Pseudo best estimator by a separable approximation of spatial covariance structures

Toshihiro Hirano*

June 17, 2014

Abstract

We consider a linear regression model with a spatially correlated error term on a lattice. When estimating coefficients in the linear regression model, the generalized least squares estimator (GLSE) is used if the covariance structures are known. However, the GLSE for large spatial data sets is computationally expensive, because it involves inverting the covariance matrix of error terms from each observations. To reduce the computational complexity, we propose a pseudo best estimator (PBE) using spatial covariance structures approximated by separable covariance functions. We derive the asymptotic covariance matrix of the PBE and compare it with those of the least squares estimator (LSE) and the GLSE through some simulations. Monte Carlo simulations demonstrate that the PBE using separable covariance functions has superior accuracy to that of the LSE, which does not contain the information of the spatial covariance structure, even if the true process has an isotropic Matérn covariance function. Additionally, our proposed PBE is computationally efficient relative to the GLSE for large spatial data sets.

Key words: Asymptotic covariance matrix, Generalized least squares, Lattice process, Pseudo best estimator, Separable process, Spatial statistics, Spectral density

1 Introduction

Recently, various statistical methods for spatial data have been investigated. Among them, a linear regression model with a spatially correlated error term has played an important role in a wide variety of scientific fields such as geostatistics, econometrics and forestry.

To estimate the coefficients of the linear regression model with a spatially dependent error term, we often use the generalized least squares estimator (GLSE) (see, e.g., Cressie

---

*Graduate School of Economics, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan, Currently at NEC Corporation, 1753, Shimonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666, Japan. E-mail: nennekog@hotmail.co.jp
However, the GLSE for large spatial data sets is computationally expensive, because it involves inverting the covariance matrix of error terms at \( n \) different spatial points, which requires \( O(n^3) \) operations. For example, the Walker Lake dataset in Isaaks and Srivastava (1989) and the soil moisture index derived from other GIS layers in Huang et al. (2010) are observed on a spatial lattice of \( 260 \times 300 \) and \( 100 \times 100 \) regular grid points respectively. Recently, much attention has been paid to the analysis of such large spatial datasets (Furrer et al. 2006; Banerjee et al. 2008; Cressie and Johannesson 2008; Lindgren et al. 2011).

To deal with this problem, we propose a pseudo best estimator (PBE) using separable covariance structures which approximate the true ones. Specifically, we expressed the covariance function as the product of the covariance functions of the causal autoregressive process. The covariance matrix of the PBE has a separable structure and this facilitates the computational procedure for large spatial datasets using the property of the Kronecker product. Moreover, unlike the PBE, the GLSE is infeasible because the true covariance matrix is unknown in practice, meaning that the PBE is also a useful estimator for real spatial data. Genton (2007) considered separable approximation to predict large space-time datasets by solving the nearest Kronecker product for a space-time covariance matrix problem and obtained good predictive performance.

The main contributions of this paper are to propose the PBE using separable covariance functions for efficient computation and derive its asymptotic covariance matrix by extending the technical method of Grenander and Rosenblatt (1957) and Anderson (1971) in time series to a lattice process. As this analysis provides the condition for the asymptotic efficiency of the least squares estimator (LSE) relative to the GLSE, it is easy in the simulation to compare the PBE with the LSE which is computationally efficient as with the PBE. Our work can be regarded as an extension of Amemiya (1973) and Engle (1974), who considered the asymptotic properties of the GLSE when the covariance structure of the true process was incorrectly identified in time series literature. Koreisha and Fang (2001) investigated the finite accuracy of the GLSE and the PBE for time series.

In this paper, we compare the finite accuracy and asymptotic variance of the PBE with those of the LSE and the GLSE in Yajima and Matsuda (2008) through some simulations. In these simulations, the effect of the misspecification of the covariance function for the GLSE is also examined. The LSE is efficient in terms of calculating the estimator for large spatial datasets and Yajima and Matsuda (2008) obtained the necessary and sufficient conditions under which the asymptotic covariance matrix of the LSE is identical to that of the GLSE. However, our simulations illustrate that the PBE outperforms the LSE when these conditions are not satisfied and shows good performance as well as the LSE even if
these conditions hold. In the simulations, the difference in mean squared error between the PBE approximated by separable covariance functions and the GLSE is small, even in the case of the true process with an isotropic Matérn covariance function. Additionally, our proposed PBE is computationally efficient relative to the GLSE.

The remainder of this paper is organized as follows. We introduce a linear regression model and some estimators in Section 2. In Section 3, we review the necessary and sufficient conditions under which the asymptotic covariance matrix of the LSE is identical to that of the GLSE and derive the asymptotic covariance matrix of the PBE. In Section 4, computer experiments are conducted to examine the finite sample performance of our asymptotic result and compare the finite accuracy and asymptotic variance of the PBE with those of the LSE and GLSE. Our conclusions and future studies are discussed in Section 5. Technical proofs of the lemmas and theorem are given in Appendices A and B.

2 Linear regression model and some estimators

For simplicity, we will consider the sampling region to be a square on a plane. Define \( Z = \{0, \pm 1, \pm 2, \ldots \} \). Let \( Z^2 \) be the integer lattice points in the two-dimensional Euclidean space. For \( t = (t_1, t_2)' \in Z^2 \), consider the regression model of the form

\[
y_t = X_t' \beta + \epsilon_t,
\]

where \( \{y_t\} \) is an observed sequence, \( X_t = (x_{t,1}, \ldots, x_{t,p})' \) is a \( p \)-vector of nonstochastic regressors, \( \beta = (\beta_1, \ldots, \beta_p)' \) is a vector of unknown regression coefficients and the prime denotes the transposition. Hereafter, it is assumed that \( (y_t, X_t) \) is observed on the square sampling domain \( P_N = \{t = (t_1, t_2)' \in Z^2 | 1 \leq t_1 \leq N, 1 \leq t_2 \leq N \} \). In this case, the sample size is \( N^2 \). The error terms \( \{\epsilon_t\} \) follow a stationary random field with mean 0 and spectral density function \( f(\lambda) \), \( \lambda = (\lambda_1, \lambda_2)' \in [-\pi, \pi]^2 \). Thus, the covariance function \( \gamma_\epsilon(h) \) of \( \{\epsilon_t\} \) is given by

\[
\gamma_\epsilon(h) = E[\epsilon_t \epsilon_{t+h}] = \int_{\Pi^2} \exp(ih'\lambda)f(\lambda) d\lambda,
\]

where \( \Pi = (-\pi, \pi] \), \( h = (h_1, h_2)' \in Z^2 \) and \( h'\lambda = h_1 \lambda_1 + h_2 \lambda_2 \).

When we estimate coefficients in the linear regression model, the GLSE is given by

\[
\hat{\beta}_{GLSE} = \left(X' \Sigma^{-1} X\right)^{-1} X' \Sigma^{-1} y,
\]

where \( X = (X_{(1,1)}, \ldots, X_{(1,N)}, X_{(2,1)}, \ldots, X_{(2,N)}, \ldots, X_{(N,1)}, \ldots, X_{(N,N)})' \), \( y = (y_{(1,1)}, \ldots, y_{(1,N)}, y_{(2,1)}, \ldots, y_{(2,N)}, \ldots, y_{(N,1)}, \ldots, y_{(N,N)})' \), \( \Sigma = E[\epsilon \epsilon'] \) and \( \epsilon = (\epsilon_{(1,1)}, \ldots, \epsilon_{(1,N)}, \epsilon_{(2,1)}, \ldots, \epsilon_{(2,N)}, \ldots, \epsilon_{(N,1)}, \ldots, \epsilon_{(N,N)})' \).
It is known that the GLSE is the best linear unbiased estimator (BLUE). However, if the true covariance structure is unknown, the GLSE is infeasible. Additionally, the operation count for computing $\Sigma^{-1} \left( N^2 \times N^2 \right)$ of $\hat{\beta}_{\text{GLSE}} \left( = \left( X' \Sigma^{-1} X \right)^{-1} X' \Sigma^{-1} y \right)$ is of order $N^6$. Hence, as the sample size increases, the computation becomes impractical. For example, the Walker Lake dataset in Isaaks and Srivastava (1989) consists of two variables measured at 78000 points on a spatial lattice of $260 \times 300$ regular grid points and the soil moisture index in Huang et al. (2010) is derived from GIS layers on a $100 \times 100$ grid.

To reduce the computational burden, we consider the approximation of the true covariance function by the product of the covariance functions of the causal autoregressive process of order $P$ (AR($P$)) in time series literature. Thus, we obtain the following estimator

$$\hat{\beta}_{\text{PBE}} = \left( X' \hat{\Sigma}^{-1} X \right)^{-1} X' \hat{\Sigma}^{-1} y,$$

where $\hat{\Sigma} = \hat{\Sigma}_1 \otimes \hat{\Sigma}_2$ and $\hat{\Sigma}_i$ is the covariance matrix of causal AR($P_i$) ($i = 1, 2$). This kind of estimator, in which the true covariance matrix is replaced by an incorrect one, is called the pseudo best estimator. Each element of $\hat{\Sigma}$ is denoted by the separable covariance function $\gamma(h_1, h_2) = \gamma_1(h_1) \gamma_2(h_2)$ where $\gamma_i(h_i)$ represents the autocovariance function of AR($P_i$) ($i = 1, 2$). The corresponding spectral density functions are denoted by $g(\lambda_1, \lambda_2)$, $g_1(\lambda_1)$ and $g_2(\lambda_2)$ respectively and it follows that $g(\lambda_1, \lambda_2) = g_1(\lambda_1) g_2(\lambda_2)$. $g(\lambda_1, \lambda_2)$ is a kind of an approximation of $f(\lambda_1, \lambda_2)$. From the property of the Kronecker product (Horn and Johnson 1991; page 244), $\hat{\Sigma}^{-1} = \left( \hat{\Sigma}_1 \otimes \hat{\Sigma}_2 \right)^{-1} = \hat{\Sigma}_1^{-1} \otimes \hat{\Sigma}_2^{-1}$ and we can obtain the exact form of the inverse of the covariance matrix $\hat{\Sigma}_i$ ($i = 1, 2$) given by the autoregressive process (e.g., Anderson 1971; page 576). This makes it much faster to calculate $\hat{\beta}_{\text{PBE}}$ by the separable approximation of the true covariance function.

