

ON THE ASYMPTOTICS OF MAXIMUM LIKELIHOOD ESTIMATION FOR SPATIAL LINEAR MODELS ON A LATTICE

BY YANBING ZHENG*, AND JUN ZHU†

University of Kentucky and University of Wisconsin - Madison†*

Spatial linear models and the corresponding likelihood-based statistical inference are important tools for the analysis of spatial lattice data and have been applied in a wide range of disciplines. However, understanding of the asymptotic properties of maximum likelihood estimates is limited. Here we consider a unified asymptotic framework that encompasses increasing domain, infill, and a combination of increasing domain and infill asymptotics. Under each type of asymptotics, we derive the asymptotic properties of maximum likelihood estimates. Our results show that the rates of convergence vary for different asymptotic types and under infill asymptotics, some of the model parameters estimates are inconsistent. A simulation study is conducted to examine the finite-sample properties of the maximum likelihood estimates.

1. Introduction. In many fields of the biological, physical, and social sciences, spatial lattice data are becoming increasingly common. For example, many remotely sensed data in ecological and environmental studies are aggregated at a certain resolution on a lattice. Spatial linear models and the corresponding likelihood-based statistical inference are important tools for the analysis of such data and have been applied in a wide range of disciplines. This paper investigates the asymptotic properties of maximum likelihood estimation under a unified asymptotic framework, which encompasses increasing domain, infill, and a hybrid of increasing domain and infill asymptotics.

A spatial linear model generally has two additive components: a linear regres-

AMS 2000 subject classifications: Primary 62M30; secondary 62F12

Keywords and phrases: Autoregressive models, increasing domain asymptotics, infill asymptotics, linear regression, spatial process

sion component that relates the response variable of interest to covariates and a random error component that is modeled by a zero-mean Gaussian process. Spatial dependence in the Gaussian process can be modeled via autoregression (AR). Two classes of AR models are commonly used in practice: simultaneous autoregressive model (SAR) and conditional autoregressive model (CAR) following a neighborhood structure on the lattice. SAR models are direct generalizations of the popular autoregressive (AR) models in time series, as the random error at one site is auto-regressed on the random errors at neighboring sites on the lattice, whereas CAR models are a subclass of the Markov random fields such that the spatial dependence is induced by conditional distributions of random errors at individual sites (see, e.g., Whittle (1954), Cressie (1993); Schabenberger and Gotway (2005)). It is worth mentioning that Bandyopadhyay and Maity (2011) combined a CAR model for the random error and a semiparametric additive model for the mean response. The resulting semiparametric model allows linear and nonlinear relations between the response and covariates. In this paper, we focus on SAR models and discuss CAR models in Section 6.

SAR models have been very popular in, for example, economics and epidemiology (see, e.g., Anselin (1988); Waller and Gotway (2004)), and are becoming even more so due to the advances of software capabilities such as the R package `spdep` (see, e.g., Bivand, Pebesma and Gomez-Rubio (2008)). The SAR model we consider here is also known as an error SAR model as versus a lag SAR model where the response variable has two additive components, one is a linear regression but the other is a linear combination of the response variables at neighboring sites (Anselin (1988)). Although widely regarded as offering intuitive interpretation in spatial econometrics, we do not consider lag SAR models here, because the expected response variable in a lag SAR model is not the linear regression and it is difficult to interpret the regression coefficients as in most statistical linear models.

For statistical inference of error SAR models, maximum likelihood estimation is often adopted. However, in terms of asymptotic properties of the maximum

likelihood estimates (MLE), the results are rather limited. In a seminal paper, Mardia and Marshall (1984) established that the MLEs of the parameters are consistent and asymptotically normal as the sample size tends to infinity for general spatial linear models. The regularity conditions for the asymptotic results to hold involve continuity, growth, and convergence of the observed information matrix. These results are known to be applicable for spatial lattice models under increasing domain asymptotics, but it is unclear how the approach taken by Mardia and Marshall (1984) can be extended to deal with infill and a hybrid of increasing domain and infill asymptotics. Thus, a more general, unified framework for studying the asymptotic properties of the MLEs is critically needed for SAR models.

In contrast, more research has been conducted on infill asymptotics in geostatistics (see, e.g., Ying (1993); Lahiri (1996); Lahiri (2003); Zhang (2004); Loh (2005); Zhang and Zimmerman (2005); Du, Zhang and Mandrekar (2009)). For example, Zhang (2004) showed that, under infill, MLEs of the parameters in the Matérn class of covariance functions are inconsistent. However, Zhang (2004) considered a model without regression (i.e., with a constant mean) and did not address estimation of the mean parameter. For hybrid asymptotics, Lahiri, Lee and Cressie (2002) considered the least-squares estimators of the covariance parameters for certain spatial variogram models and established their consistency and asymptotic normality. Recently Matsuda and Yajima (2009) considered hybrid asymptotics for observations that are randomly sampled in \mathbb{R}^d and provided asymptotic results for the spectral density estimator.

Here we focus on SAR models with statistical inference based on MLE. In particular, we consider a measure of number of neighbors for any given site on the lattice. Depending on the rate of convergence of this measure, we attain different asymptotic properties of the MLEs. This idea bears similarity to Lee (2004), who studied the asymptotic properties of MLEs for a lag SAR model with the restriction to one lag only in space. Although the modeling framework in Lee (2004) is quite different and appears to be more restrictive in parameterization than

ours, we will utilize some of the techniques there when establishing the theoretical results. In addition, Robinson and Thawornkaiwong (2010) established the asymptotic normality of instrumental variables estimates of the regression coefficients for linear and semiparametric partly linear regression models and discussed the consistency of the estimates of the spatial covariance matrix, although it is not clear how general the asymptotic framework is. Under hybrid asymptotics, Robinson (2010) considered a semiparametric lag SAR model and established asymptotic properties of the adaptive estimates.

In the paper, we define increasing domain, infill, and a combination of increasing domain and infill asymptotics, which appear to have not been systematically developed for lattice models. There turns out to be an interesting and direct connection between the measure of number of neighbors and the type of asymptotics under consideration here. Our scope is broad, as we examine all three asymptotics under a unified framework and all the parameters in the spatial linear models including the regression coefficients. As we will show, the MLEs are consistent and asymptotically normal under the increasing domain and the rate of convergence is square root of the sample size. Under the infill, however, consistency and asymptotic normality are attained for the MLEs of the regression coefficients and the variance component only, while the MLEs of the autoregressive coefficients are inconsistent. When combining increasing domain and infill, consistency and asymptotic normality are attained for the MLEs of all the parameters as under increasing domain, but the rate of convergence is slower for the autoregressive coefficients.

The implications of our results are several folds. On the one hand, they reveal the difficulties in estimating autoregressive coefficients under infill. This may not be surprising, as a merely denser lattice should not be expected to provide additional information about the relations among neighboring sites. On the other hand, these results show that estimation of the regression coefficients still has desired, sound asymptotic properties. We find this particularly useful, as in many studies involving spatial lattice data, regression coefficients are of primary in-

terests and by either increasing the domain or infilling sampling sites, better estimates in terms of accuracy and precision are attainable for the regression coefficients. Furthermore, our results are instrumental from a theoretical point of view, as we strive for a comprehensive scope by examining all three types of asymptotics under a unified framework and all the parameters in the spatial linear models, including both the regression coefficients and autoregressive coefficients.

The remainder of the paper is organized as follows. In Section 2, we describe the SAR models in detail. In Section 3, we develop maximum likelihood for inference and establish three types of asymptotics. The main theoretical results are given in Section 4. A simulation study is performed in Section 5. Conclusions and discussion are given in Section 6. Technical proofs are shown in the Appendices A-C. In Appendix D, we compare our asymptotic framework and results with those in Mardia and Marshall (1984).

2. Model.

2.1. A Spatial Lattice Model. Let $\mathcal{D} \subset \mathbb{R}^d$ denote a spatial domain of interest, where $d \in \mathbb{N}$. Partitioning $\mathcal{D} = \cup_{i=1}^N \mathcal{D}_i$ into a collection of cells results in a spatial lattice $\{\mathcal{D}_i : i = 1, \dots, N\}$, where $N \in \mathbb{N}$. Let $\mathbf{s}_i \in \mathcal{D}_i$ denote a representative site in the i th cell for $i = 1, \dots, N$. The collection of sites $\{\mathbf{s}_i : i = 1, \dots, N\}$ gives an alternative way of denoting the spatial lattice. A spatial lattice may be regular or irregular. For example, an imagery in remote sensing is often a regular lattice formed by square pixels or the centroids of individual square pixels, whereas a risk map in disease mapping is often an irregular lattice formed by counties within a state or states within a country. In this paper, we deal with both regular and irregular spatial lattices.

Let (Ω, \mathcal{F}, P) denote a probability space. For modeling a given response variable on the spatial lattice, let $\{Y(\mathbf{s}_i) : \mathbf{s}_i \in \mathcal{D}_i, i = 1, \dots, N\}$ denote a random spatial process consisting of $Y(\cdot)$ defined on (Ω, \mathcal{F}, P) . For ease of notation, let $Y_i \equiv Y(\mathbf{s}_i)$ for $i = 1, \dots, N$. Under a suitable asymptotic framework to be specified in Section 3.2, we let n index the stage of the asymptotics for $n \in \mathbb{N}$. Thus

let $\mathcal{D} \equiv \mathcal{D}_n$, $N \equiv N_n$, and $\mathcal{D}_i \equiv \mathcal{D}_{n,i}$ for $i = 1, \dots, N_n$ such that $\mathcal{D}_n = \cup_{i=1}^{N_n} \mathcal{D}_{n,i}$ for $n \in \mathbb{N}$.

Now, let $\mathbf{Y}_n = (Y_1, \dots, Y_{N_n})'$ denote the vector of response variables on the lattice $\{\mathcal{D}_{n,i} : i = 1, \dots, N_n\}$. We consider a spatial linear model in the form of

$$(2.1) \quad \mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \boldsymbol{\epsilon}_n$$

where \mathbf{X}_n is an $N_n \times p$ design matrix and $\boldsymbol{\beta}$ is a p -dimensional vector of regression coefficients. Furthermore, $\boldsymbol{\epsilon}_n$ is an N_n -dimensional vector of random errors such that

$$(2.2) \quad \boldsymbol{\epsilon}_n = \mathbf{W}_n(\boldsymbol{\theta}) \boldsymbol{\epsilon}_n + \boldsymbol{\nu}_n$$

where $\mathbf{W}_n(\boldsymbol{\theta})$ is an $N_n \times N_n$ spatial weight matrix parameterized by a q -dimensional vector $\boldsymbol{\theta}$ such that the diagonal elements of $\mathbf{W}_n(\boldsymbol{\theta})$ are 0's and $\mathbf{I}_n - \mathbf{W}_n(\boldsymbol{\theta})$ is non-singular, where \mathbf{I}_n is an $N_n \times N_n$ identity matrix. Further, $\boldsymbol{\nu}_n$ is an N_n -dimensional vector of disturbances that follow $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. In model (2.2), the random error at a given site is auto-regressed on those at other sites on the spatial lattice, which induces spatial dependence.

2.2. Model Parameterization. The spatial weight matrix $\mathbf{W}_n(\boldsymbol{\theta})$ can be specified in different ways, but usually involves specification of a neighborhood structure. We let $\mathcal{N}_n(i) = \{j : \text{site } j \text{ is a neighbor of site } i\}$ denote the neighborhood of site i . The neighborhood of site i can be further partitioned into q orders, such that $\mathcal{N}_n(i) = \cup_{k=1}^q \mathcal{N}_{n,k}(i)$ where $\mathcal{N}_{n,k}(i) = \{j : \text{site } j \text{ is a } k\text{th order neighbor of site } i\}$. We then consider the following parameterization of the spatial weights matrix $\mathbf{W}_n(\boldsymbol{\theta})$:

$$(2.3) \quad \mathbf{W}_n(\boldsymbol{\theta}) = \sum_{k=1}^q \theta_k \mathbf{W}_{n,k}$$

where the diagonal elements of $\mathbf{W}_{n,k}$ are 0's for all k and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$ is a q -dimensional vector of autoregressive coefficients such that $\mathbf{I}_n - \sum_{k=1}^q \theta_k \mathbf{W}_{n,k}$ is nonsingular. For row standardized weight matrices, if $\sum_{k=1}^q |\theta_k| < 1$, then $\mathbf{I}_n - \sum_{k=1}^q \theta_k \mathbf{W}_{n,k}$ is nonsingular and thus the covariance matrix in the SAR

model is positive definite (Corollary 5.6.16, Horn and Johnson (1985)). We let $\boldsymbol{\theta} \in \Theta$, where Θ is a compact subset of \mathbb{R}^q . The parameterization of spatial weight matrix (2.3) is flexible, as it can accommodate different orders of neighborhoods to be associated with different autoregressive coefficients (Zhu, Huang and Reyes (2010)).

3. Maximum Likelihood Estimation.

3.1. *Likelihood Function.* Let $\boldsymbol{\eta} = (\boldsymbol{\beta}', \boldsymbol{\theta}', \sigma^2)'$ denote a $(p+q+1)$ -dimensional vector of parameters under the model specified in (2.1)–(2.3). Let $\mathbf{S}_n(\boldsymbol{\theta}) = \mathbf{I}_n - \sum_{k=1}^q \theta_k \mathbf{W}_{n,k}$ for $\boldsymbol{\theta} \in \Theta$. The log-likelihood function, up to a constant, is

$$(3.1) \quad \ell(\boldsymbol{\eta}) = -(N_n/2) \log \sigma^2 + \log |\mathbf{S}_n(\boldsymbol{\theta})| - (2\sigma^2)^{-1} \boldsymbol{\nu}_n' \boldsymbol{\nu}_n,$$

where $\boldsymbol{\nu}_n = \mathbf{S}_n(\boldsymbol{\theta})(\mathbf{Y}_n - \mathbf{X}_n \boldsymbol{\beta})$. The first-order derivatives of $\ell(\boldsymbol{\eta})$ with respect to $\boldsymbol{\beta}$ and σ^2 are, respectively,

$$\frac{\partial \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\beta}} = (\sigma^2)^{-1} \mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \boldsymbol{\nu}_n, \quad \frac{\partial \ell(\boldsymbol{\eta})}{\partial \sigma^2} = (2\sigma^4)^{-1} (\boldsymbol{\nu}_n' \boldsymbol{\nu}_n - N_n \sigma^2).$$

Thus, written in terms of $\boldsymbol{\theta}$, the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ and σ^2 are, respectively,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_n(\boldsymbol{\theta}) &= \{\mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{X}_n\}^{-1} \mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{Y}_n, \\ \hat{\sigma}_n^2(\boldsymbol{\theta}) &= N_n^{-1} \{\mathbf{Y}_n - \mathbf{X}_n \hat{\boldsymbol{\beta}}_n(\boldsymbol{\theta})\}' \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \{\mathbf{Y}_n - \mathbf{X}_n \hat{\boldsymbol{\beta}}_n(\boldsymbol{\theta})\}. \end{aligned}$$

We define a profile log-likelihood function of $\boldsymbol{\theta}$ as

$$\ell(\boldsymbol{\theta}) = \ell\{\hat{\boldsymbol{\beta}}_n(\boldsymbol{\theta}), \boldsymbol{\theta}, \hat{\sigma}_n^2(\boldsymbol{\theta})\} = -(N_n/2) \log \hat{\sigma}_n^2(\boldsymbol{\theta}) + \log |\mathbf{S}_n(\boldsymbol{\theta})| - N_n/2.$$

Then the MLE of $\boldsymbol{\theta}$ maximizes the profile log-likelihood $\ell(\boldsymbol{\theta})$ and is denoted as $\hat{\boldsymbol{\theta}}_n$.

