\Gamma}(k) = \frac{c_1}{2} \sum_{j=0}^{\infty} \Psi_j [\psi^{-1}(W_n)\, \psi^{-1}(W_n')] \, \Psi_{j+k}'.$$

It is clearly that $\tilde{\Gamma}(k)$ is the covariance matrix function of a certain stationary vector process. Define

$$\tilde{f}(\omega, \mu) = (2\pi)^{-1} \Phi(e^{i\omega})^{-1} [\psi^{-1}(W_n)\, \psi^{-1}(W_n')] \, \Phi^*(e^{i\omega})^{-1}.$$

Then $\tilde{f}(\omega, \mu)$ is a spectral density matrix function and

$$\tilde{\Gamma}(k) = \frac{c_1}{2} \int_{-\pi}^{\pi} e^{ik\omega} \tilde{f}(\omega, \mu)\, d\omega.$$

Let

$$\Gamma = \begin{bmatrix} \bar{\Gamma}(1,1) & \cdots & \bar{\Gamma}(1,p) \\ \cdots & \cdots & \\ \bar{\Gamma}(p,1) & \cdots & \bar{\Gamma}(p,p) \end{bmatrix}, \qquad \tilde{\Gamma} = \begin{bmatrix} \tilde{\Gamma}(0) & \cdots & \tilde{\Gamma}(p-1) \\ \cdots & \cdots & \\ \tilde{\Gamma}'(p-1) & \cdots & \tilde{\Gamma}(0) \end{bmatrix}.$$

Under condition E3, there exists a constant τ_3 independent of μ such that $0 < \tau_3 \leqslant \lambda_*(f(\omega, \mu))$, which implies that $c_1 \pi \tau_3 \leqslant \lambda_*(\tilde{\Gamma})$. Since $\bar{\Gamma} = \Gamma - \tilde{\Gamma}$ is a non-negative definite matrix, we have

$$\lambda_*(\Gamma) \geqslant \lambda_*(\tilde{\Gamma}) \geqslant c_1 \pi \tau_3 > 0.$$

Similarly, we can show that there exists a constant τ_4 independent of μ such that

$$\lambda^*(\Gamma) \leqslant c_2 \pi \tau_4 < \infty.$$

Now we have

$$0 < (c_1 \pi \tau_3)^{12np} \leqslant |\Gamma| \leqslant (c_2 \pi \tau_4)^{12np} < \infty$$

and

$$0 \leqslant \xi_*' \Gamma^{-1} \xi_* \leqslant \lambda^*(\Gamma^{-1}) \, \xi_*' \xi_* \leqslant \frac{1}{c_1 \pi \tau_3} \sum_{t=1}^{p} \xi(t)' \, \xi(t).$$

Hence (3.11) holds. ∎

We now discuss the strong consistency of $\hat{\boldsymbol{\beta}}_T$ and $\hat{\boldsymbol{\mu}}$. For this purpose, we need to impose some conditions on the regressors. Define

$$d_T(j) = \sum_{t=1}^{T} x_j^2(t).$$

The conditions imposed on the regressors are

(C1)   There exists a constant $0 < \rho \leqslant 1/2$ such that $\lim \sup_{T \to \infty} T^{\rho}/d_T(j)$
$\leqslant N_0 < \infty$, for $j = 1, 2, ..., m$;

(C2) $\lim_{T \to \infty} x_j^2(T)/d_T(j) = 0$, for $j = 1, 2, ..., m$;

(C3)   For all integers h and $j, k = 1, 2, ..., m$,

$$\lim_{T \to \infty} \frac{\sum_{t=1}^{T} x_j(t)\, x_k(t+h)}{\sqrt{d_T(j)\, d_T(k)}} = r_{jk}(h).$$

(C4)   Let $R(h)$ be the matrix with entries $r_{jk}(h)$. Then $R(0)$ is positive definite.

We will refer to the above conditions as conditions C.

The conditions C are essentially the same as the well-known Grenander's conditions, except condition C1 is replaced by $\lim_{T \to \infty} d_T(j) = \infty$ in Grenander's setting. For polynomial and trigonometric regressions, it is easy to show that condition C1 is satisfied. Grenander and Rosenblatt [5, Chap. 7] proved that for these regression variables, conditions C2–C4 are satisfied too. Furthermore, if $\mathbf{a}$ is a vector of complex numbers, $\{\mathbf{a}^* R(h)\, \mathbf{a}\}$ forms a non-negative definite sequence. Therefore $R(h)$ can be expressed as

$$R(h) = \int_{-\pi}^{\pi} e^{ih\omega} M(d\omega), \tag{3.12}$$

where $M(\omega)$ is a function of ω taking as values $m \times m$ matrices and $\Delta M(\omega) = M(\omega_1) - M(\omega_2)$ is a non-negative definite matrix for every interval (ω_1, ω_2). $M(\omega)$ is called the spectral distribution function of the regression variables.