The LSE is another alternative to the GLSE, because it does not require the inversion of the covariance matrix. The LSE is defined as

$$\hat{\beta}_{\text{LSE}} = \left( X' X \right)^{-1} X' y.$$

We will compare the accuracy of the GLSE, PBE and LSE by evaluating the asymptotic covariance matrix.

3 **Asymptotic properties of $\hat{\beta}_{\text{GLSE}}, \hat{\beta}_{\text{LSE}}$ and $\hat{\beta}_{\text{PBE}}$**

In this section, we derive the asymptotic covariance matrices of $\hat{\beta}_{\text{GLSE}}, \hat{\beta}_{\text{LSE}}$ and $\hat{\beta}_{\text{PBE}}$. First, we introduce some assumptions.
Define

\[ a_{ij}^{(N,N)}(h_1, h_2) = \sum_{t_1=1}^{N-h_1} \sum_{t_2=1}^{N-h_2} x_{(t_1+h_1, t_2+h_2),i} x_{(t_1, t_2),j}, \quad h_1, h_2 \geq 0, \]

\[ = \sum_{t_1=1-h_1}^{N} \sum_{t_2=1-h_2}^{N} x_{(t_1+h_1, t_2+h_2),i} x_{(t_1, t_2),j}, \quad h_1, h_2 \leq 0, \]

\[ = \sum_{t_1=1-h_1}^{N} \sum_{t_2=1}^{N} x_{(t_1+h_1, t_2+h_2),i} x_{(t_1, t_2),j}, \quad h_1 \leq 0, h_2 \geq 0, \]

\[ = \sum_{t_1=1}^{N} \sum_{t_2=1-h_2}^{N} x_{(t_1+h_1, t_2+h_2),i} x_{(t_1, t_2),j}, \quad h_1 \geq 0, h_2 \leq 0. \]

(a) \( a_{ii}^{(N,N)}(0, 0) \to \infty \) as \( N \to \infty \), \( i = 1, \ldots, p \).

(b) \( \lim_{N \to \infty} a_{ii}^{(N+h_1, N+h_2)}(0, 0)/a_{ii}^{(N,N)}(0, 0) = 1 \) for every \( i \) and \( h_1, h_2, i = 1, \ldots, p \) and \( h_1, h_2 \in \mathbb{Z} \).

(c) The limit of

\[ \gamma_{ij}^{(N,N)}(h_1, h_2) = \frac{a_{ij}^{(N,N)}(h_1, h_2)}{\{a_{ii}^{(N,N)}(0,0)\} a_{jj}^{(N,N)}(0,0)} \]

as \( N \to \infty \) exists for every \( i, j \) and \( h_1, h_2 \), \( (i, j) = 1, \ldots, p \) and \( h_1, h_2 \in \mathbb{Z} \).

Let

\[ \rho_{ij}(h_1, h_2) = \lim_{N \to \infty} \gamma_{ij}^{(N,N)}(h_1, h_2) \]

and \( R(h_1, h_2) \) be the \( p \times p \) matrix with \( (i, j) \)th element \( \rho_{ij}(h_1, h_2) \).

(d) \( R(0, 0) \) is nonsingular.

(e) \( \{\epsilon_t\} \) is a unilateral moving average process,

\[ \epsilon_{(t_1, t_2)} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \theta_{l,m} \eta_{(t_1-l, t_2-m)}, \]

where \( \{\theta_{l,m}\} \) satisfies \( \sum_{l,m} |\theta_{l,m}|^2 < \infty \) with \( \theta_{0,0} = 1 \) and \( \{\eta_{(t_1, t_2)}\} \) is white noise with \( \text{Var} (\eta_{(t_1, t_2)}) = \sigma^2_{\eta} \).

Moreover if we define \( \theta(z_1, z_2) = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \theta_{l,m} z_1^l z_2^m \) and \( \phi(z_1, z_2) = \theta(z_1, z_2)^{-1} = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \phi_{l,m} z_1^l z_2^m \), then \( \{\phi_{l,m}\} \) satisfies \( \sum_{l,m} |\phi_{l,m}| < \infty \) with \( \phi_{0,0} = 1 \).

(f) \( f(\lambda) \) is a positive continuous function in \([–\pi, \pi]^2\).

(g) \( N \max_{1 \leq t_1, t_2 \leq N} (x_{(t_1, t_2),i})^2 / a_{ii}^{(N,N)}(0, 0) \to 0 \) as \( N \to \infty \) \( (i = 1, \ldots, p) \).

(h) \( f(\lambda_1, \lambda_2) = f(-\lambda_1, \lambda_2) = f(\lambda_1, -\lambda_2) = f(-\lambda_1, -\lambda_2) \).

We make brief comments on the assumptions. We can view (a)-(d) as a two-dimensional version of so-called Grenander’s conditions on \( X_t \) (Grenander and Rosenblatt 1957; Anderson
1971). Under (a)-(d), there exists a Hermitian matrix function $M(\lambda_1, \lambda_2)$ with positive semidefinite increments such that

$$R(h_1, h_2) = \int_{\Pi^2} \exp(i(h_1\lambda_1 + h_2\lambda_2))dM(\lambda_1, \lambda_2)$$

(see Cohen and Francos 2002). Put

$$m^{(N,N)}_{kl}(\lambda_1, \lambda_2) = \left( \sum_{t_1=1}^{N} \sum_{t_2=1}^{N} x_{(t_1,t_2)} e^{-i(t_1\lambda_1 + t_2\lambda_2)} \right) \left( \sum_{t_1=1}^{N} \sum_{t_2=1}^{N} x_{(t_1,t_2)} e^{i(t_1\lambda_1 + t_2\lambda_2)} \right) \left( \frac{(2\pi)^2}{(a_{kk}^{(N,N)}(0,0)a_{ll}^{(N,N)}(0,0))^{1/2}} \right)$$

and define

$$M^{(N,N)}_{kl}(\lambda_1, \lambda_2) = \int_{-\pi}^{\lambda_1} \int_{-\pi}^{\lambda_2} m^{(N,N)}_{kl}(\omega_1, \omega_2) d\omega_1 d\omega_2.$$

Let $M_{kl}(\lambda_1, \lambda_2)$ and $M^{(N,N)}_{kl}(\lambda_1, \lambda_2)$ be $p \times p$ matrices with $(k,l)$th element of $M(\lambda_1, \lambda_2)$ and $M^{(N,N)}(\lambda_1, \lambda_2)$ respectively for $k,l = 1, \ldots, p$. If we regard $M^{(N,N)}(\lambda_1, \lambda_2)$ and $M(\lambda_1, \lambda_2)$ as matrix measures in $\Pi^2$, $R^{(N,N)}(h_1, h_2) = [\gamma^{(N,N)}_{kl}(h_1, h_2)]$ and $R(h_1, h_2)$ are their characteristic functions respectively. Then (c) implies

$$M^{(N,N)}(\lambda_1, \lambda_2) \xrightarrow{w} M(\lambda_1, \lambda_2)$$

as $N \to \infty$, where $\xrightarrow{w}$ means $M^{(N,N)}(\lambda_1, \lambda_2)$ converges weakly to $M(\lambda_1, \lambda_2)$ (see Ibragimov and Rozanov 1978). Consequently, for any continuous bounded function $\phi(\lambda)$ in $\Pi^2$,

$$\lim_{N \to \infty} \int_{\Pi^2} \phi(\lambda) dM^{(N,N)}(\lambda_1, \lambda_2) = \int_{\Pi^2} \phi(\lambda) dM(\lambda_1, \lambda_2).$$

In accordance with the time series literature, $M(\lambda_1, \lambda_2)$ is called the regression spectral measure of $X_t$ (Taniguchi et al. 2008).

From Yajima and Matsuda (2008), under (e), $\{\epsilon_t\}$ has a unilateral AR representation,

$$\sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \phi_{l,m} \epsilon_{(t_1-l,t_2-m)} = \eta(t_1, t_2),$$

which has a quarter-plane dependence. Bronars and Jansen (1987) applied such a model to unemployment rate fluctuations in the United States. Moreover, (e) holds for a random process with separable covariance functions such as the product of one-dimensional Matérn covariance functions, but a random process with an isotropic Matérn covariance function
does not satisfy (e) (Yajima and Matsuda 2008). The spectral density functions of separable covariance function and isotropic one satisfy (h), which is often referred to as axial symmetry.

Next we derive the asymptotic covariance matrix of the three estimators, that is \( \hat{\beta}_{GLSE} \), \( \hat{\beta}_{PBE} \) and \( \hat{\beta}_{LSE} \).

Define
\[
D_{N^2} = \text{diag}(\|x_1\|, \ldots, \|x_p\|)
\]
where
\[
x_i = (x_{(1,1),i}, \ldots, x_{(1,N),i}, x_{(2,1),i}, \ldots, x_{(2,N),i}, \ldots, x_{(N,1),i}, \ldots, x_{(N,N),i})',
\]
and \(\|x_i\| = (\sum_{t_1=1}^N \sum_{t_2=1}^N (x_{(t_1,t_2),i})^2)^{1/2} (i = 1, \ldots, p)\).