3.2. *Asymptotic Types.* For any set $A \subset \mathbb{R}^d$, let $\text{vol}(A)$ denote the volume (i.e., the Lebesgue measure) of A . We define three asymptotic types in terms of the volume of the spatial domain $\text{vol}(\mathcal{D}_n)$ and that of the individual cells $\text{vol}(\mathcal{D}_{n,i})$ as follows.

- *Increasing domain asymptotics*: The volume of the spatial lattice tends to infinity $\text{vol}(\mathcal{D}_n) \rightarrow \infty$ as $n \rightarrow \infty$, while the volume of each cell on the lattice is fixed $\text{vol}(\mathcal{D}_{n,i}) \equiv \text{vol}(\mathcal{D}_{1,i})$ for $i = 1, \dots, N_n, n = 2, 3, \dots$
- *Infill asymptotics*: The volume of the spatial lattice is fixed $\text{vol}(\mathcal{D}_n) \equiv \text{vol}(\mathcal{D}_1)$ for $n = 2, 3, \dots$, while the volume of each cell on the lattice tends to zero $\max\{\text{vol}(\mathcal{D}_{n,i}) : i = 1, \dots, N_n\} \rightarrow 0$ as $n \rightarrow \infty$.
- *Hybrid asymptotics (increasing domain combined with infill asymptotics)*: The volume of the spatial lattice tends to infinity $\text{vol}(\mathcal{D}_n) \rightarrow \infty$ and the volume of each cell on the lattice tends to zero $\max\{\text{vol}(\mathcal{D}_{n,i}) : i = 1, \dots, N_n\} \rightarrow 0$ as $n \rightarrow \infty$.

In the context of (2.3), an essential element in the specification of an asymptotic type is the order of the elements in the spatial weight matrix, denoted as m_n^{-1} . That is, $w_{n,k}^{i,j} = \mathcal{O}(m_n^{-1})$, where $\{w_{n,k}^{i,j}\}$ are elements of the weight matrix $\mathbf{W}_{n,k}$. Consider the following examples on a regular spatial lattice.

EXAMPLE 1. *Nearest neighbors*: Let $v_{n,k}^{i,j} = \mathcal{I}\{\text{site } j \text{ is the } k\text{th nearest neighbor of site } i\}$, where $\mathcal{I}(\cdot)$ is an indicator function, and $\mathbf{V}_{n,k} = [v_{n,k}^{i,j}]_{i,j=1}^{N_n}$. Row standardize $\mathbf{V}_{n,k}$ to attain a spatial weight matrix $\mathbf{W}_{n,k}$ such that $w_{n,k}^{i,j} = v_{n,k}^{i,j} / \sum_{j=1}^{N_n} v_{n,k}^{i,j}$.

EXAMPLE 2. *Distance-based neighbors*: Let d_{ij} denote the Euclidean distance between sites i and j . Let $v_{n,k}^{i,j} = \mathcal{I}\{d_{ij} \in [\delta_{k-1}, \delta_k)\}$, where $\delta_0 = 0 < \delta_1 < \dots < \delta_q$ are prespecified threshold values, and $\mathbf{V}_{n,k} = [v_{n,k}^{i,j}]_{i,j=1}^{N_n}$. Row standardize $\mathbf{V}_{n,k}$ to attain a spatial weight matrix $\mathbf{W}_{n,k}$ such that $w_{n,k}^{i,j} = v_{n,k}^{i,j} / \sum_{j=1}^{N_n} v_{n,k}^{i,j}$, as in Example 1.

In Example 1, $m_n = \mathcal{O}(\max\{\sum_{j=1}^{N_n} v_{n,k}^{i,j} : k = 1, \dots, q, i = 1, \dots, N_n\})$ is $\mathcal{O}(1)$ by the definition of the neighborhood. This order is the same for all n and all three asymptotic types. For example, consider $q = 1$ and thus the four nearest neighbors on a regular lattice. The infill asymptotics can be thought of as the reverse of the increasing domain asymptotics, even though the spatial domain is fixed and finite. For this reason, it suffices to consider only the increasing domain for Example 1. In Example 2, $m_n = \mathcal{O}(\max\{\sum_{j=1}^{N_n} v_{n,k}^{i,j} : k = 1, \dots, q, i = 1, \dots, N_n\})$. Under the

increasing domain asymptotics, m_n is $\mathcal{O}(1)$. However, under the infill asymptotics, $m_n \rightarrow \infty$ and m_n/N_n does not tend to 0, whereas under the hybrid asymptotics, $m_n \rightarrow \infty$ and $m_n/N_n \rightarrow 0$. Here, q is assumed to be fixed. This is a reasonable assumption even for the hybrid and infill asymptotics, as in Example 2, for a given site, the number of neighbors for each order of neighborhood structure tends to infinity even though q remains fixed.

For infill asymptotics and hybrid asymptotics, the autoregressive coefficients θ_k , $k = 1, \dots, q$ are assumed to be fixed for different cell resolutions. This assumption is reasonable under row standardization of the weight matrices. In Example 2, it means that the average of the neighboring cells in a neighborhood of a particular order contributes the same across resolutions.

4. Asymptotic Properties.

4.1. *Notation.* Let $\boldsymbol{\eta}_0 = (\beta'_0, \theta'_0, \sigma_0^2)'$ denote the $(p+q+1)$ -dimensional vector of true model parameters. Let $\mathbf{S}_{0n} \equiv \mathbf{S}_n(\boldsymbol{\theta}_0)$. The model (2.1)–(2.2) evaluated at the true parameters $\boldsymbol{\eta}_0$ is $\mathbf{Y}_n = \mathbf{X}_n \beta_0 + \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n}$, where $\boldsymbol{\nu}_{0n} \sim N(\mathbf{0}, \sigma_0^2 \mathbf{I}_n)$. Let $\hat{\beta}_n = \hat{\beta}_n(\hat{\boldsymbol{\theta}}_n)$, $\hat{\sigma}_n^2 = \hat{\sigma}_n^2(\hat{\boldsymbol{\theta}}_n)$, and $\hat{\boldsymbol{\eta}}_n = (\hat{\beta}'_n, \hat{\theta}'_n, \hat{\sigma}_n^2)'$ denote the MLE of $\boldsymbol{\eta}$. For any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, let $\sigma_n^{*2}(\boldsymbol{\theta}) = N_n^{-1} \sigma_0^2 \text{tr} \left\{ \mathbf{S}_{0n}^{-1} \mathbf{S}'_n(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{S}_{0n}^{-1} \right\}$.

Let \mathbf{A}_n denote an $N_n \times N_n$ matrix with elements $[a_n^{i,j}]_{i,j=1}^{N_n}$. The sequence of matrices \mathbf{A}_n is uniformly bounded in matrix norm $\|\cdot\|_\infty$, if $\sup_{1 \leq i \leq N_n, n \geq 1} \sum_{j=1}^{N_n} |a_n^{i,j}| < \infty$. The sequence of matrices \mathbf{A}_n is uniformly bounded in matrix norm $\|\cdot\|_1$, if $\sup_{1 \leq j \leq N_n, n \geq 1} \sum_{i=1}^{N_n} |a_n^{i,j}| < \infty$. Furthermore, the sequence of matrices \mathbf{A}_n is uniformly bounded in ℓ_∞ norm if $\sup_{1 \leq i, j \leq N_n, n \geq 1} |a_n^{i,j}| < \infty$ (Horn and Johnson (1985)).

Let Γ denote a compact subset of \mathbb{R}^m , $m \in \mathbb{N}$, and let Γ_n denote a non-empty compact subset of Γ , $n = 1, 2, \dots$. Let $g_n : \Gamma \rightarrow \bar{\mathbb{R}}$ denote a continuous function on Γ . Suppose that $g_n(\boldsymbol{\gamma})$ has a maximum on Γ_n at $\boldsymbol{\gamma}_n^*$, $n = 1, 2, \dots$. Let $S_n(\delta)$ denote an open ball in \mathbb{R}^m centered at $\boldsymbol{\gamma}_n^*$ with fixed radius $\delta > 0$. For each $n = 1, 2, \dots$, define a neighborhood $\Gamma_n(\delta) = S_n(\delta) \cap \Gamma_n$ with compact complement $\Gamma_n^c(\delta)$ in Γ_n . The sequence of maximizers $\{\boldsymbol{\gamma}_n^*\}$ is said to be identifiably unique

on $\{\Gamma_n\}$, if either for all $\delta > 0$ and all n , $\Gamma_n^c(\delta)$ is empty, or for all $\delta > 0$ $\lim_{n \rightarrow \infty} \sup \left\{ \max_{\gamma \in \Gamma_n^c(\delta)} g_n(\gamma) - g_n(\gamma_n^*) \right\} < 0$ (White (1994)).

4.2. *Assumptions.* To establish the asymptotic properties of the MLEs of the model parameters, we impose the following regularity conditions.

- (A.1) The elements $w_{n,k}^{i,j}$ of the spatial weight matrix $\mathbf{W}_{n,k}$ are at most of order m_n^{-1} uniformly for all $j \neq i$ and $w_{n,k}^{i,i} = 0$, for all $i = 1, \dots, N_n$ and $k = 1, \dots, q$ and m_n is bounded away from zero uniformly.
- (A.2) The sequence of spatial weight matrices $\{\mathbf{W}_{n,k} : k = 1, \dots, q\}$ are uniformly bounded in matrix norms $||| \cdot |||_1$ and $||| \cdot |||_\infty$.
- (A.3) The matrix $\mathbf{S}_n(\boldsymbol{\theta})$ is nonsingular for $\boldsymbol{\theta} \in \Theta$ and $n \in \mathbb{N}$.
- (A.4) The sequence of matrices $\{\mathbf{S}_n^{-1}(\boldsymbol{\theta})\}$ is uniformly bounded in matrix norms $||| \cdot |||_1$ and $||| \cdot |||_\infty$ for $\boldsymbol{\theta} \in \Theta$. The true parameter $\boldsymbol{\theta}_0$ is in the interior of Θ .
- (A.5) The elements of \mathbf{X}_n are uniformly bounded constants. The limit of $N_n^{-1} \mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{X}_n$ as $n \rightarrow \infty$ exists and is nonsingular for $\boldsymbol{\theta} \in \Theta$.
- (A.6) For $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\lim_{n \rightarrow \infty} N_n^{-1} \{ \log |\sigma_n^{*2}(\boldsymbol{\theta}) \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta})| - \log |\sigma_0^2 \mathbf{S}_{0n}' \mathbf{S}_{0n}| \} \neq 0$.
- (A.6') For $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\lim_{n \rightarrow \infty} N_n^{-1} m_n \{ \log |\sigma_n^{*2}(\boldsymbol{\theta}) \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta})| - \log |\sigma_0^2 \mathbf{S}_{0n}' \mathbf{S}_{0n}| \} \neq 0$.

Assumptions (A.1) and (A.2) are regularity conditions on the spatial weight matrices. Note that, in (A.1), the sequence of rates $\{m_n : n = 1, 2, \dots\}$ can be bounded or divergent in general. Assumption (A.2) is generally satisfied under row standardization. Indeed, by Lemma 6, it can be shown that (A.1) implies (A.2). Assumptions (A.3) and (A.4) are made about the sequence of matrices $\mathbf{S}_n(\boldsymbol{\theta})$. Assumption (A.3) is standard for the spatial lattice model under consideration. Assumption (A.4) is needed to ensure that the variance of \mathbf{Y}_n is bounded. In fact, one can show that (A.4) holds under fairly weak conditions by applying Lemma 7. Assumption (A.5) is a standard assumption of the design matrix and implies that the elements of $N_n \{ \mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{X}_n \}^{-1}$ are uniformly bounded. Assumption (A.6) (or (A.6')) is needed to establish identifiable uniqueness when establishing consistency of the MLE.

4.3. Results.

THEOREM 1. Assume that (A.1)–(A.6) hold and $m_n = \mathcal{O}(1)$. Then the MLE of $\boldsymbol{\eta}$ is consistent such that, as $n \rightarrow \infty$,

$$\hat{\boldsymbol{\eta}}_n \xrightarrow{P} \boldsymbol{\eta}_0.$$

If, in addition, the limit of $N_n^{-1} E \left\{ -\frac{\partial^2 \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right\}$ as $n \rightarrow \infty$ exists and is positive definite for $\boldsymbol{\eta} \in \mathbb{R}^p \times \Theta \times \mathbb{R}^+$, then the MLE of $\boldsymbol{\eta}$ is asymptotically normal such that, as $n \rightarrow \infty$,

$$N_n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_0})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_0}^{-1} = \lim_{n \rightarrow \infty} -N_n^{-1} E \left\{ \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right\}$.

Theorem 1 shows that, under the regularity conditions and when $m_n = \mathcal{O}(1)$, the MLE $\hat{\boldsymbol{\eta}}_n$ is consistent and asymptotically normal at a convergence rate of square root of the sample size N_n . The condition $m_n = \mathcal{O}(1)$ corresponds to increasing domain asymptotics in Examples 1–2.

THEOREM 2. Assume (A.1)–(A.5) and (A.6') hold, $m_n \rightarrow \infty$ and $m_n/N_n \rightarrow 0$, as $n \rightarrow \infty$. Then the MLE of $\boldsymbol{\eta}$ is consistent such that, as $n \rightarrow \infty$,

$$\hat{\boldsymbol{\eta}}_n \xrightarrow{P} \boldsymbol{\eta}_0.$$

If, in addition, $\lim_{n \rightarrow \infty} N_n^{-1} m_n^{1+\delta} = 0$ for some $\delta > 0$ and if the limit of $(m_n/N_n) E \left\{ -\frac{\partial^2 \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\}$ as $n \rightarrow \infty$ exists and is positive definite for $\boldsymbol{\theta} \in \Theta$, then the MLE of $\boldsymbol{\eta}$ is asymptotically normal such that, as $n \rightarrow \infty$,

$$\begin{aligned} N_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) &\xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}), & N_n^{1/2}(\hat{\sigma}_n^2 - \sigma_0^2) &\xrightarrow{D} N(0, 2\sigma_0^4) \\ (N_n/m_n)^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) &\xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}), \end{aligned}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_0} = \sigma_0^2 \lim_{n \rightarrow \infty} N_n (\mathbf{X}'_n \mathbf{S}'_{0n} \mathbf{S}_{0n} \mathbf{X}_n)^{-1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_0}^{-1} = \lim_{n \rightarrow \infty} -(m_n/N_n) E \left\{ \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\}$. Furthermore, $\hat{\boldsymbol{\beta}}_n$ is asymptotically independent of $\hat{\sigma}_n^2$ and $\hat{\boldsymbol{\theta}}_n$.

Theorem 2 shows that, under the regularity conditions and when $m_n \rightarrow \infty$ and $m_n/N_n \rightarrow 0$, the MLE $\hat{\boldsymbol{\eta}}_n$ is consistent and asymptotically normal. The

conditions $m_n \rightarrow \infty$ and $m_n/N_n \rightarrow 0$ correspond to the hybrid asymptotics in Example 2. The rate of convergence is the square root of the sample size N_n for the regression coefficients β and the variance component σ^2 . However, the rate of convergence is slower at $\sqrt{N_n/m_n}$ for the autoregressive coefficients θ . It appears that only the increasing domain asymptotics improve the inference for θ , whereas the infill asymptotics do not yield additional information.