THEOREM 3.2.   *Suppose that the regression variables satisfy conditions* C. *Then for* $\tilde{\boldsymbol{\beta}}_T$ *defined in* (2.8), *we have*

$$\lim_{T \to \infty} \tilde{\boldsymbol{\beta}}_T = \boldsymbol{\beta}_0 \qquad a.s. \tag{3.13}$$

uniformly in $\boldsymbol{\mu} \in \bar{\boldsymbol{\Theta}}$.

Proof. Define

$$\mathbf{d}_T = \mathrm{Diag}\left(\frac{1}{\sqrt{d_T(1)}}, \ldots, \frac{1}{\sqrt{d_T(m)}}\right), \qquad D_T = I_n \otimes \mathbf{d}_T,$$

$$V_T = D_T X' FX D_T, \qquad \zeta_T = T^{-p/2} D_T X' F\xi.$$

Then we have

$$\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0 = (X'FX)^{-1} X'F\xi = (T^{p/2}D_T) V_T^{-1}\zeta_T$$

By condition C1, $\{\lambda^*(T^pD_T^2)\}$ are bounded and independent of $\boldsymbol{\mu}$. It can be shown that for a certain large number $T_0 > 0$, $\{\lambda^*(V_T^{-1}), \ T \geq T_0\}$ are bounded uniformly in $\boldsymbol{\mu} \in \bar{\boldsymbol{\Theta}}$ and that $\lim_{T \to \infty} \zeta_T = 0$, a.s. uniformly in $\boldsymbol{\mu} \in \bar{\boldsymbol{\Theta}}$ (see Niu [11]). Hence (3.13) follows. ∎

For the consistency properties of $\hat{\boldsymbol{\beta}}_T$ and $\hat{\boldsymbol{\mu}}_T$, we have the following result.

THEOREM 3.3. *Suppose that the regression variables satisfy conditions C. Then we have*

$$\lim_{T \to \infty} \hat{\boldsymbol{\mu}}_T = \boldsymbol{\mu}_0 \qquad a.s. \tag{3.14}$$

and

$$\lim_{T \to \infty} \hat{\boldsymbol{\beta}}_T = \boldsymbol{\beta}_0 \qquad a.s. \tag{3.15}$$

Proof. Note that

$$(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta})' X'FX(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) = (\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta})' X'F(Y - X\boldsymbol{\beta}).$$

Hence

$$S(Y - X\tilde{\boldsymbol{\beta}}_T) = (Y - X\tilde{\boldsymbol{\beta}}_T)' F(Y - X\tilde{\boldsymbol{\beta}}_T)$$

$$= (Y - X\boldsymbol{\beta})' F(Y - X\boldsymbol{\beta}) - (\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta})' X'FX(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta})$$

$$= S(\xi, \xi_*) - (\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta})' X'FX(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}). \tag{3.16}$$

By Lemma 3.1, in order to show $\lim_{T \to \infty} \hat{\boldsymbol{\mu}}_T \overset{a.s.}{=} \boldsymbol{\mu}_0$, it suffices to prove that

$$\lim_{T \to \infty} (\tilde{l}_T(\boldsymbol{\mu} \mid \tilde{\boldsymbol{\beta}}_T, Y) - \tilde{l}_T(\boldsymbol{\mu} \mid \xi, \xi_*)) = \frac{1}{T}(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0)' X'FX(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) = 0 \qquad a.s.$$

uniformly in $\mu \in \bar{\Theta}$. This follows from $\tilde{\beta}_T \xrightarrow{\text{a.s.}} \beta_0$ uniformly in $\mu \in \bar{\Theta}$ and $\lambda^*(D_T X' F X D_T)$ uniformly bounded in $\mu \in \bar{\Theta}$.