**Theorem 1 (Yajima and Matsuda (2008)).** Under (a)-(g),
\[
\lim_{N \to \infty} D_{N^2} E \left[ (\hat{\beta}_{GLSE} - \beta)(\hat{\beta}_{GLSE} - \beta)' \right] D_{N^2} = (2\pi)^2 \left( \int_{\Pi^2} \frac{1}{f(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right)^{-1}.
\]

**Theorem 2 (Yajima and Matsuda (2008)).** Under (a)-(d) and (f),
\[
\lim_{N \to \infty} D_{N^2} E \left[ (\hat{\beta}_{LSE} - \beta)(\hat{\beta}_{LSE} - \beta)' \right] D_{N^2} = (2\pi)^2 R(0, 0)^{-1} \int_{\Pi^2} f(\lambda_1, \lambda_2) dM(\lambda_1, \lambda_2) R(0, 0)^{-1}.
\]

Moreover, Yajima and Matsuda (2008) gave the following necessary and sufficient conditions for the LSE to be asymptotically efficient relative to the GLSE, which means that the asymptotic covariance matrix of the LSE is identical to that of the GLSE.

**Theorem 3 (Yajima and Matsuda (2008)).** Under (a)-(g), the LSE is asymptotically efficient relative to the GLSE if and only if \( M(\lambda_1, \lambda_2) \) increases at not more than \( p \) values of \( \lambda \) \((\in [0, \pi]^2)\) and the sum of the ranks of the increases in \( M(\lambda_1, \lambda_2) \) is \( p \).

Theorems 1-3 can be regarded as an extension of Grenander and Rosenblatt (1957) for \( d = 1 \) to spatial processes. The following theorem is our main result. The proof is given in Appendix B.

**Theorem 4.** Under (a)-(d) and (f)-(h),
\[
\lim_{N \to \infty} D_{N^2} E \left[ (\hat{\beta}_{PBE} - \beta)(\hat{\beta}_{PBE} - \beta)' \right] D_{N^2} = (2\pi)^2 \left( \int_{\Pi^2} \frac{1}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right)^{-1} \left( \int_{\Pi^2} \frac{f(\lambda_1, \lambda_2)}{g^2(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right) \times \left( \int_{\Pi^2} \frac{1}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right)^{-1}.
\]
Note that the conditions required in Theorems 1-4 are not in common. In particular, unlike Theorems 2 and 4, Theorems 1 and 3 do not hold due to (e) if the random process has an isotropic Matérn covariance function. That kind of separable covariance function satisfies (e), (f) and (h) of Theorems 1-4.

Amemiya (1973) and Engle (1974) investigated the asymptotic properties of \( \hat{\beta}_{PBE} \) for the case \( d = 1 \) and Theorem 4 is an extension of their theoretical results to the spatial case. Additionally, it can be regarded as an extension of Theorem 1 or 2 because the asymptotic covariance matrix in Theorem 4 is identical to that in Theorem 1 or 2 under the appropriate assumptions when \( g(\lambda_1, \lambda_2) = f(\lambda_1, \lambda_2) \) or \( g(\lambda_1, \lambda_2) \) is a constant.

4 Computational experiments

We conduct Monte Carlo simulations using MATLAB. The first experiment examines the convergence and finite sample accuracy of \( \hat{\beta}_{PBE} \) for different sample sizes, separable approximations and true covariance functions. We consider linear regression models with one regressor

\[
y(t_1, t_2) = \beta x(t_1, t_2) + \epsilon(t_1, t_2)
\]

for \( 1 \leq t_1, t_2 \leq N \) and \( \beta = 2 \). For the regressor \( x(t_1, t_2) \) satisfying (a)-(d) and (g), our computational experiments consider the polynomial trend \( x(t_1, t_2) = t_1 t_2 \), the harmonic trend \( x(t_1, t_2) = \cos((\pi/2)t_1) \cos((\pi/2)t_2) \) and the polynomial plus harmonic trend \( x(t_1, t_2) = 1 + \cos((\pi/2)t_1) \cos((\pi/2)t_2) \) (see Toyooka 1985 for the case \( d = 1 \)). These regressors are useful for obtaining row and column effects and a kind of periodicity. The jumps of \( M(\lambda_1, \lambda_2) \) are 1 at \( (\lambda_1, \lambda_2) = (0, 0) \) and 1/4 at \( (\lambda_1, \lambda_2) = (\pi/2, \pi/2), (-\pi/2, \pi/2), (\pi/2, -\pi/2), (-\pi/2, -\pi/2) \) for the polynomial and harmonic trends respectively. For these two regressors, it follows from a routine calculation that the asymptotic variance of the PBE does not depend on separable approximations and the asymptotic variance of the PBE is identical to that of the LSE. As a result, the asymptotic variance of the PBE in each true case is the same for the first and second regressors. In the third regressor, the jumps of \( M(\lambda_1, \lambda_2) \) are 4/5 at \( (\lambda_1, \lambda_2) = (0, 0) \) and 1/20 at \( (\lambda_1, \lambda_2) = (\pi/2, \pi/2), (-\pi/2, \pi/2), (\pi/2, -\pi/2), (-\pi/2, -\pi/2) \). These jump values are used to calculate the asymptotic variance of the GLSE, LSE and PBE.

For the true covariance function \( \gamma_\epsilon(h) \), we consider the following six models. Some models include the Matérn covariance function which is popular in spatial statistics because of its flexibility with spatial data. The first and second models are

\[
\gamma_\epsilon(h) = c(||h||) \quad ||h|| = \sqrt{h_1^2 + h_2^2},
\]
where

\[ c(x) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left( \frac{2\sqrt{\nu}|x|}{\rho} \right)^\nu K_\nu \left( \frac{2\sqrt{\nu}|x|}{\rho} \right) \]

and \( K_\nu(\cdot) \) is the modified Bessel function of the second kind of order \( \nu \). This is the isotropic Matérn covariance function and we set \( \nu = 2, \rho = 3, \sigma^2 = 1 \) and \( \nu = 1, \rho = 3, \sigma^2 = 1 \) respectively. The third model is the product of one-dimensional Matérn covariance functions with \( \nu = 2, \rho = 3, \sigma^2 = 1 \) and \( \nu = 1, \rho = 3, \sigma^2 = 1 \) respectively. The fourth model is the product of the one-dimensional Matérn covariance function with \( \nu = 1, \rho = 3, \sigma^2 = 1 \) and the autocovariance function of AR(2)

\[ c^*(x) = \frac{\sigma^2 \xi_1 \xi_2^2}{(\xi_1 \xi_2 - 1)(\xi_2 - \xi_1)} \left[ \frac{\xi_1^{1-|x|}}{\xi_1 - 1} - \frac{\xi_2^{1-|x|}}{\xi_2 - 1} \right], \]

where \( \xi_1 = (2/3)(1 + \sqrt{3}i), \xi_2 = (2/3)(1 - \sqrt{3}i) \) and \( \sigma^2_* \) is chosen such that \( c^*(0) = 1 \). It is inappropriate to approximate \( c^*(x) \) by that of AR(1) because a successive negative correlation exists. The fifth model is the product of the autocovariance function of AR(2) in the fourth one and that of AR(1)

\[ c^{**}(x) = \frac{\sigma^2_{**} \phi^{|x|}}{1 - \phi^2}, \]

where \( \phi = 0.5 \) and \( \sigma^2_{**} \) is chosen such that \( c^{**}(0) = 1 \). The sixth model is the product of two autocovariance functions of AR(1) with \( \phi = 0.9 \) and a scale parameter such that \( \gamma_0(0) = 1 \). These separable covariance functions are used to check the convergence and finite sample accuracy of Theorem 4 and compare the asymptotic variance of the PBE with that of the GLSE. Because all the models satisfy (f) and (h), Theorem 4 holds in these settings.

Next, we explain the method of the separable approximation to obtain the approximated covariance function \( \gamma(h_1, h_2) \). This is similar to the Yule-Walker estimator in time series literature. Three types are adopted, namely AR(1)×AR(1), AR(1)×AR(2) and AR(2)×AR(2). The first approximation is expressed by

\[ \gamma(h_1, h_2) = \gamma_1(h_1) \gamma_2(h_2) = \frac{\sigma_1^2}{1 - \phi_1^2} \frac{\sigma_2^2}{1 - \phi_2^2} \phi_{|h_1|} \phi_{|h_2|}. \]

Each parameter is given by

\[ \hat{\phi}_1 = \frac{\hat{\gamma}(1,0)}{\hat{\gamma}(0,0)}, \hat{\phi}_2 = \frac{\hat{\gamma}(0,1)}{\hat{\gamma}(0,0)} \quad \text{and} \quad \hat{\sigma}_1^2 = \frac{\hat{\gamma}(0,0)}{\gamma'(0,0; \hat{\phi}_1, \hat{\phi}_2)}, \]
allowing us to use the fixed 
g from 1000 iterations, as well as the asymptotic variance, for \( \hat{\beta} \) with the average parameter values, we calculate the empirical variance multiplied by

\[
\hat{\beta} = \frac{1}{N} \sum_{t_1, t_2} \hat{\epsilon}(t_1, t_2)
\]

and

\[
\gamma'(0, 0; \hat{\phi}_1, \hat{\phi}_2) = \frac{1}{(1 - \hat{\phi}_1^2)(1 - \hat{\phi}_2^2)}.
\]

\( \hat{\sigma}_{12}^2 \) is an estimator of \( \sigma^2 \sigma_2^2 \) and is necessary for the calculation of the approximated spectral density function \( g(\lambda_1, \lambda_2) \). For the second one, the covariance function used in the separable approximation is

\[
\gamma(h_1, h_2) = \gamma_1(h_1)\gamma_2(h_2)
\]

\[
= \frac{\sigma_1^2}{1 - \phi_1^2} \frac{\sigma_2^2}{\xi_1^{-1}} \frac{\xi_2^2}{\xi_2} \left[ \frac{\xi_1^{1-|h_2|}}{\xi_1^{-1}} - \frac{\xi_2^{1-|h_2|}}{\xi_2^{-1}} \right],
\]

where

\[
\hat{\xi}_1 = \hat{a} + \sqrt{\hat{a}^2 + 4\hat{b}}
\]

and

\[
\hat{\xi}_2 = \hat{a} - \sqrt{\hat{a}^2 + 4\hat{b}}.
\]