THEOREM 3. *Assume (A.1)–(A.5) hold, $m_n \rightarrow \infty$ and $m_n/N_n \rightarrow c \in (0, \infty]$ as $n \rightarrow \infty$. Then the MLEs of β and σ^2 are consistent such that, as $n \rightarrow \infty$,*

$$\hat{\beta}_n(\theta) \xrightarrow{P} \beta_0 \quad \text{and} \quad \hat{\sigma}_n^2(\theta) \xrightarrow{P} \sigma_0^2$$

and asymptotically normal such that, as $n \rightarrow \infty$,

$$N_n^{1/2}\{\hat{\beta}_n(\theta) - \beta_0\} \xrightarrow{D} N(0, \Sigma_\beta), \quad N_n^{1/2}\{\hat{\sigma}_n^2(\theta) - \sigma_0^2\} \xrightarrow{D} N(0, 2\sigma_0^4)$$

where $\Sigma_\beta = \sigma_0^2 \lim_{n \rightarrow \infty} N_n \{X_n' S_n(\theta)' S_n(\theta) X_n\}^{-1} X_n' S_n'(\theta) S_n(\theta) S_{0n}^{-1} S_{0n}'^{-1} S_n'(\theta) S_n(\theta) X_n \{X_n' S_n'(\theta) S_n(\theta) X_n\}^{-1}$, for a given $\theta \in \Theta$. Furthermore, $\hat{\beta}_n(\theta)$ and $\hat{\sigma}_n^2(\theta)$ are asymptotically independent. However, the consistency of the MLE of θ is not guaranteed.

Theorem 3 shows that, under the regularity conditions and when $m_n \rightarrow \infty$ and $m_n/N_n \rightarrow c > 0$, the MLEs $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ are consistent and asymptotically normal at a rate of $\sqrt{N_n}$. However, the MLE $\hat{\theta}_n$ is inconsistent. The conditions $m_n \rightarrow \infty$ and $m_n/N_n \rightarrow c > 0$ correspond to the infill asymptotics in Example 2. In other words, without increasing the domain, it is not possible to infer about θ consistently using MLE. This makes sense intuitively, as infill alone does not yield additional information about the relations among neighboring sites.

5. Simulation Study. We now conduct a simulation study to examine the finite-sample properties of the MLEs under three types of asymptotics. We consider an $r \times r$ square lattice with a unit resolution. We vary the lattice size by letting $r = 4, 8$, or 16 , while keeping the area of each cell at 1 unit². For each lattice size, we further divide each cell into an $r^* \times r^*$ sub-lattice with sub-cells.

We then vary the sub-lattice size by letting $r^* = 1, 2$, or 4 . The sample size N thus ranges from 16 ($r = 4, r^* = 1$) to 4096 ($r = 16, r^* = 4$).

For a given lattice size r and sub-lattice size r^* , we simulate data from a SAR model defined in (2.1) and (2.2). For the linear regression, we let $E(Y_i) = \beta_0 + \beta_1 X_i$, where $X_i = \sin(i)$, $\beta_0 = 2$, and $\beta_1 = 2$ for the i th cell, $i = 1, \dots, N$. For the spatial dependence, we consider distance-based neighborhood with one order $q = 1$ and a unit threshold distance $\delta_1 = 1$. The parameter values are set at $\theta_1 = 0.2$ and $\sigma^2 = 1$. For each simulated data, we estimate the model parameters by maximum likelihood and obtain $\hat{\beta}_0, \hat{\beta}_1, \hat{\theta}_1$, and $\hat{\sigma}^2$. We repeat this procedure 100 times. Tables 1–3 give the means and the standard deviations of the MLEs.

First, for a given sub-lattice size r^* , we examine the results across different lattice sizes r , which corresponds to increasing domain asymptotics. We note that the biases and standard deviations of all four parameter estimates decrease as the lattice size r increases from 4 to 16 , for any given sub-lattice size $r^* = 1, 2$, or 4 . Next, for a given lattice size r , we examine the results over different sub-lattice sizes r^* , which corresponds to infill asymptotics. The biases and standard deviations of the regression coefficient estimates $\hat{\beta}_0, \hat{\beta}_1$ and the variance component estimate $\hat{\sigma}^2$ decrease as the sub-lattice size r^* increases from 1 to 4 , for any given lattice size $r = 4, 8$, or 16 . However, for the autoregressive coefficient estimate $\hat{\theta}_1$, the biases and standard deviations remain similar in values as r^* increases from 1 to 4 , indicating inconsistency. Last, we examine the results when both the lattice size r and the sub-lattice size r^* increase, which correspond to the hybrid asymptotics. We note that the biases and standard deviations of all four parameters decrease. However, the biases and standard deviations of $\hat{\theta}_1$ decrease at a slower rate in comparison to the results of the other three parameter estimates. In summary, under all three types of asymptotics, the finite-sample properties support the theoretical results given in Theorems 1–3.

6. Conclusions and Discussion. The modeling approach in SAR is by construction in the sense that spatial dependence is induced from autoregression in formulas like (2.2). This is a technique widely used in other areas of statis-

tics such as time series, dynamic modeling, and graphical models. The resulting SAR model has a rational spectral density under suitable conditions (Gaetan and Guyon (2010)). It is also intuitive, as one can interpret the model parameters via equation (2.2) and thus SAR model has been one of the most popular models used for spatial lattice data. However, we believe that there lacks an adequate framework for studying the asymptotic properties of the parameter estimates for SAR, despite its immense popularity in practical use in many scientific disciplines. What we have achieved here is an attempt to bridge some of this gap between theory and practice.

The asymptotic framework considered here is unified covering all three types of asymptotics. This is in contrast to Mardia and Marshall (1984), which is often referenced as the theoretical backing of SAR model parameter estimation (see, e.g., Cressie (1993)). It can be shown that if the assumptions of our Theorem 1 or 2 hold, then the assumptions for Mardia and Marshall (1984)'s results are satisfied. The details are given in Appendix D. However, the asymptotic results in Mardia and Marshall (1984) do not distinguish the rates of convergence as we do for increasing domain in Theorem 1 and for hybrid asymptotics in Theorem 2. Furthermore, the approach taken in Mardia and Marshall (1984) cannot deal with infill asymptotics in a straightforward manner. Thus, we believe that the approach taken here offers a viable alternative for studying the asymptotics of SAR models.

As pointed out by Cressie (1993), any SAR model can be represented as a CAR model, which is a Gaussian Markov random field model. Research has been conducted to study the link between geostatistical Gaussian random fields and Gaussian Markov random fields. Rue and Tjelmeland (2002) demonstrated empirically that Gaussian Markov random fields approximate well commonly used Gaussian fields. More recently, Lindgren, Lindström and Rue (2010) showed that for some Gaussian fields with the Matérn class of covariance functions, an explicit link can be established to Gaussian Markov random fields using an approximate stochastic weak solution to stochastic partial differential equations. Taken to-

gether, we conjecture that SAR models approximate a class of Gaussian random fields, which would then provide a closer connection between asymptotic frameworks considered here and those in geostatistics. We leave this for future research.

Appendix A: Proof of Theorem 1.

PROOF. From (3.1), it can be shown that, under $\boldsymbol{\eta}_0$,

$$\begin{aligned} E\{\ell(\boldsymbol{\eta})\} &= -(N_n/2) \log \sigma^2 + \log |\mathbf{S}_n(\boldsymbol{\theta})| \\ &\quad - (2\sigma^2)^{-1} \left[(\beta_0 - \beta)' \mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{X}_n (\beta_0 - \beta) + \sigma_0^2 \text{tr} \left\{ \mathbf{S}_{0n}'^{-1} \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{S}_{0n}^{-1} \right\} \right] \end{aligned}$$

Thus the first-order derivatives of $E\{\ell(\boldsymbol{\eta})\}$ with respect to $\boldsymbol{\beta}$ and σ^2 are, respectively,

$$\frac{\partial E\{\ell(\boldsymbol{\eta})\}}{\partial \boldsymbol{\beta}} = (\sigma^2)^{-1} \mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{X}_n (\beta_0 - \boldsymbol{\beta}), \quad \frac{\partial E\{\ell(\boldsymbol{\eta})\}}{\partial \sigma^2} = (2\sigma^4)^{-1} \{E(\boldsymbol{\nu}_n' \boldsymbol{\nu}_n) - N_n \sigma^2\}.$$

The maximizers of $E\{\ell(\boldsymbol{\eta})\}$ are $\boldsymbol{\beta}_n^*(\boldsymbol{\theta}) = \beta_0$ and $\sigma_n^{*2}(\boldsymbol{\theta}) = N_n^{-1} \sigma_0^2 \text{tr} \left\{ \mathbf{S}_{0n}'^{-1} \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{S}_{0n}^{-1} \right\}$.

Let $g_n(\boldsymbol{\theta}) = E[\ell\{\boldsymbol{\beta}_n^*(\boldsymbol{\theta}), \boldsymbol{\theta}, \sigma_n^{*2}(\boldsymbol{\theta})\}]$. Thus $g_n(\boldsymbol{\theta}) = -(N_n/2) \log \sigma_n^{*2}(\boldsymbol{\theta}) + \log |\mathbf{S}_n(\boldsymbol{\theta})| - N_n/2$. We establish the consistency of $\hat{\boldsymbol{\theta}}_n$ by showing that $\sup_{\boldsymbol{\theta} \in \Theta} N_n^{-1} |\ell(\boldsymbol{\theta}) - g_n(\boldsymbol{\theta})| = o_p(1)$ and that $N_n^{-1} g_n(\boldsymbol{\theta})$ is identifiably unique. We then establish the consistency of $\hat{\boldsymbol{\beta}}_n$ and $\hat{\sigma}_n^2$ in terms of $\hat{\boldsymbol{\theta}}_n$.

Note that,

$$N_n^{-1} \{\ell(\boldsymbol{\theta}) - g_n(\boldsymbol{\theta})\} = -1/2 \{\log \hat{\sigma}_n^2(\boldsymbol{\theta}) - \log \sigma_n^{*2}(\boldsymbol{\theta})\} = -\{2\tilde{\sigma}_n^2(\boldsymbol{\theta})\}^{-1} \{\hat{\sigma}_n^2(\boldsymbol{\theta}) - \sigma_n^{*2}(\boldsymbol{\theta})\}$$

where $\tilde{\sigma}_n^2(\boldsymbol{\theta}) = \alpha \sigma_n^{*2}(\boldsymbol{\theta}) + (1 - \alpha) \hat{\sigma}_n^2(\boldsymbol{\theta})$ for some $\alpha \in (0, 1)$ and $\hat{\sigma}_n^2(\boldsymbol{\theta}) = N_n^{-1} \boldsymbol{\nu}_{0n}' \mathbf{B}_n(\boldsymbol{\theta}) \boldsymbol{\nu}_{0n}$, with

$$\begin{aligned} \mathbf{B}_n(\boldsymbol{\theta}) &= \mathbf{S}_{0n}'^{-1} \mathbf{S}_n'(\boldsymbol{\theta}) \left[\mathbf{I}_n - \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{X}_n \{ \mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{X}_n \}^{-1} \mathbf{X}_n' \mathbf{S}_n'(\boldsymbol{\theta}) \right] \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{S}_{0n}^{-1}. \\ (6.1) \end{aligned}$$

Also, by (A.1)–(A.4) and Lemma 5,

$$N_n^{-1} \left[\boldsymbol{\nu}_{0n}' \mathbf{S}_{0n}'^{-1} \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} - E \left\{ \boldsymbol{\nu}_{0n}' \mathbf{S}_{0n}'^{-1} \mathbf{S}_n'(\boldsymbol{\theta}) \mathbf{S}_n(\boldsymbol{\theta}) \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} \right\} \right] = o_p(1).$$

Furthermore, the convergence is uniform on Θ , because of the linear-quadratic form in θ and by Corollary 2.2 of Newey (1991). By (A.1)–(A.5) and Lemma 4,

$$\begin{aligned} & N_n^{-1} \nu'_{0n} S_{0n}'^{-1} S_n'(\theta) S_n(\theta) X_n \{X_n' S_n'(\theta) S_n(\theta) X_n\}^{-1} X_n' S_n'(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \\ &= N_n^{-1} \left\{ N_n^{-1/2} X_n' S_n'(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \right\}' \left\{ N_n^{-1} X_n' S_n'(\theta) S_n(\theta) X_n \right\}^{-1} \\ & \quad \times \left\{ N_n^{-1/2} X_n' S_n'(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \right\} = o_p(1) \end{aligned}$$

Again, the convergence is uniform on Θ , since $N_n^{-1} X_n' S_n'(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} = o_p(1)$ uniformly on Θ , $\{N_n^{-1} X_n' S_n'(\theta) S_n(\theta) X_n\}^{-1}$ is uniformly bounded in ℓ_∞ , and the boundedness is uniform on Θ by (A.5). Thus it follows that, uniformly on Θ ,

$$\begin{aligned} & \hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta) \\ &= N_n^{-1} \left[\nu'_{0n} S_{0n}'^{-1} S_n'(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} - E \left\{ \nu'_{0n} S_{0n}'^{-1} S_n'(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \right\} \right] \\ & \quad - N_n^{-1} \nu'_{0n} S_{0n}'^{-1} S_n'(\theta) S_n(\theta) X_n \{X_n' S_n'(\theta) S_n(\theta) X_n\}^{-1} X_n' S_n'(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} = o_p(1). \end{aligned} \tag{6.2}$$

Further, by Jensen's inequality,

$$\begin{aligned} N_n^{-1} \{g_n(\theta) - g_n(\theta_0)\} &= N_n^{-1} (\log |S_n(\theta)| - \log |S_{0n}|) - 1/2 \{\log \sigma_n^{*2}(\theta) - \log \sigma_0^2\} \\ &= (1/2) \log \frac{|S_{0n}'^{-1} S_n'(\theta) S_n(\theta) S_{0n}^{-1}|^{N_n^{-1}}}{N_n^{-1} \text{tr}\{S_{0n}'^{-1} S_n'(\theta) S_n(\theta) S_{0n}^{-1}\}} \leq 0 \end{aligned} \tag{6.3}$$

for $\theta \in \Theta$. Under (A.1)–(A.4),

$$\begin{aligned} N_n^{-1} (\log |S_n(\theta)| - \log |S_{0n}|) &= -N_n^{-1} (\text{tr}\{S_n(\tilde{\theta})^{-1} W_{n,1}\}, \dots, \text{tr}\{S_n(\tilde{\theta})^{-1} W_{n,q}\})(\theta - \theta_0) \\ &= -\sum_{k=1}^q \mathcal{O}(m_n^{-1})(\theta_k - \theta_{0,k}), \end{aligned} \tag{6.4}$$

where $\tilde{\theta} = \alpha\theta + (1 - \alpha)\theta_0$ for some $\alpha \in (0, 1)$. Thus

$$\begin{aligned} \log \sigma_n^{*2}(\theta) &= -2N_n^{-1} \{g_n(\theta) - g_n(\theta_0)\} + 2N_n^{-1} \{\log |S_n(\theta)| - \log |S_{0n}|\} + \log \sigma_0^2 \\ &\geq 2N_n^{-1} \{\log |S_n(\theta)| - \log |S_{0n}|\} + \log \sigma_0^2, \end{aligned}$$

which is bounded from below uniformly on Θ , and $\sigma_n^{*2}(\theta)$ is bounded away from zero uniformly on Θ . Since $\hat{\sigma}_n^2(\theta) - \sigma_n^{*2}(\theta) = o_p(1)$ uniformly on Θ , $\hat{\sigma}_n^2(\theta)$ is bounded away from zero in probability uniformly on Θ . Hence, $\sup_{\theta \in \Theta} N_n^{-1} |\ell(\theta) - g_n(\theta)| = o_p(1)$.