Since $\tilde{\beta}_T$ is a continuous function of μ, $\tilde{\beta}_T \xrightarrow{\text{a.s.}} \beta_0$ uniformly in $\mu \in \bar{\Theta}$ and $\hat{\mu}_T \xrightarrow{\text{a.s.}} \mu_0$, by (2.8), (2.10), and the continuous mapping theorem, we have

$$\lim_{T \to \infty} \hat{\beta}_T = \beta_0 \qquad \text{a.s.} \quad \blacksquare$$

COROLLARY 3.1. *Suppose that the regression variables satisfy conditions* C. *Let* β_T^* *and* μ_T^* *be the estimates obtained by maximizing the log-likelihood function defined in* (2.6). *Then we have*

$$\lim_{T \to \infty} \mu_T^* = \mu_0, \qquad a.s. \tag{3.17}$$

and

$$\lim_{T \to \infty} \beta_T^* = \beta_0, \qquad a.s. \tag{3.18}$$

Proof. It can be shown that $\beta_T^* - \hat{\beta}_T$ converges to zero almost surely. Therefore (3.18) follows from (3.15). The proof of (3.17) is essentially the same as that of (3.14), hence it is omitted. $\blacksquare$

4. THE CENTRAL LIMIT THEOREM

When $\xi(t)$ is a general linear process with the spectral density defined in (1.6), where the matrix functions $h^{-1}g$ and K are specified by different sets of parameters, Dunsmuir and Hannan [4] established the central limit theorem for the maximum likelihood type estimates of the parameters in model (1.5). In a followup paper, Dunsmuir [3] established the central limit theorem for parameter estimates obtained by maximizing two specific frequency domain approximations to the Gaussian likelihood, in which the innovation covariance matrix and the linear transfer function need not be separately parametrized. In this section, we study the central limit theorem for the estimates of the parameters in the space-time regression models. The maximum likelihood estimates and the maximum conditional likelihood estimates have exactly the same limit distributions. Since the proof of the central limit theorems for the two types of estimator are essentially the same, we only present the results for the maximum conditional likelihood estimates. First, we state the following two lemmas which proofs were given in Niu [11].

LEMMA 4.1. *Suppose that the regression variables satisfy conditions* C. *Then we have*

$$\lim_{T \to \infty} \frac{1}{\sqrt{T}} \frac{\partial}{\partial \mu_k} (\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \, X' FX(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) = 0 \qquad a.s. \qquad (4.1)$$

uniformly in $\boldsymbol{\mu} \in \bar{\boldsymbol{\Theta}}$ *and*

$$\lim_{T \to \infty} \frac{1}{\sqrt{T}} \frac{\partial}{\partial \mu_k \partial \mu_l} (\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \, X' FX(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) = 0 \qquad a.s. \qquad (4.2)$$

uniformly in $\boldsymbol{\mu} \in \bar{\boldsymbol{\Theta}}$.

LEMMA 4.2. *Suppose that the yearly vector process* $\{\bar{\xi}(t)\}$ *has the spectral density matrix* $f(\omega, \boldsymbol{\mu}_0)$ *defined in* (3.3). *Then we have*

$$\lim_{T \to \infty} \frac{1}{T} S(\xi, \xi_*) \stackrel{a.s.}{=} \frac{1}{12} \times \text{trace} \left(\int_{-\pi}^{\pi} \frac{1}{2\pi} f(\omega, \boldsymbol{\mu})^{-1} f(\omega, \boldsymbol{\mu}_0) \, d\omega \right), \qquad (4.3)$$

$$\lim_{T \to \infty} \frac{1}{T} \frac{\partial}{\partial \mu_k} S(\xi, \xi_*) \stackrel{a.s.}{=} \frac{1}{12} \times \text{trace} \left(\int_{-\pi}^{\pi} \frac{1}{2\pi} \frac{\partial f(\omega, \boldsymbol{\mu})^{-1}}{\partial \mu_k} f(\omega, \boldsymbol{\mu}_0) \, d\omega \right), \qquad (4.4)$$

and

$$\lim_{T \to \infty} \frac{1}{T} \frac{\partial^2}{\partial \mu_k \partial \mu_l} S(\xi, \xi_*) \stackrel{a.s.}{=} \frac{1}{12} \times \text{trace} \left(\int_{-\pi}^{\pi} \frac{1}{2\pi} \frac{\partial^2 f(\omega, \boldsymbol{\mu})^{-1}}{\partial \mu_k \partial \mu_l} f(\omega, \boldsymbol{\mu}_0) \, d\omega \right). \qquad (4.5)$$

Furthermore, the convergences in (4.3), (4.4), *and* (4.5) *are uniformly in* $\boldsymbol{\mu} \in \bar{\boldsymbol{\Theta}}$.