We generate the error terms \( \{\epsilon(t_1, t_2)\} \) by a multivariate normal distribution with mean \( 0 \) and covariance matrix \( \Sigma \) and the sample size is \( N^2 = 20^2 \) or \( 60^2 \). Next, \( \hat{\beta}_{LSE} \) is calculated 1000 times for \( N = 60 \) and we obtain 1000 sets of parameters in \( \gamma \) from \( \hat{\epsilon}(t_1, t_2) = y(t_1, t_2) - \hat{\beta}_{LSE} x(t_1, t_2) \) by the preceding separable approximation. Using \( \gamma \) and \( g(\lambda_1, \lambda_2) \) with the average parameter values, we calculate the empirical variance multiplied by \( D_{N^2}^2 \) from 1000 iterations, as well as the asymptotic variance, for \( \hat{\beta}_{PBE} \) and \( \hat{\beta}_{LSE} \). Note that the average parameter values calculated from \( \hat{\beta}_{LSE} \) for \( N = 60 \) are used for the case \( N = 20 \), allowing us to use the fixed \( g(\lambda_1, \lambda_2) \) in Theorem 4 throughout Section 4. This improves the
accuracy of the PBE in the case $N = 20$, but does not change the conclusion of the computer experiments, that is, the PBE outperforms the LSE when the average parameter value is calculated from $\hat{\beta}_{LSE}$ for $N = 20$. These simulations are omitted from this paper. When calculating the spectral density function of the Matérn covariance function, we must consider the aliasing phenomenon (Fuentes 2005) because the observations exist on an integer lattice.

Table 1: Summary of results from the first experiment for the polynomial trend.

| case        | $\nu = 2$ | $\nu = 1$ |
|-------------|-----------|-----------|
| approximation | AR(1) × AR(1) | AR(1) × AR(2) | AR(2) × AR(2) | AR(1) × AR(1) | AR(1) × AR(2) | AR(2) × AR(2) |
| $N = 20$    | 18.073    | 16.922    | 16.015    | 16.619    | 16.557    | 16.351    |
| $N = 60$    | 24.288    | 22.934    | 22.085    | 23.315    | 23.283    | 23.334    |
| asymptotic variance | 28.276 | 28.276 | 28.276 | 28.296 | 28.296 | 28.296 |

Table 2: Summary of results from the first experiment for the harmonic trend.

| case        | $\nu = 2$ | $\nu = 1$ |
|-------------|-----------|-----------|
| approximation | AR(1) × AR(1) | AR(1) × AR(2) | AR(2) × AR(2) | AR(1) × AR(1) | AR(1) × AR(2) | AR(2) × AR(2) |
| $N = 20$    | 0.101     | 0.101     | 0.102     | 0.240     | 0.242     | 0.244     |
| $N = 60$    | 0.108     | 0.108     | 0.108     | 0.223     | 0.222     | 0.222     |
| asymptotic variance | 0.103 | 0.103 | 0.103 | 0.222 | 0.222 | 0.222 |

Tables 1 - 3 show that as $N$ increases, the empirical variance multiplied by $D_{N^2}$ goes to the asymptotic variance in all cases, as in Theorem 4. However, in some cases with the polynomial trend regressor, the sample size seems to be insufficient.

Moreover, we can plot the approximated spectral density function $g(\lambda_1, \lambda_2)$ with the average parameter values given in each case for the polynomial plus harmonic trend to check
Table 3: Summary of results from the first experiment for the polynomial plus harmonic trend.

| case | \( \nu = 2 \) | \( \nu = 1 \) |
|------|----------------|----------------|
|      | AR(1) × AR(1) | AR(1) × AR(2)  | AR(2) × AR(2) | AR(1) × AR(1) | AR(1) × AR(2)  | AR(2) × AR(2) |
| N=20 | 0.538          | 0.542          | 0.547         | 1.011         | 1.005          | 1.003         |
| N=60 | 0.497          | 0.498          | 0.498         | 1.069         | 1.066          | 1.063         |
| asymptotic variance | 0.516 | 0.515 | 0.513 | 1.097 | 1.095 | 1.088 |

The fit of \( g \) to the true spectral density function \( f(\lambda_1, \lambda_2) \). Figures 1-3 show \( f \) and \( g \) on \([0, \pi]^2\) because \( f \) and \( g \) are axial symmetry. As the order of the autoregressive process increases, the fit of the spectral density function \( g \) to \( f \) is better at the points \((\lambda_1, \lambda_2)\) where \( M(\lambda_1, \lambda_2) \) has jumps. This is consistent with the accuracy of the PBE in Table 3.

![Figure 1: Approximated spectral density functions](image-url)

Figure 1: Approximated spectral density functions \( g(\lambda_1, \lambda_2) \) on \([0, \pi]^2\) in the approximation of AR(1)×AR(1). The true covariance function corresponds to the isotropic Matérn \((\nu = 2)\), the isotropic Matérn \((\nu = 1)\) and \((\nu = 2)\)×\((\nu = 1)\) for the left, middle and right column in the top row respectively. Similarly, it corresponds to \((\nu = 1)\)×AR(2), AR(1)×AR(2) and AR(1)×AR(1) for the left, middle and right column in the bottom row respectively.
Figure 2: Approximated spectral density functions \( g(\lambda_1, \lambda_2) \) on \([0, \pi]^2\) in the approximation case of AR(1)×AR(2). The true covariance function corresponds to the isotropic Matérn \((\nu = 2)\), the isotropic Matérn \((\nu = 1)\) and \((\nu = 2)\times(\nu = 1)\) for the left, middle and right column in the top row respectively. Similarly, it corresponds to \((\nu = 1)\times\text{AR}(2), \text{AR}(1)\times\text{AR}(2)\) and \(\text{AR}(1)\times\text{AR}(1)\) for the left, middle and right column in the bottom row respectively.

The second experiment compares the finite accuracy and asymptotic variance of the PBE with those of the LSE and the GLSE under the same settings as the first experiment. To see the finite accuracy and asymptotic variance of the three estimators, we calculate the ratios of the empirical variance of the PBE and LSE to that of the GLSE, as well as the theoretical ratios given by the asymptotic variances in Theorems 1, 2 and 4. The empirical variances of the GLSE, LSE and PBE and \(g(\lambda_1, \lambda_2)\) are as in the first experiment. The random processes with true covariance functions except the isotropic Matérn covariance function and \((\nu = 1)\times\text{AR}(2)\) satisfy (e) and all the models satisfy (f) and (h). Therefore, with the isotropic Matérn covariance function and \((\nu = 1)\times\text{AR}(2)\) case, Theorems 1 and 3 do not hold for each regressor. Moreover, for the first and second regressors, it follows from Theorem 3 and a routine calculation that the LSE and PBE are asymptotically efficient except with the isotropic Matérn covariance function and \((\nu = 1)\times\text{AR}(2)\) case. Therefore, we show only the empirical ratios in tables for the first and second regressors and the isotropic Matérn covariance function and \((\nu = 1)\times\text{AR}(2)\) case for the third regressor. Similarly, from Theorem 3, the LSE is not asymptotically efficient for the third regressor except with the isotropic Matérn covariance function and \((\nu = 1)\times\text{AR}(2)\), so that the asymptotic variance of the LSE is different from that of the GLSE. The asymptotic efficiency of the LSE in
Figure 3: Approximated spectral density functions $g(\lambda_1, \lambda_2)$ on $[0, \pi]^2$ in the approximation case of AR(2)×AR(2). The true covariance function corresponds to the isotropic Matérn ($\nu = 2$), the isotropic Matérn ($\nu = 1$) and ($\nu = 2$)×($\nu = 1$) for the left, middle and right column in the top row respectively. Similarly, it corresponds to ($\nu = 1$)×AR(2), AR(1)×AR(2) and AR(1)×AR(1) for the left, middle and right column in the bottom row respectively.

the isotropic Matérn cases and ($\nu = 1$)×AR(2) case is a subject for future work. In subsequent tables, we omit the empirical or asymptotic variances of the LSE, except for the separable approximation AR(1)×AR(1) because they do not depend on the three separable approximations.

Tables 4 and 5 summarize the results for the first and second regressors. As $N$ increases, the finite efficiency goes to 1 based on theoretical results except with the isotropic Matérn covariance function and ($\nu = 1$)×AR(2) case. In the isotropic and ($\nu = 1$)×AR(2) cases of this simulation, the finite efficiency is close to 1. When AR(1)×AR(2) and AR(1)×AR(1) are the true covariance functions, the PBE of the corresponding approximation is often superior to those in other cases. Moreover, although the asymptotic variances of the LSE and PBE are equal to each other, the PBE outperforms the LSE in many of the $N = 60$ cases in terms of the ratio of the empirical variances. This is because the PBE includes information about the approximated spatial correlation structure. Throughout these simulations, as the order of the autoregressive process in the separable approximation is larger, the efficiency tends to be often better.