To show the identifiable uniqueness of $N_n^{-1}g_n(\boldsymbol{\theta})$, we note that $N_n^{-1}g_n(\boldsymbol{\theta})$ is uniformly equicontinuous. In

$$\begin{aligned} & N_n^{-1}\{g_n(\boldsymbol{\theta}_1) - g_n(\boldsymbol{\theta}_2)\} \\ &= N_n^{-1}\{\log |\mathbf{S}_n(\boldsymbol{\theta}_1)| - \log |\mathbf{S}_n(\boldsymbol{\theta}_2)|\} - (1/2)\{\log \sigma_n^{*2}(\boldsymbol{\theta}_1) - \log \sigma_n^{*2}(\boldsymbol{\theta}_2)\} \\ &= N_n^{-1}\{\log |\mathbf{S}_n(\boldsymbol{\theta}_1)| - \log |\mathbf{S}_n(\boldsymbol{\theta}_2)|\} - \{2\bar{\sigma}_n^{*2}(\boldsymbol{\theta})\}^{-1}\{\sigma_n^{*2}(\boldsymbol{\theta}_1) - \sigma_n^{*2}(\boldsymbol{\theta}_2)\}, \end{aligned}$$

where $\bar{\sigma}_n^{*2}(\boldsymbol{\theta}) = \alpha\sigma_n^{*2}(\boldsymbol{\theta}_1) + (1-\alpha)\sigma_n^{*2}(\boldsymbol{\theta}_2)$ for some $\alpha \in (0, 1)$ and is bounded away

from zero, both terms are uniformly equicontinuous, since by (6.4), $N_n^{-1}(\log |\mathbf{S}_n(\boldsymbol{\theta}_1)| - \log |\mathbf{S}_n(\boldsymbol{\theta}_2)|) = -\sum_{k=1}^q \mathcal{O}(m_n^{-1})(\theta_{1,k} - \theta_{2,k})$, and with $\tilde{\boldsymbol{\theta}} = \alpha\boldsymbol{\theta}_1 + (1-\alpha)\boldsymbol{\theta}_2$ for some $\alpha \in (0, 1)$,

$$\begin{aligned} & \sigma_n^{*2}(\boldsymbol{\theta}_1) - \sigma_n^{*2}(\boldsymbol{\theta}_2) \\ &= -N_n^{-1}\sigma_0^2 \sum_{k=1}^q \text{tr} \left\{ \mathbf{S}_{0n}'^{-1} \mathbf{W}_{n,k}' \mathbf{S}_n(\tilde{\boldsymbol{\theta}}) \mathbf{S}_{0n}^{-1} + \mathbf{S}_{0n}'^{-1} \mathbf{S}_n'(\tilde{\boldsymbol{\theta}}) \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \right\} (\theta_{1,k} - \theta_{2,k}) \\ &= -\sigma_0^2 \mathcal{O}(m_n^{-1}) \sum_{k=1}^q (\theta_{1,k} - \theta_{2,k}). \end{aligned}$$

Together with (A.6) and (6.3), $N_n^{-1}g_n(\boldsymbol{\theta})$ is identifiably unique (Kelly (1955)).

Thus, by Theorem 3.4 of White (1994), the MLE of $\boldsymbol{\theta}$ is a consistent estimator.

By (6.2), the consistency of $\hat{\sigma}_n^2(\hat{\boldsymbol{\theta}}_n)$ can be derived from the consistency of $\sigma_n^{*2}(\hat{\boldsymbol{\theta}}_n)$, which is obvious from its definition. Further,

$$\begin{aligned} \hat{\beta}_n(\hat{\boldsymbol{\theta}}_n) &= \left\{ \mathbf{X}_n' \mathbf{S}_n'(\hat{\boldsymbol{\theta}}_n) \mathbf{S}_n(\hat{\boldsymbol{\theta}}_n) \mathbf{X}_n \right\}^{-1} \mathbf{X}_n' \mathbf{S}_n'(\hat{\boldsymbol{\theta}}_n) \mathbf{S}_n(\hat{\boldsymbol{\theta}}_n) \mathbf{Y}_n \\ &= \beta_0 + \sum_{k=1}^q (\theta_{0,k} - \hat{\theta}_{n,k}) N_n^{-1/2} \left\{ N_n^{-1} \mathbf{X}_n' \mathbf{S}_n'(\hat{\boldsymbol{\theta}}_n) \mathbf{S}_n(\hat{\boldsymbol{\theta}}_n) \mathbf{X}_n \right\}^{-1} \left\{ N_n^{-1/2} \mathbf{X}_n' \mathbf{S}_n'(\hat{\boldsymbol{\theta}}_n) \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \nu_{0n} \right\} \\ &\quad + N_n^{-1/2} \left\{ N_n^{-1} \mathbf{X}_n' \mathbf{S}_n'(\hat{\boldsymbol{\theta}}_n) \mathbf{S}_n(\hat{\boldsymbol{\theta}}_n) \mathbf{X}_n \right\}^{-1} \left\{ N_n^{-1/2} \mathbf{X}_n' \mathbf{S}_n'(\hat{\boldsymbol{\theta}}_n) \nu_{0n} \right\} \end{aligned}$$

where the last two terms are of order $o_p(1)$ by (A.1)–(A.5) and Lemma 4.

By (A.4), $\boldsymbol{\theta}_0$ is in the interior of Θ . Thus, for sufficiently small $\epsilon > 0$, we have $A_\epsilon = \{\boldsymbol{\eta} : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| < \epsilon\} \subset \mathbb{R}^p \times \Theta \times \mathbb{R}^+$ and $P(\hat{\boldsymbol{\eta}}_n \in A_\epsilon) \rightarrow 1$ as $n \rightarrow \infty$, where $\|\cdot\|$ denotes the Euclidean norm. Here we focus on $\hat{\boldsymbol{\eta}}_n \in A_\epsilon$. Now, we establish the asymptotic normality of the MLE by showing asymptotic normality of $N_n^{-1/2} \frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}}$ and convergence in probability of $N_n^{-1} \frac{\partial^2 \ell(\tilde{\boldsymbol{\eta}}_n)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}$, where $\tilde{\boldsymbol{\eta}}_n = \alpha \boldsymbol{\eta}_0 + (1-\alpha) \hat{\boldsymbol{\eta}}_n$ for $\alpha \in (0, 1)$ converges to $\boldsymbol{\eta}_0$ in probability.

For convergence of $N_n^{-1} \frac{\partial^2 \ell(\tilde{\boldsymbol{\eta}}_n)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}$, we show that, under (A.1)–(A.5), $N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\boldsymbol{\eta}}_n)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} - \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right\} = o_p(1)$ and $N_n^{-1} \left\{ \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} - E \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right\} = o_p(1)$. Here for an $a \times a$ random matrix

A_n , we let $A_n = \mathcal{O}_p(1)$ (or $o_p(1)$) if all the elements of A_n is of order $\mathcal{O}_p(1)$ (or $o_p(1)$). The second-order derivatives of $\ell(\eta)$ are

$$\begin{aligned}\frac{\partial^2 \ell(\eta)}{\partial \beta \partial \beta'} &= -(\sigma^2)^{-1} \mathbf{X}'_n \mathbf{S}'_n(\theta) \mathbf{S}_n(\theta) \mathbf{X}_n, \quad \frac{\partial^2 \ell(\eta)}{\partial \beta \partial \sigma^2} = -(\sigma^4)^{-1} \mathbf{X}'_n \mathbf{S}'_n(\theta) \boldsymbol{\nu}_n, \\ \frac{\partial^2 \ell(\eta)}{\partial \beta \partial \theta_k} &= -(\sigma^2)^{-1} \mathbf{X}'_n \{ \mathbf{W}'_k \mathbf{S}_n(\theta) + \mathbf{S}'_n(\theta) \mathbf{W}_{n,k} \} \mathbf{S}_n^{-1}(\theta) \boldsymbol{\nu}_n, \quad k = 1, \dots, q. \\ \frac{\partial^2 \ell(\eta)}{\partial \theta_k \partial \theta_l} &= -\text{tr} \{ \mathbf{W}_{n,k} \mathbf{S}_n^{-1}(\theta) \mathbf{W}_{n,l} \mathbf{S}_n^{-1}(\theta) \} - (\sigma^2)^{-1} \boldsymbol{\nu}'_n \mathbf{S}_n'^{-1}(\theta) \mathbf{W}_{n,l}' \mathbf{W}_{n,k} \mathbf{S}_n^{-1}(\theta) \boldsymbol{\nu}_n, \\ &\quad k, l = 1, \dots, q. \\ \frac{\partial^2 \ell(\eta)}{\partial \theta_k \partial \sigma^2} &= -(\sigma^4)^{-1} \boldsymbol{\nu}'_n \mathbf{W}_{n,k} \mathbf{S}_n^{-1}(\theta) \boldsymbol{\nu}_n, \quad k = 1, \dots, q. \quad \frac{\partial^2 \ell(\eta)}{\partial (\sigma^2)^2} = (2\sigma^6)^{-1} (-2\boldsymbol{\nu}'_n \boldsymbol{\nu}_n + N_n \sigma^2).\end{aligned}$$

By (A.1)–(A.5), we have

$$\begin{aligned}& N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\eta}_n)}{\partial \beta \partial \beta'} - \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \beta'} \right\} \\ &= N_n^{-1} \mathbf{X}'_n \mathbf{S}'_{0n} \mathbf{S}_{0n} \mathbf{X}_n (1/\sigma_0^2 - 1/\tilde{\sigma}_n^2) - (N_n \tilde{\sigma}_n^2)^{-1} \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{X}'_n (\mathbf{S}'_{0n} \mathbf{W}_{n,k} + \mathbf{W}_{n,k}' \mathbf{S}_{0n}) \mathbf{X}_n \\ &\quad + (N_n \tilde{\sigma}_n^2)^{-1} \mathbf{X}'_n \left\{ \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{W}_{n,k}' \right\} \left\{ \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{W}_{n,k} \right\} \mathbf{X}_n = o_p(1), \\ & N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\eta}_n)}{\partial \beta \partial \sigma^2} - \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \sigma^2} \right\} \\ &= N_n^{-1} \mathbf{X}'_n \mathbf{S}'_{0n} \boldsymbol{\nu}_{0n} (1/\sigma_0^4 - 1/\tilde{\sigma}_n^4) - (N_n \tilde{\sigma}_n^4)^{-1} \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{X}'_n (\mathbf{S}'_{0n} \mathbf{W}_{n,k} + \mathbf{W}_{n,k}' \mathbf{S}_{0n}) \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} \\ &\quad + (N_n \tilde{\sigma}_n^4)^{-1} \mathbf{X}'_n \left\{ \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{W}_{n,k}' \right\} \left\{ \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{W}_{n,k} \right\} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} \\ &\quad - (\tilde{\sigma}_n^4)^{-1} \mathbf{X}'_n \mathbf{S}'_n(\tilde{\theta}_n) \mathbf{S}_n(\tilde{\theta}_n) \mathbf{X}_n (\beta_0 - \tilde{\beta}_n) = o_p(1), \\ & N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\eta}_n)}{\partial \beta \partial \theta_k} - \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \theta_k} \right\} = -(N_n \tilde{\sigma}_n^2)^{-1} \mathbf{X}'_n \{ \mathbf{W}'_{n,k} \mathbf{S}_n(\tilde{\theta}_n) + \mathbf{S}'_n(\tilde{\theta}_n) \mathbf{W}_{n,k} \} (\mathbf{Y}_n - \mathbf{X}_n \beta_0) \\ &\quad + (N_n \sigma_0^2)^{-1} \mathbf{X}'_n (\mathbf{W}'_{n,k} \mathbf{S}_{0n} + \mathbf{S}'_{0n} \mathbf{W}_{n,k}) (\mathbf{Y}_n - \mathbf{X}_n \beta_0) \\ &\quad - (N_n \tilde{\sigma}_n^2)^{-1} \mathbf{X}'_n \{ \mathbf{W}'_{n,k} \mathbf{S}_n(\tilde{\theta}_n) + \mathbf{S}'_n(\tilde{\theta}_n) \mathbf{W}_{n,k} \} \mathbf{X}_n (\beta_0 - \tilde{\beta}_n) \\ &= o_p(1), \\ & N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\eta}_n)}{\partial \theta_k \partial \theta_l} - \frac{\partial^2 \ell(\eta_0)}{\partial \theta_k \partial \theta_l} \right\} \\ &= N_n^{-1} \left[-\text{tr} \{ \mathbf{W}_{n,k} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \mathbf{W}_{n,l} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \} + \text{tr} \{ \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \mathbf{W}_{n,l} \mathbf{S}_{0n}^{-1} \} \right] \\ &\quad - N_n^{-1} \boldsymbol{\nu}'_{0n} \mathbf{S}_{0n}'^{-1} \mathbf{W}_{n,l}' \mathbf{W}_{n,k} \mathbf{S}_{0n} \boldsymbol{\nu}_{0n} (1/\tilde{\sigma}_n^2 - 1/\sigma_0^2) - (N_n \tilde{\sigma}_n^2)^{-1} (\beta_0 - \tilde{\beta}_n)' \mathbf{X}'_n \mathbf{W}_{n,l}' \mathbf{W}_{n,k} \mathbf{X}_n \\ &\quad \times (\beta_0 - \tilde{\beta}_n) - 2(N_n \tilde{\sigma}_n^2)^{-1} (\beta_0 - \tilde{\beta}_n)' \mathbf{X}'_n \mathbf{W}_{n,l}' \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} = o_p(1),\end{aligned}$$

where

$$\begin{aligned}
& \text{tr} \left\{ \mathbf{W}_{n,k} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \mathbf{W}_{n,l} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \right\} \\
&= \text{tr} \left\{ \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \mathbf{W}_{n,l} \mathbf{S}_{0n}^{-1} \right\} + \sum_{j=1}^q \text{tr} \left\{ \mathbf{W}_{n,k} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \mathbf{W}_{n,j} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \mathbf{W}_{n,l} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \right. \\
&\quad \left. + \mathbf{W}_{n,k} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \mathbf{W}_{n,l} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \mathbf{W}_{n,j} \mathbf{S}_n^{-1}(\tilde{\theta}_n) \right\} (\tilde{\theta}_{n,j} - \theta_{0,j}) \\
&= \text{tr} \left\{ \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \mathbf{W}_{n,l} \mathbf{S}_{0n}^{-1} \right\} + \mathcal{O}(N_n/m_n) o_p(1), \\
& N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\boldsymbol{\eta}}_n)}{\partial \theta_k \partial \sigma^2} - \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \theta_k \partial \sigma^2} \right\} \\
&= N_n^{-1} \left\{ -(\tilde{\sigma}_n^4)^{-1} \boldsymbol{\nu}'_{0n} \mathbf{S}_{0n}^{'-1} \mathbf{S}'_n(\tilde{\theta}_n) \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} + (\sigma_0^4)^{-1} \boldsymbol{\nu}'_{0n} \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} \right\} \\
&\quad - (N_n \tilde{\sigma}_n^4)^{-1} (\beta_0 - \tilde{\beta}_n)' \mathbf{X}'_n \mathbf{S}'_n(\tilde{\theta}_n) \mathbf{W}_{n,k} \mathbf{X}_n (\beta_0 - \tilde{\beta}_n) \\
&\quad - (N_n \tilde{\sigma}_n^4)^{-1} (\beta_0 - \tilde{\beta}_n)' \mathbf{X}'_n \mathbf{S}'_n(\tilde{\theta}_n) \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} - (N_n \tilde{\sigma}_n^4)^{-1} \boldsymbol{\nu}'_{0n} \mathbf{S}_{0n}^{'-1} \mathbf{S}'_n(\tilde{\theta}_n) \mathbf{W}_{n,k} \mathbf{X}_n (\beta_0 - \tilde{\beta}_n) \\
&= o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
& N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\boldsymbol{\eta}}_n)}{\partial (\sigma^2)^2} - \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial (\sigma^2)^2} \right\} \\
&= N_n^{-1} \boldsymbol{\nu}'_{0n} \boldsymbol{\nu}_{0n} (1/\sigma_0^6 - 1/\tilde{\sigma}_n^6) - (N_n \tilde{\sigma}_n^6)^{-1} \boldsymbol{\nu}'_{0n} \mathbf{S}_{0n}^{'-1} \left\{ \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{W}'_{n,k} \right\} \\
&\quad \left\{ \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{W}_{n,k} \right\} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} - 2(N_n \tilde{\sigma}_n^6)^{-1} \boldsymbol{\nu}'_{0n} \left\{ \sum_{k=1}^q (\theta_{0,k} - \tilde{\theta}_{n,k}) \mathbf{W}_{n,k} \right\} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} \\
&\quad + \{(2\sigma_0^4)^{-1} - (2\tilde{\sigma}_n^4)^{-1}\} = o_p(1).
\end{aligned}$$