From now on, we suppose that $T/12$ is an integer. Define

$$\mathbf{k}_t = \sum_{s=1}^{n} \log |\psi(e^{i2s\pi/n})|^2 - n \log \sigma^2(t),$$

$$S_t = \frac{1}{\sigma^2(t)} \left(\psi(W_n) \xi(t) - \sum_{l=1}^{p} \phi_l \xi(t-l) \right)' \left(\psi(W_n) \xi(t) - \sum_{l=1}^{p} \phi_l \xi(t-l) \right),$$

and

$$\mathbf{k}(\boldsymbol{\mu}) = (T-p) \sum_{s=1}^{n} \log |\psi(e^{i2s\pi/n})|^2 - n \sum_{t=p+1}^{T} \log \sigma^2(t).$$

Then

$$S(\xi, \xi_*) = \sum_{t=p+1}^{T} S_t, \qquad \mathbf{k}(\mathbf{\mu}) = \sum_{t=p+1}^{T} \mathbf{k}_t,$$

and

$$\lim_{T \to \infty} \frac{1}{T} \mathbf{k}(\mathbf{\mu}) = \frac{1}{12} \log |K(\mathbf{\mu})| = \sum_{s=1}^{n} \log |\psi(\varepsilon^{i2s\pi/n})|^2 - n \sum_{s=1}^{12} \log \sigma^2(s).$$

Furthermore, define

$$u_t(k) = \frac{\partial}{\partial \alpha_k} (\mathbf{k}_t + S_t)|_{\mathbf{\mu}_0} = \frac{2}{\sigma_0^2(t)} \xi'(t)(W_n')^k \varepsilon(t)$$

$$+ \sum_{s=1}^{n} \frac{\partial}{\partial \alpha_k} (\log |\psi(\varepsilon^{i2s\pi/n})|^2)|_{\mathbf{\mu}_0},$$

$$u_t(q+k) = \frac{\partial}{\partial \theta_k} (\mathbf{k}_t + S_t)|_{\mathbf{\mu}_0} = \frac{2}{\sigma_0^2(t)} \xi'(t) W_n^k \varepsilon(t)$$

$$+ \sum_{s=1}^{n} \frac{\partial}{\partial \theta_k} (\log |\psi(e^{i2s\pi/n})|^2)|_{\mathbf{\mu}_0},$$

$$u_t(2q+k) = \frac{\partial}{\partial \phi_k} (\mathbf{k}_t + S_t)|_{\mathbf{\mu}_0} = \frac{2}{\sigma_0^2(t)} \xi'(t-k) \varepsilon(t),$$

$$u_t(2q+p+k) = \frac{\partial}{\partial \sigma_k^2} (\mathbf{k}_t + S_t)|_{\mathbf{\mu}_0} = \left\{ \frac{1}{\sigma_0^4(t)} \varepsilon'(t) \varepsilon(t) \right\}_{t \mid \mathrm{mod}\ 12 = k} - \frac{n}{\sigma_0^2(k)}$$

for $\quad 1 \leq k \leq 12,$

and

$$Z_T(k) = \frac{\partial}{\partial \mu_k} \left(\frac{1}{T} \mathbf{k}(\mathbf{\mu}) - \frac{1}{T} S(\xi, \xi_*) \right) \Bigg|_{\mathbf{\mu}_0}.$$

It is easy to see that $Z_T(k) = (1/T) \sum_{t=p+1}^{T} u_t(k)$.

Let $\mathscr{F}_t$ be the σ-field generated by $\{\varepsilon(s), s \leq t\}$. It can be shown that $Eu_k(t) = 0$ and $\{\mu_k(t)\}$ is a sequence of martingale differences with respect to $\{\mathscr{F}_t\}$ and satisfies the conditional Lindeberg condition, i.e.,

$$\frac{1}{T} \sum_{t=p+1}^{T} E\{u_k(t)^2 I_{(|u_k(t)| > \sqrt{T}\varepsilon)} \mid \mathscr{F}_{t-1}\} \xrightarrow{\text{a.s.}} 0 \qquad \text{for all} \quad \varepsilon > 0. \quad (4.6)$$

Furthermore, we have

$$E\{u_k(t)\,u_l(s) \mid \mathscr{F}_{t-1}\} = 0 \qquad \text{for} \quad s \neq t,$$

$$E\{u_k(t)\,u_l(t) \mid \mathscr{F}_{t-1}\} = 0 \qquad \text{for} \quad k \neq l; \quad k, l \geqslant 2q+1,$$

and for each pair (k, l), the limit of $(1/T) \sum_{t=p+1}^{T} E\{u_k(t)\,u_l(t) \mid \mathscr{F}_{t-1}\}$ exists as $T \to \infty$. The exact expression of the limits was given in Niu [11]. Define

$$V_0 = \left(\lim_{T \to \infty} \frac{1}{T} \sum_{t=p+1}^{T} E\{u_k(t)\,u_l(t) \mid \mathscr{F}_{t-1}\} \right),$$

$$\mathbf{Z}_T = (Z_T(1), Z_T(2), ..., Z_T(2q+p+12))'.$$

By the martingale central limit theorem, we have the following result.