Table 6 shows the results for the polynomial plus harmonic trend. (·) denotes the theoretical ratio of the asymptotic variances of the LSE and PBE in Theorems 2 and 4 to that of
Table 4: Summary of results from the second experiment for the polynomial trend.

| case            | Matérn (ν = 2) | Matérn (ν = 1) | (ν = 2)×(ν = 1) |
|-----------------|----------------|----------------|-----------------|
|                 | approximation | N              | LSE            | PBE            | LSE          | PBE            |
| AR(1) × AR(1)   | 20             | 1.328          | 1.161          | 1.214          | 1.039        | 1.187          | 1.159          |
|                 | 60             | 1.108          | 1.116          | 1.137          | 1.022        | 1.066          | 1.107          |
| AR(1) × AR(2)   | 20             | -              | 1.087          | -              | 1.047        | -              | 1.131          |
|                 | 60             | -              | 1.054          | -              | 1.021        | -              | 1.081          |
| AR(2) × AR(2)   | 20             | -              | 1.029          | -              | 1.035        | -              | 1.001          |
|                 | 60             | -              | 1.015          | -              | 1.023        | -              | 1.001          |
| (ν = 1)×AR(2)   | AR(1)×AR(2)   | AR(1)×AR(1)   |                 |                |               |                |                |
|                 | approximation | N              | LSE            | PBE            | LSE          | PBE            |
| AR(1) × AR(1)   | 20             | 1.145          | 1.204          | 1.066          | 1.172        | 1.929          | 1.013          |
|                 | 60             | 1.073          | 1.099          | 1.053          | 1.090        | 1.379          | 1.002          |
| AR(1) × AR(2)   | 20             | -              | 1.023          | -              | 1.0          | -              | 1.015          |
|                 | 60             | -              | 1.007          | -              | 1.0          | -              | 1.005          |
| AR(2) × AR(2)   | 20             | -              | 1.001          | -              | 1.004        | -              | 1.012          |
|                 | 60             | -              | 1.001          | -              | 1.0          | -              | 1.007          |

Table 5: Summary of results from the second experiment for the harmonic trend.

| case            | Matérn (ν = 2) | Matérn (ν = 1) | (ν = 2)×(ν = 1) |
|-----------------|----------------|----------------|-----------------|
|                 | approximation | N              | LSE            | PBE            | LSE          | PBE            |
| AR(1) × AR(1)   | 20             | 1.178          | 1.034          | 1.040          | 1.028        | 1.483          | 1.044          |
|                 | 60             | 1.064          | 1.008          | 1.032          | 1.003        | 1.151          | 1.016          |
| AR(1) × AR(2)   | 20             | -              | 1.034          | -              | 1.038        | -              | 1.036          |
|                 | 60             | -              | 1.012          | -              | 1.001        | -              | 1.017          |
| AR(2) × AR(2)   | 20             | -              | 1.042          | -              | 1.047        | -              | 1.011          |
|                 | 60             | -              | 1.011          | -              | 1.004        | -              | 1.004          |
| (ν = 1)×AR(2)   | AR(1)×AR(2)   | AR(1)×AR(1)   |                 |                |               |                |                |
|                 | approximation | N              | LSE            | PBE            | LSE          | PBE            |
| AR(1) × AR(1)   | 20             | 1.343          | 1.103          | 1.203          | 1.059        | 1.770          | 1.0            |
|                 | 60             | 1.119          | 1.033          | 1.075          | 1.032        | 1.271          | 1.0            |
| AR(1) × AR(2)   | 20             | -              | 1.008          | -              | 1.0          | -              | 1.0            |
|                 | 60             | -              | 1.003          | -              | 1.0          | -              | 1.0            |
| AR(2) × AR(2)   | 20             | -              | 1.001          | -              | 1.0          | -              | 1.007          |
|                 | 60             | -              | 1.0            | -              | 1.0          | -              | 1.005          |
Table 6: Summary of results from the second experiment for the polynomial plus harmonic trend.

| approximation | case | Matérn (ν = 2) | Matérn (ν = 1) | (ν = 2) × (ν = 1) |
|---------------|------|----------------|----------------|------------------|
|               | N    | LSE PBE        | LSE PBE        | LSE PBE          |
| AR(1) × AR(1) | 20   | 37.347 1.045  | 18.060 1.049  | 54.357 1.044     |
|               | 60   | 42.525 1.008  | 20.083 1.007  | 65.797 1.018     |
| AR(1) × AR(2) | 20   | - 1.054      | - 1.043      | - 1.044          |
|               | 60   | - 1.009      | - 1.005      | - 1.014          |
| AR(2) × AR(2) | 20   | - 1.064      | - 1.041      | - 1.015          |
|               | 60   | - 1.010      | - 1.002      | - 1.006          |

Table 7: Time required to calculate each estimator for the three approximations (seconds).

| estimation procedure (sec.) | LSE | GLSE | PBE(AR(1)×AR(1)) | PBE(AR(1)×AR(2)) | PBE(AR(2)×AR(2)) |
|-----------------------------|-----|------|-------------------|-------------------|-------------------|
| calculation time of estimators (sec.) | 4.43 × 10⁻⁴ | 384.793 | 2.368 | 2.404 | 2.399 |

the GLSE in Theorem 1. As N increases, the finite efficiency goes to the value of (.) except in the isotropic Matérn cases and (ν = 1) × AR(2) case. Because the LSE is not asymptotically efficient for the third regressor, its performance is poor in this case. However, the efficiency of the PBE is close to 1 in both the empirical and theoretical ratios. In particular, the PBE shows good performance even in the cases of the isotropic Matérn class. In fact, the asymptotic variances of the LSE in Table 6, which are 22.642 and 22.681 with ν = 2 and ν = 1 respectively, are much larger than those of the PBE in Table 3. Other properties are similar to those in the preceding simulations.

Finally we examine the computational time of $\hat{\beta}_{GLSE}$, $\hat{\beta}_{LSE}$ and $\hat{\beta}_{PBE}$. We set N = 100 and adopt the isotropic Matérn covariance function with ν = 2, ρ = 3, σ² = 1 for the polynomial plus harmonic trend. All computations are carried out on Linux powered 3.33GHz Xeon processor with 8 Gbytes RAM. From Table 7, we see that the LSE is computationally efficient, but Table 8 shows it can suffer poor performance. The computation time of the PBE including the estimation procedure is faster than that of the GLSE and the empirical ratios of the PBE relative to the GLSE are close to 1.0 in Table 9.


5 Conclusion and future studies

We have proposed the PBE using the separable approximation of the covariance function on lattice data as an alternative estimator of the GLSE which is practically infeasible owing to its computational burden and the unknown covariance structure. We derived the asymptotic covariance matrix of the PBE and examined the effect of the misspecification of the covariance function in the GLSE. The PBE by the separable approximation works well in many simulations even if the true covariance function is isotropic. In particular, when the LSE is not asymptotically efficient, the PBE exhibits superior performance. Moreover, the PBE substantially reduces the computation time because of its separable structure.

In future work, we will present a theoretical comparison of the asymptotic covariance matrices of the PBE and LSE. In addition, the extension of the true process to strongly dependent random fields should be considered.

Appendix A : Technical Lemmas

To derive the asymptotic covariance matrix of \( \hat{\beta}_{PBE} \), we shall prove the following two lemmas.

**Lemma A.1.** Let \( (\pi_{1,1}^{m_1,m_2}, \pi_{1,2}^{m_1,m_2}, \ldots, \pi_{1,N}^{m_1,m_2}, \pi_{2,1}^{m_1,m_2}, \ldots, \pi_{N,1}^{m_1,m_2}, \ldots, \pi_{N,N}^{m_1,m_2})' \) be the solution of

\[
\hat{\Sigma} \begin{pmatrix}
\pi_{1,1}^{m_1,m_2} \\
\vdots \\
\pi_{N,N}^{m_1,m_2}
\end{pmatrix} = \begin{pmatrix}
\gamma(m_1, m_2) \\
\gamma(m_1, m_2 + 1) \\
\vdots \\
\gamma(m_1, m_2 + N - 1) \\
\gamma(m_1 + 1, m_2) \\
\vdots \\
\gamma(m_1 + N - 1, m_2 + N - 1)
\end{pmatrix}, \quad m_1, m_2 \in \mathbb{Z}, \quad (A.1)
\]

where \( \hat{\Sigma} \) and \( \gamma \) correspond to those in the definition of \( \hat{\beta}_{PBE} \) in Section 2. For fixed \( m_1 \geq -N + 1 \) and \( m_2 \geq -N + 1 \),

\[
\sup_N \sum_{i=1}^{N} \sum_{j=1}^{N} |\pi_{i,j}^{m_1,m_2}| < \infty.
\]

**Proof.** (A.1) is expressed by

\[
\left( \hat{\Sigma}_1 \otimes \hat{\Sigma}_2 \right) \begin{pmatrix}
\pi_{1,1}^{m_1,m_2} \\
\vdots \\
\pi_{N,N}^{m_1,m_2}
\end{pmatrix} = \begin{pmatrix}
\gamma_1(m_1) \\
\vdots \\
\gamma_1(m_1 + N - 1)
\end{pmatrix} \otimes \begin{pmatrix}
\gamma_2(m_2) \\
\vdots \\
\gamma_2(m_2 + N - 1)
\end{pmatrix}.
\]
Then, from the property of the Kronecker product (see Horn and Johnson 1991; page 244),

\[
\begin{pmatrix}
\pi_{1,1}^{m_1,m_2} \\
\vdots \\
\pi_{N,N}^{m_1,m_2}
\end{pmatrix} = \left( \hat{\Sigma}_1^{-1} \begin{pmatrix} \gamma_1(m_1) \\ \vdots \\ \gamma_1(m_1 + N - 1) \end{pmatrix} \right) \otimes \left( \hat{\Sigma}_2^{-1} \begin{pmatrix} \gamma_2(m_2) \\ \vdots \\ \gamma_2(m_2 + N - 1) \end{pmatrix} \right).
\]