Thus $N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\boldsymbol{\eta}}_n)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} - \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right\} = o_p(1)$.

Further, under (A.1)–(A.5), we have

$$\begin{aligned}
N_n^{-1} \left\{ \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \beta'} - E \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \beta'} \right\} &= \mathbf{0}, \\
N_n^{-1} \left\{ \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \sigma^2} - E \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \sigma^2} \right\} &= (N_n \sigma_0^4)^{-1} \mathbf{X}'_n \mathbf{S}'_{0n} \boldsymbol{\nu}_{0n} = N_n^{-1/2} \mathcal{O}_p(1) = o_p(1), \\
N_n^{-1} \left\{ \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \theta_k} - E \frac{\partial^2 \ell(\eta_0)}{\partial \beta \partial \theta_k} \right\} &= (N_n \sigma_0^2)^{-1} \mathbf{X}'_n (\mathbf{W}'_{n,k} \mathbf{S}_{0n} + \mathbf{S}'_{0n} \mathbf{W}_{n,k}) \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} \\
&= N_n^{-1/2} \mathcal{O}_p(1) = o_p(1), \\
N_n^{-1} \left\{ \frac{\partial^2 \ell(\eta_0)}{\partial \theta_k \partial \theta_l} - E \frac{\partial^2 \ell(\eta_0)}{\partial \theta_k \partial \theta_l} \right\} &= -(N_n \sigma_0^2)^{-1} \left\{ \boldsymbol{\nu}'_{0n} \mathbf{S}_{0n}^{-1} \mathbf{W}'_{n,l} \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} \right. \\
&\quad \left. - \sigma_0^2 \text{tr}(\mathbf{S}_{0n}^{-1} \mathbf{W}'_{n,l} \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1}) \right\} = N_n^{-1} o_p(N_n/m_n) = o_p(1), \\
N_n^{-1} \left\{ \frac{\partial^2 \ell(\eta_0)}{\partial \theta_k \partial \sigma^2} - E \frac{\partial^2 \ell(\eta_0)}{\partial \theta_k \partial \sigma^2} \right\} &= (N_n \sigma_0^4)^{-1} \left\{ \boldsymbol{\nu}'_{0n} \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} - \sigma_0^2 \text{tr}(\mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1}) \right\} = o_p(1), \\
N_n^{-1} \left\{ \frac{\partial^2 \ell(\eta_0)}{(\partial \sigma^2)^2} - E \frac{\partial^2 \ell(\eta_0)}{(\partial \sigma^2)^2} \right\} &= -(N_n \sigma_0^6)^{-1} (\boldsymbol{\nu}'_{0n} \boldsymbol{\nu}_{0n} - N_n) = o_p(1).
\end{aligned}$$

Thus $N_n^{-1} \left\{ \frac{\partial^2 \ell(\eta_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} - E \frac{\partial^2 \ell(\eta_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right\} = o_p(1)$. It follows that $N_n^{-1} \frac{\partial^2 \ell(\tilde{\eta}_n)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \rightarrow \boldsymbol{\Sigma}_{\eta_0}^{-1} = \lim_{n \rightarrow \infty} N_n^{-1} E \frac{\partial^2 \ell(\eta_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}$.

Furthermore, the first-order derivatives of $\ell(\boldsymbol{\eta})$ at $\boldsymbol{\theta}_0$ are linear or quadratic forms of $\boldsymbol{\nu}_{0n}$, since

$$\begin{aligned}
\frac{\partial \ell(\eta_0)}{\partial \beta} &= (\sigma_0^2)^{-1} \mathbf{X}'_n \mathbf{S}'_{0n} \boldsymbol{\nu}_{0n}, \quad \frac{\partial \ell(\eta_0)}{\partial \sigma^2} = (2\sigma_0^4)^{-1} (\boldsymbol{\nu}'_{0n} \boldsymbol{\nu}_{0n} - N_n \sigma_0^2) \\
\frac{\partial \ell(\eta_0)}{\partial \theta_k} &= -\text{tr} \left\{ \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \right\} + (\sigma_0^2)^{-1} \boldsymbol{\nu}'_{0n} \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1} \boldsymbol{\nu}_{0n} = -\text{tr}(\mathbf{G}_k) + (\sigma_0^2)^{-1} \boldsymbol{\nu}'_{0n} \mathbf{G}_k \boldsymbol{\nu}_{0n},
\end{aligned}$$

where $\mathbf{G}_k = \mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1}$ for $k = 1, \dots, q$. By (A.5) and Lemma 4, we have

$$N_n^{-1/2} \frac{\partial \ell(\eta_0)}{\partial \beta} \xrightarrow{D} N(0, \lim_{n \rightarrow \infty} (N_n \sigma_0^2)^{-1} \mathbf{X}'_n \mathbf{S}'_{0n} \mathbf{S}_{0n} \mathbf{X}_n).$$

By a classic central limit theorem,

$$N_n^{-1/2} \frac{\partial \ell(\eta_0)}{\partial \sigma^2} \xrightarrow{D} N(0, (2\sigma_0^4)^{-1}).$$

For the asymptotic normality of $N_n^{-1/2} \frac{\partial \ell(\eta_0)}{\partial \theta_k}$, (A.2) and (A.4) ensure that \mathbf{G}_k is uniformly bounded in matrix norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ and the positive definiteness of $\boldsymbol{\Sigma}_{\eta_0}^{-1}$ ensures that $N_n^{-1} \text{Var}(\frac{\partial \ell(\eta_0)}{\partial \theta_k}) = N_n^{-1} \text{tr}(\mathbf{G}_k^2 + \mathbf{G}'_k \mathbf{G}_k)$ is bounded away from zero. By a central limit theorem for linear-quadratic forms (Theorem 1, Kelejian

and Prucha (2001)), we have

$$N_n^{-1/2} \frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \theta_k} \xrightarrow{D} N(0, \lim_{n \rightarrow \infty} N_n^{-1} \text{tr}(\mathbf{G}_k^2 + \mathbf{G}_k' \mathbf{G}_k))$$

for $k = 1, \dots, q$. Consider a linear combination of $\frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \sigma^2}$ and $\frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \theta_k}$, $k = 1, \dots, q$ with coefficients c_1, \dots, c_{q+1} , we have

$$\begin{aligned} \sum_{k=1}^q c_k N_n^{-1/2} \frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \theta_k} + c_{q+1} N_n^{-1/2} \frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \sigma^2} &= N_n^{-1/2} (\sigma_0^2)^{-1} \boldsymbol{\nu}_{0n}' \left\{ \sum_{k=1}^q c_k \mathbf{G}_k + c_{q+1} (2\sigma_0^2)^{-1} \mathbf{I}_n \right\} \boldsymbol{\nu}_{0n} \\ &\quad - \sum_{k=1}^q c_k N_n^{-1/2} \text{tr}(\mathbf{G}_k) - c_{q+1} N_n^{1/2} (2\sigma_0^2)^{-1} \end{aligned}$$

which is again a linear-quadratic form of $\boldsymbol{\nu}_{0n}$. Thus its convergence in distribution holds, again by Theorem 1 of Kelejian and Prucha (2001). Then by the Cramér-Wold theorem and the fact that $\frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \beta}$ is asymptotically independent of $\frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \sigma^2}$ and $\frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \theta_k}$, $k = 1, \dots, q$, we have

$$N_n^{-1/2} \frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_0}^{-1})$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\eta}_0}^{-1} = \lim_{n \rightarrow \infty} E \left(-N_n^{-1} \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right)$ and $E \left(-N_n^{-1} \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right)$ is

$$\begin{bmatrix} (N_n \sigma_0^2)^{-1} \mathbf{X}_n' \mathbf{S}_{0n}' \mathbf{S}_{0n} \mathbf{X}_n & 0 & \dots & 0 & 0 \\ 0 & N_n^{-1} \text{tr}(\mathbf{G}_1^2 + \mathbf{G}_1' \mathbf{G}_1) & \dots & N_n^{-1} \text{tr}(\mathbf{G}_1 \mathbf{G}_q + \mathbf{G}_1' \mathbf{G}_q) & (N_n \sigma_0^2)^{-1} \text{tr}(\mathbf{G}_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & N_n^{-1} \text{tr}(\mathbf{G}_q \mathbf{G}_1 + \mathbf{G}_1' \mathbf{G}_q) & \dots & N_n^{-1} \text{tr}(\mathbf{G}_q^2 + \mathbf{G}_q' \mathbf{G}_q) & (N_n \sigma_0^2)^{-1} \text{tr}(\mathbf{G}_q) \\ 0 & (N_n \sigma_0^2)^{-1} \text{tr}(\mathbf{G}_1) & \dots & (N_n \sigma_0^2)^{-1} \text{tr}(\mathbf{G}_q) & (2\sigma_0^4)^{-1} \end{bmatrix}$$

It follows that

$$N_n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = - \left\{ N_n^{-1} \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} + o_p(1) \right\}^{-1} N_n^{-1/2} \frac{\partial \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_0})$$

and thus the result of this theorem holds. Note that m_n needs to be bounded to ensure that the limiting variance-covariance matrix is nonsingular. When $m_n \rightarrow \infty$, for example, $N_n^{-1} \text{tr}(\mathbf{G}_1^2 + \mathbf{G}_1' \mathbf{G}_1) = N_n^{-1} \mathcal{O}(N_n/m_n) = o(1)$, which will result in a singular limiting variance-covariance matrix. \square

Appendix B: Proof of Theorem 2.

PROOF. We establish consistency of the MLE as in the proof of Theorem 1.

Note that

$$m_n N_n^{-1} \{\ell(\boldsymbol{\theta}) - g_n(\boldsymbol{\theta})\} = -m_n \{2\tilde{\sigma}_n^2(\boldsymbol{\theta})\}^{-1} \{\hat{\sigma}_n^2(\boldsymbol{\theta}) - \sigma_n^{*2}(\boldsymbol{\theta})\}$$

is of order $o_p(1)$ uniformly on Θ under (A.1)–(A.5), where $\tilde{\sigma}_n^2(\boldsymbol{\theta}) = \alpha \hat{\sigma}_n^2(\boldsymbol{\theta}) + (1 - \alpha) \sigma_n^{*2}(\boldsymbol{\theta})$ for some $\alpha \in (0, 1)$. Thus $\sup_{\boldsymbol{\theta} \in \Theta} m_n N_n^{-1} |\ell(\boldsymbol{\theta}) - g_n(\boldsymbol{\theta})| = o_p(1)$. To show the identifiable uniqueness of $m_n N_n^{-1} g_n(\boldsymbol{\theta})$, it suffices to show that it is uniformly equicontinuous. Note that

$$\begin{aligned} & m_n N_n^{-1} \{g_n(\boldsymbol{\theta}_1) - g_n(\boldsymbol{\theta}_2)\} \\ &= m_n N_n^{-1} \{\log |S_n(\boldsymbol{\theta}_1)| - \log |S_n(\boldsymbol{\theta}_2)|\} - m_n \{2\tilde{\sigma}_n^{*2}(\boldsymbol{\theta})\}^{-1} \{\sigma_n^{*2}(\boldsymbol{\theta}_1) - \sigma_n^{*2}(\boldsymbol{\theta}_2)\} \end{aligned}$$

where $\tilde{\sigma}_n^{*2}(\boldsymbol{\theta}) = \alpha \sigma_n^{*2}(\boldsymbol{\theta}_1) + (1 - \alpha) \sigma_n^{*2}(\boldsymbol{\theta}_2)$ for some $\alpha \in (0, 1)$ and is bounded away from zero. With arguments similar to those in the proof of Theorem 1, we have the uniform equicontinuity of $m_n N_n^{-1} g_n(\boldsymbol{\theta})$. Thus together with (A.6') and the fact that $m_n N_n^{-1} \{g_n(\boldsymbol{\theta}) - g_n(\boldsymbol{\theta}_0)\} \leq 0$, we have the identifiable uniqueness of $m_n N_n^{-1} g_n(\boldsymbol{\theta})$. By Theorem 3.4 of White (1994), the MLE of $\boldsymbol{\theta}$ is consistent. The consistency of MLE of β and σ^2 can be shown using arguments similar to those in the proof of Theorem 1. We omit the details.

For sufficiently small $\epsilon > 0$, we have $A_\epsilon = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \epsilon\} \subset \Theta$ and $P(\hat{\boldsymbol{\theta}}_n \in A_\epsilon) \rightarrow 1$ as $n \rightarrow \infty$. Here we focus on $\hat{\boldsymbol{\theta}}_n \in A_\epsilon$. Now, we establish the asymptotic normality of the MLE $\hat{\boldsymbol{\theta}}_n$ by showing the asymptotic normality of $(m_n/N_n)^{1/2} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ and the convergence in probability of $m_n N_n^{-1} \frac{\partial^2 \ell(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, where $\tilde{\boldsymbol{\theta}}_n = \alpha \boldsymbol{\theta}_0 + (1 - \alpha) \hat{\boldsymbol{\theta}}_n$ for $\alpha \in (0, 1)$ converges to $\boldsymbol{\theta}_0$ in probability.