LEMMA 4.3. *Suppose that the vector process* $\{\xi(t)\}$ *is generated by model* (1.4) *with the true parameter* $\mu_0 \in \Theta$. *Then* $\sqrt{T}\,\mathbf{Z}_T$ *is asymptotically normal with zero mean and covariance matrix* V_0. *Furthermore, the random variables* $\{\sqrt{T}\,Z_T(k); \; k \geqslant 2q+1\}$ *are asymptotically independent.*

Based on Lemmas 4.1, 4.2, and 4.3, we now obtain the central limit theorem for the maximum conditional likelihood estimates $\hat{\mu}$. Let

$$v_{kl} \overset{\text{def}}{=} \frac{1}{12} \text{trace} \left(\int_{-\pi}^{\pi} \frac{1}{2\pi} f(\omega, \mu)^{-1} \frac{\partial f(\omega, \mu)}{\partial \mu_k} f(\omega, \mu)^{-1} \frac{\partial f(\omega, \mu)}{\partial \mu_l} \, d\omega \Big|_{\mu_0} \right) \tag{4.7}$$

and V be the matrix with v_{kl} as its typical elements. Then we have the following result.

THEOREM 4.1. *Suppose that the vector process* $\{\xi(t)\}$ *is generated by model* (1.4) *with true parameter* $\mu_0 \in \Theta$ *and the regression variables satisfy conditions C. Then* $\sqrt{T}\,(\hat{\mu} - \mu_0)$ *is asymptotically normal with zero mean and covariance matrix* $V^{-1}V_0 V^{-1}$.

Proof. Since $\hat{\mu}$ maximizes the conditional likelihood function $\tilde{l}_T(\mu \mid \tilde{\beta}_T, Y)$ defined in (2.9), we have

$$0 = \sqrt{T} \left(\frac{\partial}{\partial \mu_l} \tilde{l}_T(\mu \mid \tilde{\beta}_T, Y) \Big|_{\hat{\mu}} \right) = \sqrt{T} \left(\frac{\partial}{\partial \mu_l} \tilde{l}_T(\mu \mid \tilde{\beta}_T, Y) \Big|_{\mu_0} \right)$$

$$+ \left(\frac{\partial^2}{\partial \mu_k \partial \mu_l} \tilde{l}_T(\mu \mid \tilde{\beta}_T, Y) \Big|_{\hat{\mu}} \right) \times \sqrt{T}\,(\hat{\mu} - \mu_0), \tag{4.8}$$

where $\|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}_0\| \leqslant \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0\|$. Noting that by (2.9) and (3.16), we have

$$\sqrt{T}\left(\frac{\partial}{\partial \mu_l} \tilde{l}_T(\boldsymbol{\mu} \mid \tilde{\boldsymbol{\beta}}_T, \mathbf{Y})\Big|_{\boldsymbol{\mu}_0}\right)$$

$$= \sqrt{T}\,\mathbf{Z}_T + \frac{1}{\sqrt{T}}\left(\frac{\partial}{\partial \mu_l}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0)\,X'FX(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0)\Big|_{\boldsymbol{\mu}_0}\right) \qquad (4.9)$$

and

$$\frac{\partial^2}{\partial \mu_k \partial \mu_l} \tilde{l}_T(\boldsymbol{\mu} \mid \tilde{\boldsymbol{\beta}}_T, \mathbf{Y}) = \left(\frac{1}{T}\frac{\partial^2}{\partial \mu_k \partial \mu_l}\mathbf{k}(\boldsymbol{\mu}) - \frac{1}{T}\frac{\partial^2}{\partial \mu_k \partial \mu_l}S(\xi, \xi_*)\right)$$

$$+ \left(\frac{1}{T}\frac{\partial}{\partial \mu_k \partial \mu_l}(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0)\,X'FX(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0)\right) \quad (4.10)$$