Now, let \((\pi_{1,k}^{m_k}, \pi_{2,k}^{m_k}, \ldots, \pi_{N,k}^{m_k})'\) be the solution of

\[
\hat{\Sigma}_k \begin{pmatrix} \pi_{1,k}^{m_k} \\ \vdots \\ \pi_{N,k}^{m_k} \end{pmatrix} = \begin{pmatrix} \gamma_k(m_k) \\ \vdots \\ \gamma_k(m_k + N - 1) \end{pmatrix}
\]

for \(k = 1, 2\). We will prove

\[
\sup_{N} \sum_{i=1}^{N} |\pi_{i,k}^{m_k}| < \infty \quad (m_k \geq -N + 1) \tag{A.2}
\]

by mathematical induction. In \(-N + 1 \leq m_k \leq 0 \ (k = 1, 2)\), \(\pi_{i,k}^{m_k} = \delta_{i,-(m_k+1)}\) from the uniqueness of \((\pi_{1,k}^{m_k}, \pi_{2,k}^{m_k}, \ldots, \pi_{N,k}^{m_k})'\) where \(\delta_{i,-(m_k+1)}\) is 1 if \(i = -m_k + 1\), otherwise 0. Therefore, (A.2) is clear in this case. When \(m_k = 1\),

\[
\hat{\Sigma}_k \begin{pmatrix} \pi_{1,k}^{1} \\ \vdots \\ \pi_{N,k}^{1} \end{pmatrix} = \begin{pmatrix} \gamma_k(1) \\ \vdots \\ \gamma_k(N) \end{pmatrix}.
\]

If \(z_t\) is AR(\(P_k\)) with the autocovariance function \(\gamma_k\), that is \(z_t = \phi_{1,k}z_{t-1} + \cdots + \phi_{P_k,k}z_{t-P_k} + \eta_t\) where \(\eta_t\) is white noise, \(\pi_{j,k}^{1} = \phi_{j,k} \ (j = 1, \ldots, P_k)\) and \(\pi_{j,k}^{1} = 0 \ (j > P_k)\). Therefore, \(\sup_{N} \sum_{i=1}^{N} |\pi_{i,k}^{1}| < \infty\). Next, assume that (A.2) holds for \(m_k = m\). Consider \(m_k = m + 1\). Then, from a routine calculation, \(P_{\mathcal{P}(z_1,\ldots,z_N)\mathbb{I}^{N+m+1}} = \sum_{l=1}^{N} \pi_{l,k}^{m+1} z_{N+1-l}\) where \(P_{\mathcal{P}(z_1,\ldots,z_N)}\) is the projection to the closed subspace \(\mathcal{P}(z_1,\ldots,z_N)\) spanned by \((z_1,\ldots,z_N)\). Moreover, from a similar calculation,

\[
P_{\mathcal{P}(z_1,\ldots,z_N)\mathbb{I}^{N+m+1}} = P_{\mathcal{P}(z_1,\ldots,z_N)} \left( \sum_{l=1}^{N+1} \pi_{l,k}^{m} z_{N+2-l} \right)
\]

\[
= P_{\mathcal{P}(z_1,\ldots,z_N)} \pi_{1,k}^{m} z_{N+1} + \sum_{l=2}^{N+1} \pi_{l,k}^{m} z_{N+2-l}
\]

\[
= \pi_{1,k}^{m} \left( \sum_{l=1}^{N} \pi_{l,k}^{1} z_{N+1-l} \right) + \sum_{l=2}^{N+1} \pi_{l,k}^{m} z_{N+2-l}.
\]
Therefore, $\pi_{i,k}^{m+1} = \pi_{i,k}^{m} \pi_{i,k}^{1} + \pi_{i+1,k}^{m} (l = 1, \ldots, N)$. In this case,

$$\sup N \sum_{i=\text{index}} |\pi_{i,k}^{m+1}| \leq \sup N \sum_{i=\text{index}} |\pi_{i,k}^{m}| |\pi_{i+1,k}^{1}| + \sup N \sum_{i=\text{index}} |\pi_{i+1,k}^{m} < \infty.$$  

(A.2) holds for $m_k = m + 1$. Note that $\pi_{i,j}^{m_1,m_2} = \pi_{i,1}^{m_1} \pi_{j,2}^{m_2}$ for $1 \leq i, j \leq N$. Then,

$$\sum_{i=\text{index}} \sum_{j=\text{index}} |\pi_{i,j}^{m_1,m_2}| = \sum_{i=\text{index}} |\pi_{i,1}^{m_1}| \sum_{j=\text{index}} |\pi_{j,2}^{m_2}|$$

is bounded for $N$. 

**Lemma A.2.** Suppose that $h(\lambda_1, \lambda_2)$ is a continuous function on $[-\pi, \pi]^2$ and $h(\lambda_1, \lambda_2) = h(-\lambda_1, \lambda_2) = h(\lambda_1, -\lambda_2) = h(-\lambda_1, -\lambda_2)$. Then, for any sufficiently small $\epsilon > 0$, there exist $h_L(\lambda_1, \lambda_2)$ and $h_U(\lambda_1, \lambda_2)$ such that

$$h_L(\lambda_1, \lambda_2) = \sum_{k_1=\text{index}} K \sum_{k_2=\text{index}} a_{k_1,k_2} e^{i(k_1\lambda_1+k_2\lambda_2)}, \quad a_{k_1,k_2} = a_{-k_1,-k_2} \in \mathbb{R},$$

$$h_U(\lambda_1, \lambda_2) = \sum_{k_1=\text{index}} K \sum_{k_2=\text{index}} b_{k_1,k_2} e^{i(k_1\lambda_1+k_2\lambda_2)}, \quad b_{k_1,k_2} = b_{-k_1,-k_2} \in \mathbb{R},$$

$$h_L(\lambda_1, \lambda_2) \leq h(\lambda_1, \lambda_2) \leq h_U(\lambda_1, \lambda_2)$$

and

$$h_U(\lambda_1, \lambda_2) - h_L(\lambda_1, \lambda_2) \leq \epsilon, \quad (\lambda_1, \lambda_2) \in [-\pi, \pi]^2.$$  

**Proof.** We set $D_{n_1,n_2}(y_1, y_2) = \sum_{|j_1| \leq n_1} \sum_{|j_2| \leq n_2} e^{i(j_1 y_1 + j_2 y_2)} = D_{n_1}(y_1) D_{n_2}(y_2)$ where $D_{n_i}(y_i) = \sum_{|j_i| \leq n_i} e^{ij_i y_i}$ ($i = 1, 2$). $D_{n_i}(y_i)$ is the Dirichlet kernel,

$$D_{n_i}(y_i) = \begin{cases} \sin[(n_i + \frac{1}{2}) y_i] & \text{if } y_i \neq 0, \\ \frac{y_i}{2n_i + 1} & \text{if } y_i = 0. \end{cases}$$

Additionally, consider

$$S_{n_1,n_2} h(\lambda_1', \lambda_2) = \sum_{|j_1| \leq n_1} \sum_{|j_2| \leq n_2} \left( \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda_1, \lambda_2) e^{-i(j_1 \lambda_1 + j_2 \lambda_2)} d\lambda_1 d\lambda_2 \right) e_{j_1,j_2}$$

$$= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda_1, \lambda_2) D_{n_1}(\lambda_1' - \lambda_1) D_{n_2}(\lambda_2' - \lambda_2) d\lambda_1 d\lambda_2,$$

where $e_{j_1,j_2} = e^{i(j_1 \lambda_1 + j_2 \lambda_2)}$. By defining $h(\lambda_1, \lambda_2) = h(\lambda_1 + 2\pi n_1', \lambda_2 + 2\pi n_2') (n_1', n_2' \in \mathbb{Z})$, this can be rewritten as

$$S_{n_1,n_2} h(\lambda_1', \lambda_2) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda_1' - \lambda_1, \lambda_2' - \lambda_2) D_{n_1}(\lambda_1) D_{n_2}(\lambda_2) d\lambda_1 d\lambda_2.$$
Moreover,
\[
\frac{1}{n_1 n_2} \sum_{i_1=0}^{n_1-1} \sum_{i_2=0}^{n_2-1} S_{i_1,i_2} h(\lambda_1', \lambda_2') = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h(\lambda_1' - \lambda_1, \lambda_2' - \lambda_2) K_{n_1}(\lambda_1) K_{n_2}(\lambda_2) d\lambda_1 d\lambda_2,
\]
where \( K_{n_i}(y_i) \) is defined by
\[
\frac{1}{2\pi n_i} \sum_{j_i=0}^{n_i-1} D_{j_i}(y_i) \quad (i = 1, 2)
\]
and is called the Fejer kernel. Finally, it follows from the argument extended from Theorem 2.11.1 of Brockwell and Davis (1991; page 69) and the axial symmetry of \( h(\lambda_1, \lambda_2) \) that for any \( \epsilon > 0 \),
\[
\left| h(\lambda_1, \lambda_2) - \frac{1}{n_1 n_2} \sum_{i_1=0}^{n_1-1} \sum_{i_2=0}^{n_2-1} S_{i_1,i_2} h(\lambda_1, \lambda_2) \right| < \epsilon,
\]
uniformly on \([-\pi, \pi]^2\) for sufficiently large \( n_1 \) and \( n_2 \). In that case,
\[
\frac{1}{n_1 n_2} \sum_{i_1=0}^{n_1-1} \sum_{i_2=0}^{n_2-1} S_{i_1,i_2} h(\lambda_1, \lambda_2) = \sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1,j_2} e^{j_1,j_2},
\]
where \( c_{j_1,j_2} = c_{-j_1,j_2} = c_{j_1,-j_2} = c_{-j_1,-j_2} \) from the axial symmetry of \( h(\lambda_1, \lambda_2) \). Therefore, \( c_{j_1,j_2}'s \) are real and
\[
\sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1,j_2} e^{j_1,j_2} = \sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1,j_2} (\cos(j_1 \lambda_1 + j_2 \lambda_2) + i \sin(j_1 \lambda_1 + j_2 \lambda_2)).
\]
Now, it follows that
\[
\sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1,j_2} \sin(j_1 \lambda_1 + j_2 \lambda_2) = 0.
\]
Therefore, \( \sum_{|j_1| \leq n_1-1} \sum_{|j_2| \leq n_2-1} c_{j_1,j_2} e^{j_1,j_2} \) is real. For any \( \epsilon > 0 \),
\[
\sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1,j_2} e^{j_1,j_2} - \frac{\epsilon}{2} \leq h(\lambda_1, \lambda_2) \leq \sum_{j_1=-(n_1-1)}^{n_1-1} \sum_{j_2=-(n_2-1)}^{n_2-1} c_{j_1,j_2} e^{j_1,j_2} + \frac{\epsilon}{2}
\]
for sufficiently large \( n_1 \) and \( n_2 \). Therefore, setting \( K_1 = n_1 - 1, K_2 = n_2 - 1, a_{k_1,k_2} = b_{k_1,k_2} = c_{k_1,k_2} \) except \( a_{0,0} = c_{0,0} - \epsilon/2, b_{0,0} = c_{0,0} + \epsilon/2 \), the proof is completed.