For convergence of $m_n N_n^{-1} \frac{\partial^2 \ell(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, we show that $m_n N_n^{-1} \{\frac{\partial^2 \ell(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\} = o_p(1)$ and $m_n N_n^{-1} \left\{ \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} - E \frac{\partial^2 \ell(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \right\} = o_p(1)$, under (A.1)–(A.5). We note that

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} &= -\{2\hat{\sigma}_n^2(\boldsymbol{\theta})\}^{-1} \boldsymbol{\nu}'_{0n} \frac{\partial B_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} \boldsymbol{\nu}_{0n} - \text{tr}\{W_{n,1} S_n^{-1}(\boldsymbol{\theta})\} \\ &= \{\hat{\sigma}_n^2(\boldsymbol{\theta})\}^{-1} \boldsymbol{\nu}'_{0n} T_{n,1}(\boldsymbol{\theta}) \boldsymbol{\nu}_{0n} - \text{tr}\{W_{n,1} S_n^{-1}(\boldsymbol{\theta})\} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1^2} &= 2\{N_n \hat{\sigma}_n^4(\boldsymbol{\theta})\}^{-1} \{\boldsymbol{\nu}'_{0n} T_{n,1}(\boldsymbol{\theta}) \boldsymbol{\nu}_{0n}\}^2 + \{\hat{\sigma}_n^2(\boldsymbol{\theta})\}^{-1} \boldsymbol{\nu}'_{0n} \frac{\partial T_{n,1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} \boldsymbol{\nu}_{0n} \\ &\quad - \text{tr}\{S_n^{-1}(\boldsymbol{\theta}) W_{n,1} S_n^{-1}(\boldsymbol{\theta}) W_{n,1}\} \end{aligned}$$

where $B_n(\theta)$ is given in (6.1), $T_{n,1}(\theta) = -(1/2) \frac{\partial B_n(\theta)}{\partial \theta_1}$ is

$$\begin{aligned} S_{0n}'^{-1} & \left[W_{n,1}' M_n S_n(\theta) - S_n'(\theta) W_{n,1} X_n \{X_n' S_n'(\theta) S_n(\theta) X_n\}^{-1} X_n' S_n'(\theta) S_n(\theta) \right. \\ & + S_n'(\theta) S_n(\theta) X_n \{X_n' S_n'(\theta) S_n(\theta) X_n\}^{-1} X_n' W_{n,1}' S_n(\theta) X_n \\ & \left. \times \{X_n' S_n'(\theta) S_n(\theta) X_n\}^{-1} X_n' S_n'(\theta) S_n(\theta) \right] S_{0n}^{-1} \end{aligned}$$

and $M_n = I_n - S_n(\theta) X_n \{X_n' S_n'(\theta) S_n(\theta) X_n\}^{-1} X_n' S_n'(\theta)$. By (A.1)–(A.5), and

Lemma 5,

$$\begin{aligned} m_n N_n^{-1} \nu_{0n}' T_{n,1}(\theta) \nu_{0n} &= \mathcal{O}_p(1) \\ m_n N_n^{-1} \nu_{0n}' \frac{\partial T_{n,1}(\theta)}{\partial \theta_1} \nu_{0n} &= m_n N_n^{-1} \sigma_0^2 \text{tr} \left\{ \frac{\partial T_{n,1}(\theta)}{\partial \theta_1} \right\} + o_p(1), \end{aligned}$$

since $T_{n,1}(\theta)$ and $\frac{\partial T_{n,1}(\theta)}{\partial \theta_1}$ are uniformly bounded in either matrix norm $\|\cdot\|_1$ or $\|\cdot\|_\infty$, which are ensured by the submultiplicative property of a matrix norm (Horn and Johnson (1985)). Thus, under (A.1)–(A.5),

$$\begin{aligned} m_n N_n^{-1} \frac{\partial^2 \ell(\theta)}{\partial \theta_1^2} &= 2 \{m_n \hat{\sigma}_n^4(\theta)\}^{-1} \left\{ m_n N_n^{-1} \nu_{0n}' T_{n,1}(\theta) \nu_{0n} \right\}^2 \\ &\quad + \{\hat{\sigma}_n^2(\theta)\}^{-1} m_n N_n^{-1} \nu_{0n}' \frac{\partial T_{n,1}(\theta)}{\partial \theta_1} \nu_{0n} \\ &\quad - m_n N_n^{-1} \text{tr} \left\{ W_{n,1} S_n^{-1}(\theta) W_{n,1} S_n^{-1}(\theta) \right\} \\ &= m_n N_n^{-1} \text{tr} \left\{ \frac{\partial T_{n,1}(\theta)}{\partial \theta_1} \right\} - m_n N_n^{-1} \text{tr} \left\{ W_{n,1} S_n^{-1}(\theta) W_{n,1} S_n^{-1}(\theta) \right\} + o_p(1) \end{aligned}$$

and for $\tilde{\theta}_n = \alpha \theta_0 + (1 - \alpha) \hat{\theta}_n$,

$$\begin{aligned} & m_n N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\theta}_n)}{\partial \theta_1^2} - \frac{\partial^2 \ell(\theta_0)}{\partial \theta_1^2} \right\} \\ &= m_n N_n^{-1} \left[\text{tr} \left\{ \frac{\partial T_{n,1}(\tilde{\theta}_n)}{\partial \theta_1} \right\} - \text{tr} \left\{ \frac{\partial T_{n,1}(\theta_0)}{\partial \theta_1} \right\} \right] \\ &\quad - m_n N_n^{-1} \left[\text{tr} \left\{ S_n^{-1}(\tilde{\theta}_n) W_{n,1} S_n^{-1}(\tilde{\theta}_n) W_{n,1} \right\} - \text{tr} \left\{ S_{0n}^{-1} W_{n,1} S_{0n}^{-1} W_{n,1} \right\} \right] + o_p(1) \\ &= (\tilde{\theta}_n - \theta_0)' \mathcal{O}(1) - (\tilde{\theta}_n - \theta_0)' \mathcal{O}(1) + o_p(1) = o_p(1). \end{aligned}$$

By similar arguments for θ_k and $\theta_{k'}$, $k, k' = 1, \dots, q$, we have $m_n N_n^{-1} \left\{ \frac{\partial^2 \ell(\tilde{\theta}_n)}{\partial \theta \partial \theta'} - \frac{\partial^2 \ell(\theta_0)}{\partial \theta \partial \theta'} \right\} = o_p(1)$.

Furthermore,

$$\begin{aligned} -m_n N_n^{-1} \frac{\partial^2 \ell(\theta_0)}{\partial \theta_1^2} &= -m_n N_n^{-1} \text{tr} \left\{ \frac{\partial T_{n,1}(\theta_0)}{\partial \theta_1} \right\} + m_n N_n^{-1} \text{tr}(\mathbf{G}_1^2) + o_p(1) \\ &= m_n N_n^{-1} \left\{ \text{tr}(\mathbf{G}_1' \mathbf{G}_1) + \text{tr}(\mathbf{G}_1^2) \right\} + o_p(1), \end{aligned}$$

since

$$\begin{aligned}
& m_n N_n^{-1} \text{tr} \left\{ \frac{\partial T_{n,1}(\theta_0)}{\partial \theta_1} \right\} \\
&= -m_n N_n^{-1} \text{tr} (S_{0n}'^{-1} W_{n,1}' W_{n,1} S_{0n}^{-1}) + m_n N_n^{-1} \text{tr} \{ (X_n' S_{0n}' S_{0n} X_n)^{-1} X_n' W_{n,1}' W_{n,1} X_n \} \\
&\quad - m_n N_n^{-1} \text{tr} \{ (X_n' S_{0n}' S_{0n} X_n)^{-1} X_n' (S_{0n}' W_{n,1} + W_{n,1}' S_{0n}) X_n (X_n' S_{0n}' S_{0n} X_n)^{-1} \\
&\quad \times X_n' (S_{0n}' W_{n,1} + W_{n,1}' S_{0n}) X_n \} + m_n N_n^{-1} \text{tr} \{ (X_n' S_{0n}' S_{0n} X_n)^{-1} X_n' (S_{0n}' W_{n,1} \\
&\quad + W_{n,1}' S_{0n}) S_{0n}^{-1} S_{0n}'^{-1} (S_{0n}' W_{n,1} + W_{n,1}' S_{0n}) X_n \} = -m_n N_n^{-1} \text{tr} \{ G_1' G_1 \} + o(1).
\end{aligned}$$

By similar arguments for θ_k and $\theta_{k'}$, $k, k' = 1, \dots, q$, we have $m_n N_n^{-1} \left\{ \frac{\partial^2 \ell(\theta_0)}{\partial \theta \partial \theta'} - E \frac{\partial^2 \ell(\eta_0)}{\partial \theta \partial \theta'} \right\} = o_p(1)$. It follows that $m_n N_n^{-1} \frac{\partial^2 \ell(\hat{\theta}_n)}{\partial \theta \partial \theta'} \rightarrow \Sigma_{\theta_0}^{-1} = \lim_{n \rightarrow \infty} m_n N_n^{-1} E \frac{\partial^2 \ell(\eta_0)}{\partial \theta \partial \theta'}$.

To establish the asymptotic normality of $(m_n/N_n)^{1/2} \frac{\partial \ell(\eta_0)}{\partial \theta}$, Theorem 1 in Kelejian and Prucha (2001) no longer applies here, since the variances of the linear-quadratic forms over N_n are not bounded away from zero. Instead we apply the central limit theorem for linear-quadratic forms in Appendix A of Lee (2004) to establish the asymptotic normality of $\hat{\theta}_n$.

Note that

$$\begin{aligned}
(m_n/N_n)^{1/2} \frac{\partial \ell(\theta_0)}{\partial \theta_1} &= (m_n/N_n)^{1/2} \{ \hat{\sigma}_n^2(\theta_0) \}^{-1} \nu_{0n}' T_{n,1}(\theta_0) \nu_{0n} - (m_n/N_n)^{1/2} \text{tr}(W_{n,1} S_{0n}^{-1}) \\
&= \{ \hat{\sigma}^2(\theta_0) \}^{-1} (m_n/N_n)^{1/2} \{ \nu_{0n}' G_1 \nu_{0n} - \sigma_0^2 \text{tr}(G_1) \} + o_p(1).
\end{aligned}$$

(A.2) and (A.4) ensure that $G_1 = W_{n,1} S_{0n}^{-1}$ is uniformly bounded in matrix norms $||| \cdot |||_1$ and $||| \cdot |||_\infty$ and the positive definiteness of $\Sigma_{\theta_0}^{-1}$ ensures that $(m_n/N_n) \text{Var}(\nu_{0n}' G_1 \nu_{0n}) = (m_n/N_n) \sigma_0^2 \text{tr}(G_1^2 + G_1' G_1)$ is bounded away from zero. By Lee (2004), we have

$$(m_n/N_n)^{1/2} \{ \nu_{0n}' G_1 \nu_{0n} - \sigma_0^2 \text{tr}(G_1) \} \xrightarrow{D} N \left(0, \lim_{n \rightarrow \infty} m_n N_n^{-1} \sigma_0^4 \{ \text{tr}(G_1 G_1') + \text{tr}(G_1^2) \} \right).$$

Hence

$$\begin{aligned}
(m_n/N_n)^{1/2} \frac{\partial \ell(\theta_0)}{\partial \theta_1} &= \{ \hat{\sigma}^2(\theta_0) \}^{-1} (m_n/N_n)^{1/2} \{ \nu_{0n}' G_1 \nu_{0n} - \sigma_0^2 \text{tr}(W_{n,1} S_{0n}^{-1}) \} + o_p(1) \\
&\xrightarrow{D} N \left(0, \lim_{n \rightarrow \infty} m_n N_n^{-1} \{ \text{tr}(G_1 G_1') + \text{tr}(G_1^2) \} \right)
\end{aligned}$$

Then similarly for θ_k , $k = 2, \dots, q$, we have

$$(m_n/N_n)^{1/2} \frac{\partial \ell(\theta_0)}{\partial \theta_k} \xrightarrow{D} N \left(0, \lim_{n \rightarrow \infty} m_n N_n^{-1} \{ \text{tr}(G_k G_k') + \text{tr}(G_k^2) \} \right).$$

For any given set of coefficients $c_k, k = 1, \dots, q$, we have

$$\begin{aligned}
& \sum_{k=1}^q c_k (m_n/N_n)^{1/2} \frac{\partial \ell(\theta_0)}{\partial \theta_k} \\
&= (m_n/N_n)^{1/2} \{\hat{\sigma}_n^2(\theta_0)\}^{-1} \nu'_{0n} \sum_{k=1}^q c_k T_{n,k}(\theta_0) \nu_{0n} - (m_n/N_n)^{1/2} \sum_{k=1}^q c_k \text{tr}(\mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1}) \\
&= \{\hat{\sigma}^2(\theta_0)\}^{-1} (m_n/N_n)^{1/2} \left\{ \nu'_{0n} \sum_{k=1}^q c_k \mathbf{G}_k \nu_{0n} - \sigma_0^2 \sum_{k=1}^q c_k \text{tr}(\mathbf{W}_{n,k} \mathbf{S}_{0n}^{-1}) \right\} + o_p(1) \\
&\xrightarrow{D} N \left(0, \lim_{n \rightarrow \infty} m_n N_n^{-1} \left\{ \sum_{k=1}^q \sum_{k'=1}^q c_k c_{k'} [\text{tr}(\mathbf{G}_i \mathbf{G}_j') + \text{tr}(\mathbf{G}_i \mathbf{G}_j)] \right\} \right).
\end{aligned}$$

Thus, by the Cramér-Wold theorem,

$$(m_n/N_n)^{1/2} \frac{\partial \ell(\theta_0)}{\partial \theta} \xrightarrow{D} N(\mathbf{0}, \Sigma_{\theta_0}^{-1})$$

where

$$\begin{aligned}
\Sigma_{\theta_0}^{-1} &= \lim_{n \rightarrow \infty} -m_n N_n^{-1} E \left\{ \frac{\partial^2 \ell(\eta_0)}{\partial \theta \partial \theta'} \right\} \\
&= \lim_{n \rightarrow \infty} \begin{bmatrix} m_n N_n^{-1} \text{tr}(\mathbf{G}_1^2 + \mathbf{G}_1' \mathbf{G}_1) & \dots & m_n N_n^{-1} \text{tr}(\mathbf{G}_1 \mathbf{G}_q + \mathbf{G}_q' \mathbf{G}_1) \\ \vdots & \ddots & \vdots \\ m_n N_n^{-1} \text{tr}(\mathbf{G}_q \mathbf{G}_1 + \mathbf{G}_1' \mathbf{G}_q) & \dots & m_n N_n^{-1} \text{tr}(\mathbf{G}_q^2 + \mathbf{G}_q' \mathbf{G}_q) \end{bmatrix}.
\end{aligned}$$

Thus it follows that

$$(N_n/m_n)^{1/2} (\hat{\theta}_n - \theta_0) = - \left\{ m_n N_n^{-1} \frac{\partial^2 \ell(\theta_0)}{\partial \theta \partial \theta'} + o_p(1) \right\}^{-1} (m_n/N_n)^{1/2} \frac{\partial \ell(\theta_0)}{\partial \theta} \xrightarrow{D} N(\mathbf{0}, \Sigma_{\theta_0}).$$