By Lemma 4.1, the second terms in the right sides of (4.9) and (4.10) converge almost surely to zero uniformly in $\boldsymbol{\mu} \in \bar{\Theta}$. By Lemma 4.2 and the proof of Theorem 2.1 in Dunsmuir [3], it can be shown that

$$\lim_{T \to \infty}\left(-\frac{1}{T}\frac{\partial^2}{\partial \mu_k \partial \mu_l}\mathbf{k}(\boldsymbol{\mu}) + \frac{1}{T}\frac{\partial^2}{\partial \mu_k \partial \mu_l}S(\xi, \xi_*)\right)\Big|_{\hat{\boldsymbol{\mu}}} \overset{\text{a.s.}}{=} V.$$

Hence we can write (4.8) as

$$\sqrt{T}\,\mathbf{Z}_T + o(1) = (V + o(1)) \times \sqrt{T}\,(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0),$$

where "$o(1)$" denotes a term which converges to zero almost surely. The result of this theorem then follows from Lemmas 4.3. ∎

To obtain the central limit theorem for $\hat{\boldsymbol{\beta}}_T$, we need modify conditions C. Define

$$\tilde{d}_T(j) = \sum_{t=1}^{T}\frac{x_j^2(t)}{\sigma^2(t)},$$

$$\tilde{\mathbf{d}}_T = \text{Diag}\left(\frac{1}{\sqrt{\tilde{d}_T(1)}}, ..., \frac{1}{\sqrt{\tilde{d}_T(m)}}\right),$$

$$\tilde{D}_T(\boldsymbol{\sigma}) = I_n \times \tilde{\mathbf{d}}_T^{-1}.$$

We modify conditions C as follows:

(C1′)   There exists a constant $0 < \rho \leqslant \frac{1}{2}$ such that $\limsup_{T \to \infty} T^\rho/\tilde{d}_T(j)$
$\leqslant N_0 < \infty$, for $j = 1, 2, ..., m$;

(C2′) $\lim_{T \to \infty} x_j^2(T)/\tilde{d}_T(j) = 0$, for $j = 1, 2, ..., m$;

(C3′) For all integers h and j, $k = 1, 2, ..., m$,

$$\lim_{T \to \infty} \frac{\sum_{t=1}^{T} x_j(t)\, x_k(t+h)/\sigma^2(t)}{\sqrt{\tilde{d}_T(j)\, \tilde{d}_T(k)}} = r_{jk}(h, \sigma)$$

uniformly for $\{\sigma^2(t)\}$ such that $0 < c_1 \leqslant \sigma^2(t) \leqslant c_2 < \infty$.

(C4′) $R(0, \sigma) = \{r_{jk}(0, \sigma)\}$ is positive definite.

We will refer to the above conditions as conditions C′.

Remark 1. It is easy to see that conditions C1′ and C2′ are equivalent to conditions C1 and C2. When the regression variables are non-negative or when $\sigma_0^2(t) \equiv \sigma_0^2$, condition C3′ is equivalent to condition C3.

Remark 2. Condition C4′ is equivalent to condition C4. In fact, it is obvious that condition C4′ implies condition C4. On other hand, let

$$R_T(0) = D_T X' X D_T, \qquad R_T(0, \sigma) = \tilde{D}_T(\sigma)\, X'(\Sigma_T^{-1} \otimes I_n)\, X\tilde{D}_T(\sigma),$$

where $\Sigma_T = \mathrm{Diag}(\sigma^2(1), ..., \sigma^2(T))$. Then

$$R_T(0, \sigma) \geqslant \frac{1}{c_1}\, (\tilde{D}_T(\sigma)\, D_T^{-1})\, R_T(0)(\tilde{D}_T(\sigma)\, D_T^{-1}).$$

Furthermore, noting that $\lambda^*(\tilde{D}_T^2(\sigma)\, D_T^{-2}) \leqslant c_1$, we have

$$\lambda_*(R_T(0, \sigma)) \geqslant \frac{1}{c_1} \inf_{a \neq 0} \frac{a' R_T(\sigma)\, a}{a'(\tilde{D}_T^2(\sigma)\, D_T^{-2})\, a} \geqslant \frac{1}{c_1^2} \lambda_*(R_T(0)).$$

Hence

$$\lambda_*(R(0, \sigma)) = \lim_{T \to \infty} \lambda_*(R_T(0, \sigma)) \geqslant \frac{1}{c_1^2} \lim_{T \to \infty} \lambda_*(R_T(0)) = \frac{1}{c_1^2} \lambda_*(R(0)).$$

Therefore if $R(0)$ is positive definite, so is $R(0, \sigma)$.