\[ \square \]

**Appendix B : Proof of Theorem 4**

We prove Theorem 4 by extending the arguments of Anderson (1971) and Yajima (1994).
Proof of Theorem 4

\[
D_N^2 E \left[ (\hat{\beta}_{PBE} - \beta)(\hat{\beta}_{PBE} - \beta)' \right] D_N^2 = \left( D_N^2 X' \hat{\Sigma}^{-1} X D_N^2 \right)^{-1} \left( \left( D_N^2 X' \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} X D_N^2 \right) \left( D_N^2 X' \hat{\Sigma}^{-1} X D_N^2 \right)^{-1} \right).
\]

From Theorem 1, the first and third terms converge to

\[
\left( D_N^2 X' \hat{\Sigma}^{-1} X D_N^2 \right)^{-1} \to \left( \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{1}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \right)^{-1}
\]
as \( N \to \infty \). Here, we put

\[
f(\lambda_1, \lambda_2) = \frac{f(\lambda_1, \lambda_2)}{g(\lambda_1, \lambda_2)} \times g(\lambda_1, \lambda_2)
\]

\[= h(\lambda_1, \lambda_2) \times g(\lambda_1, \lambda_2).\]

First, consider the case of

\[
h(\lambda_1, \lambda_2) = \sum_{l_1=-L_1}^{L_1} \sum_{l_2=-L_2}^{L_2} b_{l_1,l_2} e^{i(l_1\lambda_1 + l_2\lambda_2)}, \quad b_{l_1,l_2} = b_{-l_1,-l_2}.
\]

In this case,

\[
E[\epsilon(j_1,j_2)\epsilon(k_1,k_2)] = \gamma_\epsilon(j_1-k_1, j_2-k_2)
\]

\[= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i((j_1-k_1)\lambda_1 + (j_2-k_2)\lambda_2)} f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2
\]

\[= \sum_{l_1=-L_1}^{L_1} \sum_{l_2=-L_2}^{L_2} b_{l_1,l_2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i((j_1-k_1+1)\lambda_1 + (j_2-k_2+1)\lambda_2)} g(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2.
\]

Therefore,

\[
\Sigma = \sum_{l_1=-L_1}^{L_1} \sum_{l_2=-L_2}^{L_2} b_{l_1,l_2} \tilde{\Sigma}(l_1, l_2),
\]

21
where

\[
\tilde{\Sigma}(l_1, l_2) = \\
\begin{pmatrix}
\gamma(l_1, l_2) & \gamma(l_1, l_2 - 1) & \cdots & \gamma(l_1 + 1 - N, l_2) \\
\gamma(l_1, l_2 + 1) & \gamma(l_1, l_2) & \cdots & \gamma(l_1 + 1 - N, l_2 + 1) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(l_1 + N - 1, l_2) & \gamma(l_1 + N - 1, l_2 - 1) & \cdots & \gamma(l_1, l_2) \\
\gamma(l_1 + N - 1, l_2 + N - 1) & \gamma(l_1 + N - 1, l_2 + N - 2) & \cdots & \gamma(l_1, l_2 + N - 1) \\
\cdots & \cdots & \cdots & \cdots \\
\gamma(l_1, l_2 + 2 - N) & \gamma(l_1, l_2 + 1 - N) & \cdots & \gamma(l_1, l_2 + 1 + N) \\
\cdots & \cdots & \cdots & \cdots \\
\gamma(l_1, l_2 + 1) & \gamma(l_1, l_2) & \cdots & \gamma(l_1, l_2 + 2 - N) \\
\end{pmatrix}.
\]

Consider the case of \(l_1 \geq 0\) and \(l_2 \geq 0\) for sufficiently large \(N\). In this case, the \((m, n)\)th element \(\left(D_{N^2}^{-1} X \tilde{\Sigma}^{-1} \Sigma(l_1, l_2) \tilde{\Sigma}^{-1} X D_{N^2}^{-1}\right)_{m,n}\) is

\[
\frac{1}{\left(\frac{(N,N)}{a_{nm}}(0,0) \frac{(N,N)}{a_{mn}}(0,0)\right)^{1/2}} \left\{ (x'_{(1,:),m}, \ldots, x'_{(N,:),m}) \tilde{\Sigma}^{-1} \\
\times \left( (x_{(1,:),l_1+1,N,1}, \ldots, x_{(1,:),l_1+1,N,1})_{l_1+1-N,1}, \cdots, \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{i,j}^{l_1+1-N,1} x_{(N+1-i,N+1-j),n} \cdots \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{i,j}^{0,l_2} x_{(N+1-i,N+1-j),n} \right) \\
\times (x_{(N+1-i,N+1-j),n}, \ldots, 0, \ldots, 0, \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{i,j}^{0,1} x_{(N+1-i,N+1-j),n} \cdots \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{i,j}^{0,l_2} x_{(N+1-i,N+1-j),n} \right) \right\}^{1/2}
\]

where \(x_{(i,:),m} = (x_{(i,1),m}, \ldots, x_{(i,N),m})'\) \((i = 1, \ldots, N)\) and \(\{\pi_{i,j}^{m_1,m_2}\}\) is defined in Lemma \(A.1\).

Now, \(g_j(\lambda_j)\) is expressed by

\[
\frac{\sigma_j^2}{2\pi} \frac{1}{\left|\phi_j(e^{-i\lambda_j})\right|^2},
\]

22
where \( \phi_j(z) = 1 - \phi_{j,1}z - \cdots - \phi_{j,P_j}z^{P_j} \) and \( \phi_j(z) \neq 0 \) for \(|z| \leq 1 \) \((j = 1, 2)\). From Anderson (1971; page 576), for \( i = 1, 2 \), \( \Sigma_i^{-1} = B'_iB_i/\sigma_i^2 \) where

\[
B_i = \begin{pmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & -\phi_{i,0} \end{pmatrix}
\]

\( \phi_{i,0} = -1 \) and \( b_{i,1}, b_{i,2}, \ldots, b_{i,P_i} \) are chosen so that \( \tilde{\Sigma}_i^{-1} = B'_iB_i/\sigma_i^2 \). Moreover, from the properties of the Kronecker product (Horn and Johnson 1991; page 244), \( \tilde{\Sigma}^{-1} = (\tilde{\Sigma}_1 \otimes \tilde{\Sigma}_2)^{-1} = \tilde{\Sigma}_1^{-1} \otimes \tilde{\Sigma}_2^{-1} = (B'_1 \otimes B'_2)(B_1 \otimes B_2)/(\sigma_1^2 \sigma_2^2) = (B'_1 \otimes B'_2)(B_1 \otimes B_2)/(\sigma_1^2 \sigma_2^2) = (B'_1 \otimes B'_2)(B_1 \otimes B_2)/(\sigma_1^2 \sigma_2^2) \).