To establish the asymptotic normality of $\hat{\beta}_n(\hat{\theta}_n)$, we have

$$\begin{aligned}
\hat{\beta}_n(\hat{\theta}_n) - \beta_0 &= \left\{ \mathbf{X}_n' \mathbf{S}_n'(\hat{\theta}_n) \mathbf{S}_n(\hat{\theta}_n) \mathbf{X}_n \right\}^{-1} \mathbf{X}_n' \mathbf{S}_n'(\hat{\theta}_n) \mathbf{S}_n(\hat{\theta}_n) \mathbf{Y}_n - \beta_0 \\
&= \left\{ \mathbf{X}_n' \mathbf{S}_n'(\hat{\theta}_n) \mathbf{S}_n(\hat{\theta}_n) \mathbf{X}_n \right\}^{-1} \mathbf{X}_n' \mathbf{S}_n'(\hat{\theta}_n) \mathbf{S}_n(\hat{\theta}_n) \mathbf{S}_{0n}^{-1} \nu_{0n}.
\end{aligned}$$

Thus under (A.2) and (A.5) and by Lemma 4, we have

$$\begin{aligned}
N_n^{1/2} (\hat{\beta}_n(\hat{\theta}_n) - \beta_0) &= N_n^{-1/2} \left\{ N_n^{-1} \mathbf{X}_n' \mathbf{S}_n'(\hat{\theta}_n) \mathbf{S}_n(\hat{\theta}_n) \mathbf{X}_n \right\}^{-1} \mathbf{X}_n' \mathbf{S}_n'(\hat{\theta}_n) \mathbf{S}_n(\hat{\theta}_n) \mathbf{S}_{0n}^{-1} \nu_{0n} \\
&= N_n^{-1/2} \left(N_n^{-1} \mathbf{X}_n' \mathbf{S}_{0n}' \mathbf{S}_{0n} \mathbf{X}_n \right)^{-1} \mathbf{X}_n' \mathbf{S}_{0n}' \nu_{0n} + o_p(1) \\
&\xrightarrow{D} N \left(\mathbf{0}, \lim_{n \rightarrow \infty} \left(N_n^{-1} \mathbf{X}_n' \mathbf{S}_{0n}' \mathbf{S}_{0n} \mathbf{X}_n \right)^{-1} \right).
\end{aligned}$$

For

$$\hat{\sigma}_n^2(\hat{\theta}_n) = N_n^{-1} \mathbf{Y}_n' \mathbf{S}_n'(\hat{\theta}_n) \left[\mathbf{I}_n - \mathbf{S}_n(\hat{\theta}_n) \mathbf{X}_n \left\{ \mathbf{X}_n' \mathbf{S}_n'(\hat{\theta}_n) \mathbf{S}_n(\hat{\theta}_n) \mathbf{X}_n \right\}^{-1} \mathbf{X}_n' \mathbf{S}_n'(\hat{\theta}_n) \right] \mathbf{S}_n(\hat{\theta}_n) \mathbf{Y}_n,$$

we have under (A.1)–(A.5),

$$\begin{aligned}
& N_n^{1/2}(\hat{\sigma}_n^2(\hat{\theta}_n) - \sigma_0^2) \\
&= N_n^{-1/2} \left[\nu'_{0n} S_{0n}'^{-1} S'_n(\hat{\theta}_n) S_n(\hat{\theta}_n) S_{0n}^{-1} \nu_{0n} \right. \\
&\quad \left. - \nu'_{0n} S_{0n}'^{-1} S'_n(\hat{\theta}_n) S_n(\hat{\theta}_n) X_n \left\{ X'_n S'_n(\hat{\theta}_n) S_n(\hat{\theta}_n) X_n \right\}^{-1} X'_n S'_n(\hat{\theta}_n) S_n(\hat{\theta}_n) S_{0n}^{-1} \nu_{0n} - N_n \sigma_0^2 \right] \\
&= N_n^{-1/2} \left(\nu'_{0n} \nu_{0n} - N_n \sigma_0^2 \right) + N_n^{-1/2} \nu'_{0n} \left\{ \sum_{k=1}^q (\theta_{0k} - \hat{\theta}_{n,k}) S_{0n}'^{-1} W'_{n,k} \right\} \\
&\quad \left\{ \sum_{k=1}^q (\theta_{0k} - \hat{\theta}_{n,k}) W_{n,k} S_{0n}^{-1} \right\} \nu_{0n} + 2N_n^{-1/2} \sum_{k=1}^q (\theta_{0k} - \hat{\theta}_{n,k}) \nu'_{0n} S_{0n}'^{-1} W'_{n,k} \nu_{0n} \\
&\quad - N_n^{-1/2} \nu'_{0n} S_{0n}'^{-1} S'_n(\hat{\theta}_n) S_n(\hat{\theta}_n) X_n \left\{ X'_n S'_n(\hat{\theta}_n) S_n(\hat{\theta}_n) X_n \right\}^{-1} X'_n S'_n(\hat{\theta}_n) S_n(\hat{\theta}_n) S_{0n}^{-1} \nu_{0n} \\
&= N_n^{-1/2} \left(\nu'_{0n} \nu_{0n} - N_n \sigma_0^2 \right) + o_p(1) \\
&\xrightarrow{D} N(0, 2\sigma_0^4).
\end{aligned}$$

□

Appendix C: Proof of Theorem 3.

PROOF. For the MLE of β , we have

$$\begin{aligned}
\hat{\beta}_n(\theta) &= \{X'_n S'_n(\theta) S_n(\theta) X_n\}^{-1} X'_n S'_n(\theta) S_n(\theta) Y_n \\
&= \beta_0 + N_n^{-1/2} \left\{ N_n^{-1} X'_n S'_n(\theta) S_n(\theta) X_n \right\}^{-1} \left\{ N_n^{-1/2} X'_n S'_n(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \right\} \\
&= \beta_0 + o_p(1),
\end{aligned}$$

since under (A.1)–(A.5), $N_n^{-1/2} X'_n S'_n(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n}$ is of order $\mathcal{O}_p(1)$ and the elements of $\{N_n^{-1} X'_n S'_n(\theta) S_n(\theta) X_n\}^{-1}$ are uniformly bounded. Furthermore, under (A.1)–(A.5) and by Lemma 4, we have

$$\begin{aligned}
N_n^{1/2}(\hat{\beta}_n(\theta) - \beta_0) &= N_n^{-1/2} \left\{ N_n^{-1} X'_n S'_n(\theta) S_n(\theta) X_n \right\}^{-1} X'_n S'_n(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \\
&\xrightarrow{D} N(0, \Sigma_\beta),
\end{aligned}$$

where Σ_β is given in the statement of Theorem 3.

For the MLE of σ^2 , under (A.1)–(A.4), we have

$$\begin{aligned}
\hat{\sigma}_n^2(\theta) &= N_n^{-1} \left[\nu'_{0n} S_{0n}'^{-1} S'_n(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \right. \\
&\quad \left. - N_n^{-1/2} \nu'_{0n} S_{0n}'^{-1} S'_n(\theta) S_n(\theta) X_n (X'_n S'_n(\theta) S_n(\theta) X_n)^{-1} X'_n S'_n(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \right] \\
&= N_n^{-1} \nu'_{0n} \nu_{0n} + N_n^{-1} \nu'_{0n} \left\{ \sum_{k=1}^q (\theta_{0k} - \theta_k) S_{0n}'^{-1} W'_{n,k} \right\} \left\{ \sum_{k=1}^q (\theta_{0k} - \theta_k) W_{n,k} S_{0n}^{-1} \right\} \nu_{0n} \\
&\quad + (2/N_n) \sum_{k=1}^q (\theta_{0k} - \theta_k) \nu'_{0n} S_{0n}'^{-1} W'_{n,k} \nu_{0n} \\
&\quad - N_n^{-1/2} \nu'_{0n} S_{0n}'^{-1} S'_n(\theta) S_n(\theta) X_n (X'_n S'_n(\theta) S_n(\theta) X_n)^{-1} X'_n S'_n(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \xrightarrow{P} \sigma_0^2,
\end{aligned}$$

where the last three terms are of order $o_p(1)$ by the Chebyshev's inequality.

Furthermore, under (A.1)–(A.5),

$$\begin{aligned}
&N_n^{1/2} (\hat{\sigma}_n^2(\theta) - \sigma_0^2) \\
&= N_n^{-1/2} \left(\nu'_{0n} \nu_{0n} - N_n \sigma_0^2 \right) + N_n^{-1/2} \nu'_{0n} \left\{ \sum_{k=1}^q (\theta_{0k} - \theta_k) S_{0n}'^{-1} W'_{n,k} \right\} \\
&\quad \times \left\{ \sum_{k=1}^q (\theta_{0k} - \theta_k) W_{n,k} S_{0n}^{-1} \right\} \nu_{0n} + 2N_n^{-1/2} \sum_{k=1}^q (\theta_{0k} - \theta_k) \nu'_{0n} S_{0n}'^{-1} W'_{n,k} \nu_{0n} \\
&\quad - N_n^{-1/2} \nu'_{0n} S_{0n}'^{-1} S'_n(\theta) S_n(\theta) X_n (X'_n S'_n(\theta) S_n(\theta) X_n)^{-1} X'_n S'_n(\theta) S_n(\theta) S_{0n}^{-1} \nu_{0n} \\
&= N_n^{-1/2} \left(\nu'_{0n} \nu_{0n} - N_n \sigma_0^2 \right) + o_p(1) \xrightarrow{D} N(0, 2\sigma_0^4),
\end{aligned}$$

where the last three terms are of order $o_p(1)$ by the Chebyshev's inequality and the fact that $\lim_{n \rightarrow \infty} m_n N_n^{-1} = c \in (0, \infty]$.

Finally, we show the inconsistency of $\hat{\theta}_n$ by considering Example 2 under the infill asymptotics. We let the threshold value be such that all the cells on the lattice are neighbors of each other. Thus $q = 1$ and $\theta = \theta$. In this case,

$$W_{n,1} = \frac{1}{N_n - 1} (\mathbf{1}_n \mathbf{1}_n' - \mathbf{I}_n), \quad m_n = N_n - 1 = \mathcal{O}(N_n),$$

where $\mathbf{1}_n$ is an N_n -dimensional vector of all 1's. Thus $m_n \rightarrow \infty$ and $m_n/N_n \rightarrow 1 \neq 0$. It follows that

$$\begin{aligned}
S_n(\theta) &= \frac{N_n - 1 + \theta}{N_n - 1} \mathbf{I}_n - \frac{\theta}{N_n - 1} \mathbf{1}_n \mathbf{1}_n', \quad S_n^{-1}(\theta) = \frac{N_n - 1}{N_n - 1 + \theta} \left\{ \mathbf{I}_n + \frac{\theta}{1 - \theta} \frac{\mathbf{1}_n \mathbf{1}_n'}{N_n - 1} \right\} \\
W_{n,1} S_n^{-1}(\theta) &= \frac{1}{N_n - 1 + \theta} \left\{ (1 - \theta)^{-1} \mathbf{1}_n \mathbf{1}_n' - \mathbf{I}_n \right\}.
\end{aligned}$$

Consider the case $p = 1$, $\beta = \beta$, and $\mathbf{X} = \mathbf{1}_n$. The profile loglikelihood function is

$$\begin{aligned}\ell(\theta) &= \ell\{\hat{\beta}(\theta), \theta, \hat{\sigma}_n^2(\theta)\} = -N_n/2 - (N_n/2) \log \hat{\sigma}_n^2(\theta) + \log |\mathbf{S}_n(\theta)| \\ &= \text{const} - (N_n/2) \log \boldsymbol{\nu}'_{0n} \mathbf{B}_n(\theta) \boldsymbol{\nu}_{0n} + \log |\mathbf{S}_n(\theta)|\end{aligned}$$

where $\mathbf{G}_1 = \mathbf{W}_{n,1} \mathbf{S}_{0n}^{-1}$ and $\mathbf{B}_n(\theta) = \{\mathbf{I}_n - (\theta - \theta_0) \mathbf{G}_1\}(\mathbf{I}_n - N_n^{-1} \mathbf{1}_n \mathbf{1}_n')\{\mathbf{I}_n - (\theta - \theta_0) \mathbf{G}_1\}$. The first-order derivative of $\ell(\theta)$ is

$$\frac{\partial \ell(\theta)}{\partial \theta} = -(N_n/2) \{\boldsymbol{\nu}'_{0n} \mathbf{B}_n(\theta) \boldsymbol{\nu}_{0n}\}^{-1} \left\{ \boldsymbol{\nu}'_{0n} \frac{\partial \mathbf{B}_n(\theta)}{\partial \theta} \boldsymbol{\nu}_{0n} \right\} - \text{tr}\{\mathbf{W}_{n,1} \mathbf{S}_n^{-1}(\theta)\},$$

where

$$\frac{\partial \mathbf{B}_n(\theta)}{\partial \theta} = -2\mathbf{G}_1 + N_n^{-1} \mathbf{1}_n \mathbf{1}_n' \mathbf{G}_1 + 2(\theta - \theta_0) \mathbf{G}_1^2 + N_n^{-1} \mathbf{G}_1 \mathbf{1}_n \mathbf{1}_n' - 2(\theta - \theta_0) N_n^{-1} \mathbf{G}_1 \mathbf{1}_n \mathbf{1}_n' \mathbf{G}_1.$$

Thus, provided that $\theta_0 \neq 1$,

$$\begin{aligned}\frac{\partial \ell(\theta_0)}{\partial \theta} &= -N_n \left\{ \boldsymbol{\nu}'_{0n} (\mathbf{I}_n - N_n^{-1} \mathbf{1}_n \mathbf{1}_n') \boldsymbol{\nu}_{0n} \right\}^{-1} \left\{ (N_n - 1 + \theta_0)^{-1} \boldsymbol{\nu}'_{0n} (\mathbf{I}_n - N_n^{-1} \mathbf{1}_n \mathbf{1}_n') \boldsymbol{\nu}_{0n} \right\} \\ &\quad - (N_n \theta_0) \{ (N_n - 1 + \theta_0)(1 - \theta_0) \}^{-1} \\ &= -N_n \{ (N_n - 1 + \theta_0)(1 - \theta_0) \}^{-1} \rightarrow -(1 - \theta_0)^{-1}, \text{ as } n \rightarrow \infty.\end{aligned}$$

The second-order derivative of $\ell(\theta)$ is

$$\begin{aligned}\frac{\partial^2 \ell(\theta)}{\partial \theta^2} &= -(N_n^2 - N_n + N_n \theta^2) \{ (N_n - 1 + \theta)(1 - \theta) \}^{-2} + (N_n/2) \{ \boldsymbol{\nu}'_{0n} \mathbf{B}_n(\theta) \boldsymbol{\nu}_{0n} \}^{-2} \\ &\quad \times \left\{ \boldsymbol{\nu}'_{0n} \frac{\partial \mathbf{B}_n(\theta)}{\partial \theta} \boldsymbol{\nu}_{0n} \right\}^2 - (N_n/2) \{ \boldsymbol{\nu}'_{0n} \mathbf{B}_n(\theta) \boldsymbol{\nu}_{0n} \}^{-1} \left\{ \boldsymbol{\nu}'_{0n} \frac{\partial^2 \mathbf{B}_n(\theta)}{\partial \theta^2} \boldsymbol{\nu}_{0n} \right\} \\ &= -\frac{N_n^2 - N_n + N_n \theta^2}{(N_n - 1 + \theta)^2 (1 - \theta)^2} + \frac{2N_n}{(N_n - 1 + \theta)^2} - \frac{N_n}{(N_n - 1 + \theta)^2} \\ &\rightarrow -(1 - \theta)^{-2}, \text{ as } n \rightarrow \infty.\end{aligned}$$

We have

$$\hat{\theta}_n = \theta_0 - \left\{ \frac{\partial^2 \ell(\tilde{\theta}_n)}{\partial \theta^2} \right\}^{-1} \frac{\partial \ell(\theta_0)}{\partial \theta},$$

where $\tilde{\theta}_n = \alpha \hat{\theta}_n + (1 - \alpha) \theta_0$ for some $\alpha \in (0, 1)$. When $\hat{\theta}_n \xrightarrow{P} \theta_0$, $\tilde{\theta}_n \xrightarrow{P} \theta_0$. Thus

$$\hat{\theta}_n \xrightarrow{P} \theta_0 - \left\{ (1 - \theta_0)^{-2} \right\}^{-1} (1 - \theta_0)^{-1} = 2\theta_0 - 1 \neq \theta_0,$$

provided that $\theta_0 \neq 1$. That is, $\hat{\theta}_n$ is not a consistent estimator of θ . \square

Appendix D: Relation to a Conventional Asymptotic Framework.