Remark 3. Let $R(h, \sigma)$ be the matrix with entries $r_{jk}(h, \sigma)$. Then $R(h, \sigma)$ can be expressed as

$$R(h, \sigma) = \int_{-\pi}^{\pi} e^{ih\omega} M(d\omega, \sigma).$$

THEOREM 4.2. *Suppose that the vector process* $\{\xi(t)\}$ *is generated by the model* (1.4) *with the true parameter* $\mu_0 \in \Theta$ *and the regression variables satisfy conditions* C′. *Then* $\tilde{D}_T(\hat{\sigma})(\hat{\beta}_T - \beta_0)$ *is asymptotically normal with zero mean and covariance matrix* $Q^{-1}(\mu_0)$, *where*

$$Q(\mu) = \int_{-\pi}^{\pi} 2\pi f_1(\omega, \mu) \otimes M(d\omega, \sigma),$$

$$f_1(\omega, \mu) = (2\pi)^{-1} h(e^{i\omega}, \mu) h^*(e^{i\omega}, \mu),$$

and

$$h(z, \mu) = \psi(W_n) - \sum_{l=1}^{p} \phi_l z^l I_n.$$

Furthermore, $\tilde{D}_T(\hat{\sigma})(\hat{\beta}_T - \beta_0)$ *and* $\sqrt{T}(\hat{\mu} - \mu_0)$ *are asymptotically independent.*

Proof. Similar to Theorem 3.2, define

$$V_T^{-1}(\mu) = (\tilde{D}_T(\sigma) X'FX\tilde{D}_T(\sigma))^{-1}, \qquad \zeta_T(\mu) = \tilde{D}_T(\mu) = \tilde{D}_T(\sigma) X'F\xi.$$

Then we have

$$\tilde{D}_T(\hat{\sigma}_T)(\hat{\beta}_T - \beta_0) = V_T^{-1}(\hat{\mu}_T) \zeta_T(\hat{\mu}_T).$$

By condition C3', we can show that $V_T^{-1} \xrightarrow{\text{a.s.}} Q^{-1}(\mu)$ uniformly in $\mu \in \bar{\Theta}$. Hence $V_T^{-1}(\hat{\mu}_T) \xrightarrow{\text{a.s.}} Q^{-1}(\mu_0)$.

Now it suffices to prove that $\zeta_T(\hat{\mu}_T)$ is asymptotically normally distributed with covariance matrix $Q(\mu_0)$. Note that

$$(\zeta_T(\hat{\mu}_T) - \zeta_T(\mu_0)) = \frac{1}{\sqrt{T}} \left(\frac{\partial}{\partial \mu} \zeta_T(\tilde{\mu}) \right) \sqrt{T}(\hat{\mu}_T - \mu_0),$$

where $\|\tilde{\mu} - \mu_0\| \leqslant \|\hat{\mu}_T - \mu_0\|$. Similar to the proof of Theorem 3.2, it can be shown that

$$\frac{1}{\sqrt{T}} \left(\frac{\partial}{\partial \mu} \zeta_T(\mu) \right) \xrightarrow{\text{a.s.}} 0$$

uniformly in $\mu \in \bar{\Theta}$. Therefore,

$$\frac{1}{\sqrt{T}} \left(\frac{\partial}{\partial \mu} \zeta_T(\tilde{\mu}) \right) \xrightarrow{\text{a.s.}} 0.$$

It then follows from Theorem 4.1 that $\zeta_T(\hat{\mu}_T) - \zeta_T(\mu_0) \xrightarrow{P} 0$. Hence, asymptotically $\zeta_T(\hat{\mu}_T)$ and $\zeta_T(\mu_0)$ have the same distribution.

Since the random vector ξ is normally distributed, so is $\zeta_T(\mu_0)$. Following the argument of Grenander and Rosenblatt [5, Chap. 7], we can show that the covariance matrix of $\zeta_T(\mu_0)$ converges to $Q(\mu_0)$. Therefore the limit distribution of $\zeta_T(\hat{\mu}_T)$ is normal with mean zero and covariance matrix $Q(\mu_0)$.