By substituting this expression into the first term of (B.1),

\[
(\text{the first term of } \text{(B.1)}) = \frac{1}{\sigma_1^2 \sigma_2^2 (\ell^2_{NM} (0, 0) b_{NM} (0, 0))^{1/2}} \times \left[ \sum_{j_1=1}^{P_1} \sum_{i_1=1}^{j_1} \sum_{i_2=1}^{j_1} b_{1,j_1-i_1} b_{1,j_1-i_2} \left\{ \sum_{k=1}^{P_2} \left( \sum_{\alpha=1}^{k} b_{2,\alpha} x_{(i_1, \alpha), m} \right) \left( \sum_{\beta=1}^{k} b_{2,\beta} x_{(i_2+i_1, \beta, \beta), n} \right) \right\} + \sum_{k=1}^{N-P_2} \left( \sum_{\alpha=0}^{P_2-1} (-\phi_{2,P_2-\alpha}) x_{(i_1+k+1), m} \right) \left( \sum_{\beta=0}^{P_2} (-\phi_{2,P_2-\beta}) x_{(i_2+i_1+k+i_2, \beta, \beta), n} \right) \right] \left[ \sum_{i_1=0}^{N-P_2} \sum_{i_2=0}^{P_1} (-\phi_{1,P_1-i_1})(-\phi_{1,P_1-i_2}) x_{(i_1+a+i_2, \alpha, m)} \right] \left[ \sum_{a=1}^{k} b_{2,\alpha} x_{(i_1+a, \alpha, m)} \right] \times \left( \sum_{\beta=0}^{N-P_2} (-\phi_{2,P_2-\beta}) x_{(i_1+a+i_2+k+i_2+\beta), m} \right) \left( \sum_{\alpha=1}^{P_2} b_{2,\alpha} x_{(i_1+a+k, \alpha, m)} \right) \right]
\]
It follows from (a)-(d) and a routine calculation that \( (B.2) \) converges to
\[
\frac{1}{\sigma_1^2 \sigma_2^2} \sum_{i_1=0}^{P_1} \sum_{i_2=0}^{P_1} (-\phi_1, P_1 - i_1)(-\phi_1, P_1 - i_2) \sum_{a=0}^{P_2} \sum_{\beta=0}^{P_2} (-\phi_2, P_2 - a)(-\phi_2, P_2 - \beta) \\
\times \int_{\mathbb{R}^2} e^{i(-l_1 - i_2 + i_1)\lambda_1 e^{i(-l_2 - \beta + \alpha)}\lambda_2} dM_{mn}(\lambda_1, \lambda_2)
\]
\[
= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \frac{e^{-i(l_1 \lambda_1 + l_2 \lambda_2)}}{g_1(\lambda_1)g_2(\lambda_2)} dM_{mn}(\lambda_1, \lambda_2)
\]
as \( N \to \infty \). Next,
\[
|(\text{the second term of } (B.1))| \leq \frac{1}{a_{mn}^{(N,N)(0,0)} a_{nn}^{(N,N)(0,0)}} \lambda_{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} x_{(i,j),m}^2 \left( \sum_{l_1}^{l_2} \sum_{l_2}^{l_2} \right) \left( \sum_{m_1}^{m_1} \sum_{m_2}^{m_2} \right)
\]
\[
\leq \lambda_{N^2} \left( \frac{N \max_{1 \leq t_1, t_2 \leq N} x_{(t_1,t_2),n}^2}{a_{nn}^{(N,N)(0,0)}} \right)^{1/2} \left\{ \sum_{m_1=1}^{l_1} \left( \sum_{j=1}^{N} |\pi_{m_1,j}^2| \right)^2 \right\}^{1/2}
\]
\[
+ \sum_{l_2=1}^{l_2} \left( \sum_{j=1}^{N} |\pi_{m_2,j}^2| \right)^2 \right\}^{1/2}
\]
where \( \lambda_{N^2} \) is the greatest eigenvalue of \( \Sigma^{-1} \) and \( (\pi_{1,k}^m, \ldots, \pi_{N,k}^m)' \) (\( k = 1, 2 \)) is defined in the proof of Lemma A.1. From \( \tilde{\Sigma} = \tilde{\Sigma}_1 \otimes \tilde{\Sigma}_2 \) and the property of the Kronecker product (see Horn and Johnson 1991; page 245), the eigenvalue of \( \tilde{\Sigma} \) is \( \lambda_{1,i}\lambda_{2,j} \) \((i, j = 1, \ldots, N) \) where \( \lambda_{1,i} \) and \( \lambda_{2,j} \) are eigenvalues of \( \tilde{\Sigma}_1 \) and \( \tilde{\Sigma}_2 \) respectively. Because \( \lambda_{1,i} \) and \( \lambda_{2,j} \) are positive,
\[
\tilde{\lambda}_{N^2} = \left\{ \min_i \lambda_{1,i} \left( \min_j \lambda_{2,j} \right) \right\}^{-1} \leq \frac{1}{(\inf \lambda_1 g(\lambda_1)) (\inf \lambda_1 g(\lambda_1))} < \infty,
\]
where the first inequality is derived by Proposition 4.5.3 of Brockwell and Davis (1991). By the boundedness of \( \lambda_{N^2} \), (B.3) converges to 0 as \( N \to \infty \) from (g), Lemma A.1 and its proof. Therefore,

\[
\left( D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma(l_1, l_2) \Sigma^{-1} X D_{N^2}^{-1} \right)_{m,n} \to \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{e^{-i(l_1 \lambda_1 + l_2 \lambda_2)}}{g_1(\lambda_1)g_2(\lambda_2)} dM_{mn}(\lambda_1, \lambda_2)
\]
as \( N \to \infty \). In a similar way, we can show the three cases of \((l_1 \geq 0, l_2 < 0), (l_1 < 0, l_2 \geq 0)\) and \((l_1 < 0, l_2 < 0)\). Then,

\[
\left( D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma \Sigma^{-1} X D_{N^2}^{-1} \right)_{m,n} = \sum_{l_1=-L_1}^{L_1} \sum_{l_2=-L_2}^{L_2} b_{l_1,l_2} \left( D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma(l_1, l_2) \Sigma^{-1} X D_{N^2}^{-1} \right)_{m,n}
\]

\[
\to \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{f(\lambda_1, \lambda_2)}{g^2(\lambda_1, \lambda_2)} dM_{mn}(\lambda_1, \lambda_2)
\]
as \( N \to \infty \). Finally, consider the case of general \( h(\lambda_1, \lambda_2) \). From (f), (h) and Lemma A.2, for any sufficiently small \( \epsilon > 0 \), there exist \( h_L(\lambda_1, \lambda_2) \) and \( h_U(\lambda_1, \lambda_2) \) such that

\[
h_L(\lambda_1, \lambda_2) = \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} a_{k_1,k_2} e^{i(k_1 \lambda_1 + k_2 \lambda_2)}, \quad a_{k_1,k_2} = a_{-k_1,-k_2},
\]
\[
h_U(\lambda_1, \lambda_2) = \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} b_{k_1,k_2} e^{i(k_1 \lambda_1 + k_2 \lambda_2)}, \quad b_{k_1,k_2} = b_{-k_1,-k_2},
\]
\[
h_L(\lambda_1, \lambda_2) \leq h(\lambda_1, \lambda_2) \leq h_U(\lambda_1, \lambda_2)
\]

and

\[
h_U(\lambda_1, \lambda_2) - h_L(\lambda_1, \lambda_2) \leq \epsilon, \quad (\lambda_1, \lambda_2) \in [-\pi, \pi]^2.
\]

Then, for any \( \gamma \in \mathbb{R}^p \),

\[
\gamma' D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma \Sigma^{-1} X D_{N^2}^{-1} \gamma \leq \gamma' D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma \Sigma^{-1} X D_{N^2}^{-1} \gamma \leq \gamma' D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma \Sigma^{-1} \Sigma U \Sigma^{-1} X D_{N^2}^{-1} \gamma,
\]

where \( \Sigma_L \) and \( \Sigma_U \) are the covariance matrices with the spectral density functions \( h_L(\lambda_1, \lambda_2)g(\lambda_1, \lambda_2) \) and \( h_U(\lambda_1, \lambda_2)g(\lambda_1, \lambda_2) \) respectively. From the above discussion,

\[
\lim_{N \to \infty} \gamma' D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma \Sigma^{-1} X D_{N^2}^{-1} \gamma \leq \lim_{N \to \infty} \gamma' D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma \Sigma^{-1} X D_{N^2}^{-1} \gamma
\]

\[
\leq \gamma' D_{N^2}^{-1}X^\prime \Sigma^{-1} \Sigma \Sigma^{-1} X D_{N^2}^{-1} \gamma \leq \gamma' \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{h_U(\lambda_1, \lambda_2)}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \gamma.
\]
Since $\epsilon$ is arbitrary, as $N \to \infty$,
\[
\gamma' D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma^{-1} XD_{N^2}^{-1} \gamma \to \gamma' \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{h(\lambda_1, \lambda_2)}{g(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \gamma
\]
\[
= \gamma' \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{f(\lambda_1, \lambda_2)}{g^2(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2) \gamma.
\]
Because this holds for every vector $\gamma$, as $N \to \infty$,
\[
D_{N^2}^{-1} X' \tilde{\Sigma}^{-1} \Sigma^{-1} XD_{N^2}^{-1} \to \frac{1}{(2\pi)^2} \int_{\Pi^2} \frac{f(\lambda_1, \lambda_2)}{g^2(\lambda_1, \lambda_2)} dM(\lambda_1, \lambda_2).
\]
The proof is completed. \(\square\)

**Acknowledgments** The author is grateful to Professor Ritei Shibata, Professor Yoshiiro Yajima and Professor Yasumasa Matsuda for their encouragement and helpful comments. I also acknowledge the suggestions from the editor and two anonymous referees that refined and improved the manuscript. All results remain the author’s responsibility.

**References**

[1] Amemiya, T. (1973). Generalized least squares with an estimated autocovariance matrix. *Econometrica*, **41**, 723-732.

[2] Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

[3] Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008). Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society: Series B*, **70**, 825-848.

[4] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, second ed. Springer, New York.

[5] Bronars, S. G. and Jansen, D. W. (1987). The geographic distribution of unemployment rates in the U.S.: a spatial-time series analysis. *Journal of Econometrics*, **36**, 251-279.

[6] Cohen, G. and Francos, J. M. (2002). Linear least squares estimation of regression models for two-dimensional random fields. *Journal of Multivariate Analysis*, **82**, 431-444.

[7] Cressie, N. (1993). *Statistics for Spatial Data*, revised ed. Wiley, New York.

[8] Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B*, **70**, 209-226.
[9] Engle, R.F. (1974). Specification of the disturbance for efficient estimation. *Econometrica*, 42, 135-146.

[10] Fuentes, M. (2005). A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis*, 96, 30-54.

[11] Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15, 502-523. *Erratum and Addendum*: *Journal of Computational and Graphical Statistics*, 21, 823-824.

[12] Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics*, 18, 681-695.

[13] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

[14] Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press.

[15] Huang, H.-C., Hsu, N.-J., Theobald, D.M. and Breidt, F.J. (2010). Spatial lasso with applications to GIS model selection. *Journal of Computational and Graphical Statistics*, 19, 963-983.

[16] Ibragimov, I. A. and Rozanov, Y.A. (1978). *Gaussian Random Processes*. Springer, New York.

[17] Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press.

[18] Koreisha, S. G. and Fang, Y. (2001). Generalized least squares with misspecified serial correlation structures. *Journal of the Royal Statistical Society: Series B*, 63, 515-531.

[19] Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). *Journal of the Royal Statistical Society: Series B*, 73, 423-498.

[20] Taniguchi, M., Hirukawa, J. and Tamaki, K. (2008). *Optimal Statistical Inference in Financial Engineering*. Financial Mathematics Series. Chapman & Hall/CRC, New York.
[21] Toyooka, Y. (1985). Second-order risk comparison of SLSE with GLSE and MLE in a regression with serial correlation. *Journal of Multivariate Analysis, 17*, 107-126.

[22] Yajima, Y. (1994). On semiparametric estimation of time series analysis and its applications. *The Journal of economics, 59*(4), 2-22 (in Japanese).

[23] Yajima, Y. and Matsuda, Y. (2008). Asymptotic properties of the LSE of a spatial regression in both weakly and strongly dependent stationary random fields. Discussion Paper CIRJE-F-587, Graduate School of Economics, University of Tokyo.