Furthermore, the asymptotic independence of $\tilde{D}_T(\hat{\sigma}_T)(\hat{\beta}_T - \beta_0)$ and $\sqrt{T}(\hat{\mu}_T - \mu_0)$ is equivalent to the asymptotic independence of $\zeta_T(\mu_0)$ and $\sqrt{T}\,\mathbf{Z}_T$. Note that

$$\zeta_T(\mu_0) = D_T(\sigma_0) \sum_{t=p+1}^{T} \frac{1}{\sigma_0^2(t)} \left(X'(t)\,\psi_0(W') - \sum_{l=1}^{p} \phi_0(l)\,X'(t-l) \right) \varepsilon(t)$$

$$\stackrel{\text{def}}{=} D_T(\sigma_0) \sum_{t=p+1}^{T} \zeta(t),$$

where $\zeta(t)$ is a vector with nm elements. Then the asymptotic independence of $\zeta_T(\mu_0)$ and $\sqrt{T}\,\mathbf{Z}_T$ follows from the fact that $\{\zeta(t)\}$ is a sequence of martingale differences with respect to $\{\mathscr{F}_t\}$ and that

$$\lim_{T \to \infty} \frac{1}{\sqrt{T}\,\tilde{d}_T(j)} \sum_{t=p+1}^{T} E\{\zeta(t)\,u_k(t) \mid \mathscr{F}_{t-1}\} = 0 \qquad \text{a.s.}$$

which is easy to establish by using condition C1′. ∎

References

1. BROCKWELL, P. J., AND DAVIS, R. A. (1991). *Time Series Theory and Methods*, 2nd ed. Springer-Verlag, New York.
2. DEISTLER, M., DUNSMUIR, W., AND HANNAN, E. J. (1978). Vector linear time series models: Corrections and extensions. *Adv. Appl. Probab.* 10 360–372.
3. DUNSMUIR, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Ann. Statist.* 7 490–506.
4. DUNSMUIR, W., AND HANNAN, E. J. (1976). Vector linear time series models. *Adv. Appl. Probab.* 8 339–364.
5. GRENANDER, U., AND ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.
6. HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.
7. HANNAN, E. J. (1971). Non-linear time series regression. *J. Appl. Probab.* 8 767–780.
8. HANNAN, E. J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.* 10 130–145.
9. HANNAN, E. J., DUNSMUIR, W., AND DEISTLER, M. (1980). Estimation of vector ARMAX models. *J. Multivariate Anal.* 10 275–295.
10. KOREZLIOGLU, H., AND LOUBATON, P. (1986). Spectral factorization of wide sense stationary processes on $\mathbf{Z}^2$. *J. Multivariate Anal.* 19 24–27.
11. NIU, X-F. (1993). *Asymptotic Properties of Maximum Likelihood Estimates in a Class of Space–Time Models*. Technical Report No. 874, Department of Statistics, Florida State University.
12. NIU, X-F., AND TIAO, G. C. (1993). Modelling satellite ozone data. *J. Amer. Statist. Assoc.*, in press.
13. PÖTSCHER, B. M. (1987). Convergence results for maximum likelihood type estimators in multivariable ARMA Models. *J. Multivariate Anal.* 21 29–52.

14. REINSEL, G. C., TIAO, G. C., WANG, M. N., LEWIS, R., AND NYCHKA, D. (1981). Statistical analysis of stratospheric ozone data for the detection of trends. *Atmos. Environ.* **15** 1569–1577.

15. REINSEL, G. C., TIAO, G. C., AHN, S. K., MARIAN, P., BASU, S., DELUISI, J. J., MATEER, C. L., MILLER, A. J., CONNELL, P. S., AND WUEBBLES, D. J. (1988). An analysis of the 7-Year record of SBUV satellite ozone data: Global profile features and trends in total ozone. *J. Geophys. Res.* **93** 1689–1703.

16. TJØSTHEIM, D. (1978). Statistical spatial series modeling. *Adv. Appl. Probab.* **10** 130–154.

17. TJØSTHEIM, D. (1981). Autoregressive modeling and spectral analysis of array data in the plane. *IEEE Trans. Geosci. Remote Sensing* **GE19** 15–24.

18. TJØSTHEIM, D. (1983). Statistical spatial series modeling II: Some further results on unilateral processes. *Adv. Appl. Probab.* **15** 562–584.

19. WALKER, A. M. (1964). Asymptotic properties of least squares estimates of the parameters of the stationary non-deterministic time series. *J. Austral. Math. Soc.* **4** 363–384.

20. WHITTLE, P. (1954). On stationary process in the plane. *Biometrika* **41** 434–449.

21. WHITTLE, P. (1963). On the fitting of multivariate autregressions and the approximate canonical factorization of a spectral density matrix. *Biometrika* **50** 129–134.