The asymptotic framework we consider is unified covering all three types of asymptotics. This is in sharp contrast to Mardia and Marshall (1984). Even though it is often referenced as the theoretical backing of SAR model parameter estimation (see, e.g., Cressie (1993)), it is not clear exactly what asymptotic scheme is assumed and how different schemes affect the rates of convergence in Mardia and Marshall (1984). In particular, the rates of convergence are indistinguishable between increasing domain and hybrid asymptotics from the results of Mardia and Marshall (1984). The approach taken there also cannot deal with infill asymptotics in a straightforward manner. Thus, we believe that our approach may be more adequate for studying the asymptotics of SAR models. We clarify these issues here by taking apart the elements in the regularity conditions of Mardia and Marshall (1984) and providing insight into the connections and differences between the two approaches.

Recall the regularity conditions for the consistency and asymptotic normality of the MLE of parameters in Mardia and Marshall (1984). Let \mathbf{V} denote the variance-covariance matrix of \mathbf{Y}_n . Let $\mathbf{B}_\beta = [b_\beta^{ij}] = -E \{ \partial^2 \ell(\boldsymbol{\eta}) / \partial \beta \partial \beta' \}$ and $\mathbf{B}_\gamma = [b_\gamma^{ij}] = -E \{ \partial^2 \ell(\boldsymbol{\eta}) / \partial \gamma \partial \gamma' \}$, where $\boldsymbol{\gamma} = (\boldsymbol{\theta}', \sigma^2)'$. Let $\mathbf{V}_i = \partial \mathbf{V} / \partial \gamma_i$, $\mathbf{V}^i = \partial \mathbf{V}^{-1} / \partial \gamma_i$, $\mathbf{V}_{ij} = \partial^2 \mathbf{V} / \partial \gamma_i \partial \gamma_j$, and $\mathbf{V}^{ij} = \partial^2 \mathbf{V}^{-1} / \partial \gamma_i \partial \gamma_j$ for $i, j = 1, \dots, q+1$. To align with the notation in Mardia and Marshall (1984), we omit n in some of the notation. Consider the following assumptions.

- (B.1) $\lim_{n \rightarrow \infty} \mathbf{B}_\beta^{-1} = \mathbf{0}$ and $\lim_{n \rightarrow \infty} \mathbf{B}_\gamma^{-1} = \mathbf{0}$.
- (B.2) $\lim_{n \rightarrow \infty} \sum_{i,j,k,l=1}^{q+1} b_\gamma^{ki} b_\gamma^{lj} \text{tr}(\mathbf{V} \mathbf{V}^{kj} \mathbf{V} \mathbf{V}^{li}) = 0$.
- (B.3) $\lim_{n \rightarrow \infty} \sum_{i,k=1}^{q+1} \sum_{j,l=1}^p b_\gamma^{ki} b_\beta^{lj} (\mathbf{x}_j' \mathbf{V}^k \mathbf{V} \mathbf{V}^i \mathbf{x}_l) = 0$.

Mardia and Marshall (1984) established that under (B.1)–(B.3), the MLE of $\boldsymbol{\eta}$ is consistent and asymptotically normal.

We now investigate the relation between our regularity conditions and those in Mardia and Marshall (1984) under the SAR model. For ease of presentation, we restrict our attention to a single order of neighbors $q = 1$ and $\boldsymbol{\gamma} = (\theta_1, \sigma^2)'$. Our results below show that, if the assumptions of our Theorem 1 or 2 hold, then

the assumptions for Mardia and Marshall (1984)'s results are satisfied. However, the asymptotic results in Mardia and Marshall (1984) do not distinguish the rates of convergence as we do for increasing domain in Theorem 1 and hybrid asymptotics in Theorem 2. Furthermore, under the assumptions of our Theorem 3, the regularity conditions in Mardia and Marshall (1984) are not all satisfied.

Increasing Domain Asymptotics. Under the assumptions in Theorem 1, we show that (B.1)–(B.3) hold. It is obvious that, by (A.5), $\lim_{n \rightarrow \infty} \mathbf{B}_\beta^{-1} = \mathbf{0}$. It can be shown that \mathbf{B}_γ^{-1} is equal to

$$\left\{ N_n(2\sigma^4)^{-1} \text{tr}(\mathbf{G}'_1 \mathbf{G}_1 + \mathbf{G}_1^2) - (\sigma^4)^{-1} \text{tr}^2(\mathbf{G}_1) \right\}^{-1} \begin{bmatrix} N_n(2\sigma^4)^{-1} & -(\sigma^2)^{-1} \text{tr}\{\mathbf{G}_1\} \\ -(\sigma^2)^{-1} \text{tr}\{\mathbf{G}_1\} & \text{tr}\{\mathbf{G}'_1 \mathbf{G}_1 + \mathbf{G}_1^2\} \end{bmatrix} \quad (6.5)$$

of which the determinant term scaled by N_n^{-2} is bounded away from zero by the positive definiteness assumption of $\Sigma_{\eta_0}^{-1}$, where $\mathbf{G}_1 = \mathbf{W}_{n,1} \mathbf{S}_n^{-1}(\theta_1)$. Also, $\text{tr}(\mathbf{G}_1) = \mathcal{O}(N_n/m_n)$, $\text{tr}(\mathbf{G}'_1 \mathbf{G}_1) = \mathcal{O}(N_n/m_n)$, $\text{tr}(\mathbf{G}_1^2) = \mathcal{O}(N_n/m_n)$ under (A.1), (A.2) and (A.4), and $m_n = \mathcal{O}(1)$. It follows that $b_\gamma^{ij} = \mathcal{O}(N_n^{-1})$ for $i, j = 1, 2$ and $\lim_{n \rightarrow \infty} \mathbf{B}_\gamma^{-1} = \mathbf{0}$. Thus, (B.1) holds.

For (B.2), the inverse of the variance-covariance matrix is

$$\mathbf{V}^{-1} = (\sigma^2)^{-1} (\mathbf{I}_n - \theta_1 \mathbf{W}'_1) (\mathbf{I}_n - \theta_1 \mathbf{W}_1),$$

where, for ease of notation, we again suppress n in \mathbf{V}_n and $\mathbf{W}_{n,1}$. The first-order derivatives are

$$\begin{aligned} \mathbf{V}^1 &= \frac{\partial \mathbf{V}^{-1}}{\partial \theta_1} = -(\sigma^2)^{-1} \{ \mathbf{W}'_1 (\mathbf{I}_n - \theta_1 \mathbf{W}_1) + (\mathbf{I}_n - \theta_1 \mathbf{W}'_1) \mathbf{W}_1 \}, \\ \mathbf{V}^2 &= \frac{\partial \mathbf{V}^{-1}}{\partial \sigma^2} = -(\sigma^2)^{-2} (\mathbf{I}_n - \theta_1 \mathbf{W}'_1) (\mathbf{I}_n - \theta_1 \mathbf{W}_1). \end{aligned}$$

The second-order derivatives are

$$\begin{aligned} \mathbf{V}^{11} &= \frac{\partial^2 \mathbf{V}^{-1}}{\partial \theta_1^2} = 2(\sigma^2)^{-1} \mathbf{W}'_1 \mathbf{W}_1, \quad \mathbf{V}^{22} = \frac{\partial^2 \mathbf{V}^{-1}}{\partial (\sigma^2)^2} = 2(\sigma^2)^{-3} (\mathbf{I}_n - \theta_1 \mathbf{W}'_1) (\mathbf{I}_n - \theta_1 \mathbf{W}_1), \\ \mathbf{V}^{12} &= \mathbf{V}^{21} = \frac{\partial^2 \mathbf{V}^{-1}}{\partial \theta_1 \partial \sigma^2} = (\sigma^2)^{-2} \{ \mathbf{W}'_1 (\mathbf{I}_n - \theta_1 \mathbf{W}_1) + (\mathbf{I}_n - \theta_1 \mathbf{W}'_1) \mathbf{W}_1 \}. \end{aligned}$$

Under (A.1)–(A.4), we have $\text{tr}(\mathbf{V}\mathbf{V}^{kj}\mathbf{V}\mathbf{V}^{li}) = \mathcal{O}(N_n/m_n)$ for $i, j, k, l = 1, 2$.

Since $b_{\gamma}^{ij} = \mathcal{O}(N_n^{-1})$ for $i, j = 1, 2$, we have

$$\left| \sum_{i,j,k,l=1}^2 b_{\gamma}^{ki} b_{\gamma}^{lj} \text{tr}(\mathbf{V}\mathbf{V}^{kj}\mathbf{V}\mathbf{V}^{li}) \right| \leq \sum_{i,j,k,l=1}^2 |b_{\gamma}^{ki}| |b_{\gamma}^{lj}| |\text{tr}(\mathbf{V}\mathbf{V}^{kj}\mathbf{V}\mathbf{V}^{li})| = \mathcal{O}(N_n^{-1}) = o(1).$$

Thus (B.2) holds.

For (B.3), we have

$$\begin{aligned} \sum_{i,k=1}^2 \sum_{j,l=1}^p b_{\gamma}^{ki} b_{\beta}^{lj} (\mathbf{x}'_j \mathbf{V}^k \mathbf{V} \mathbf{V}^i \mathbf{x}_l) &= \sum_{i,k=1}^2 b_{\gamma}^{ki} N_n^{-1} \text{tr} \left\{ (N_n^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} (\mathbf{X}' \mathbf{V}^k \mathbf{V} \mathbf{V}^i \mathbf{X}) \right\} \\ &= \sum_{i,k=1}^2 b_{\gamma}^{ki} \mathcal{O}(1) \end{aligned}$$

under (A.2) and (A.5). Since $b_{\gamma}^{ik} = \mathcal{O}(N_n^{-1})$ for $i, k = 1, 2$, we have

$$\sum_{i,k=1}^2 \sum_{j,l=1}^p b_{\gamma}^{ki} b_{\beta}^{lj} (\mathbf{x}'_j \mathbf{V}^k \mathbf{V} \mathbf{V}^i \mathbf{x}_l) = \mathcal{O}(N_n^{-1}) = o(1).$$

Thus (B.3) holds.

Hybrid Asymptotics. Under the assumptions in Theorem 2, we show that (B.1)–(B.3) hold. With arguments similar to those for the increasing domain asymptotics, we have $\lim_{n \rightarrow \infty} \mathbf{B}_{\beta}^{-1} = \mathbf{0}$. For \mathbf{B}_{γ}^{-1} , the determinant term in (6.5) multiplied by m_n/N_n^2 is bounded away from 0 by the positive definiteness assumption of $\Sigma_{\theta_0}^{-1}$. Also, $\text{tr}(\mathbf{G}_1) = \mathcal{O}(N_n/m_n)$, $\text{tr}(\mathbf{G}'\mathbf{G}_1) = \mathcal{O}(N_n/m_n)$, $\text{tr}(\mathbf{G}_1^2) = \mathcal{O}(N_n/m_n)$ under (A.1), (A.2) and (A.4), $m_n \rightarrow \infty$, and $m_n N_n^{-1} = o(1)$. It follows that $b_{\gamma}^{11} = \mathcal{O}(m_n/N_n)$, $b_{\gamma}^{12} = b_{\gamma}^{21} = \mathcal{O}(N_n^{-1})$, and $b_{\gamma}^{22} = \mathcal{O}(N_n^{-1})$. Thus $\lim_{n \rightarrow \infty} \mathbf{B}_{\gamma}^{-1} = \mathbf{0}$ and (B.1) holds.

Using arguments similar to those for the increasing domain asymptotics, we can show that (B.2) and (B.3) hold since $b_{\gamma}^{11} = \mathcal{O}(m_n/N_n)$, $b_{\gamma}^{12} = b_{\gamma}^{21} = \mathcal{O}(N_n^{-1})$, and $b_{\gamma}^{22} = \mathcal{O}(N_n^{-1})$. However, as mentioned above, Mardia and Marshall (1984) do not distinguish the rates of convergence between increasing domain and hybrid asymptotics, but we do in Theorems 1 and 2.

Infill Asymptotics. Under the infill asymptotics, we continue to use the example in Appendix C to illustrate inconsistency of the MLE. In this example,

$\mathbf{G}_1 = (N_n - 1 + \theta_1)^{-1} \{(1 - \theta_1)^{-1} \mathbf{1}_n \mathbf{1}_n' - \mathbf{I}_n\} = \mathbf{G}_1'$. Since

$$\lim_{n \rightarrow \infty} N_n (2\sigma^4)^{-1} \left[N_n \sigma^{-4} \text{tr}\{\mathbf{G}_1^2\} - \sigma^{-4} \text{tr}^2\{\mathbf{G}_1\} \right]^{-1} = (1 - \theta_1)^2 / 2 \neq 0$$

provided that $\theta_1 \neq 1$, we attain $\lim_{n \rightarrow \infty} \mathbf{B}_\gamma^{-1} \neq \mathbf{0}$, where \mathbf{B}_γ^{-1} is given by (6.5).

Thus (B.1) does not hold.

Appendix E: Preparation Lemmas. Let $\mathbf{A}_n = [a_n^{i,j}]_{i,j=1}^{N_n}$ and $\mathbf{B}_n = [b_n^{i,j}]_{i,j=1}^{N_n}$ denote $N_n \times N_n$ matrices. $\boldsymbol{\nu}_n \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

LEMMA 1. *If \mathbf{A}_n and \mathbf{B}_n are uniformly bounded in both matrix norms $||| \cdot |||_1$ and $||| \cdot |||_\infty$, then $\mathbf{A}_n \mathbf{B}_n$ and $\mathbf{B}_n \mathbf{A}_n$ are also uniformly bounded in both matrix norms.*

PROOF. Let $\mathbf{C}_n = [c_n^{i,j}]_{i,j=1}^{N_n} = \mathbf{A}_n \mathbf{B}_n$. Then $\sum_{j=1}^{N_n} c_n^{1,j} = \sum_{j=1}^{N_n} \sum_{i=1}^{N_n} a_n^{1,i} b_n^{i,j} = \sum_{i=1}^{N_n} a_n^{1,i} \sum_{j=1}^{N_n} b_n^{i,j} \leq b_r \sum_{i=1}^{N_n} a_n^{1,i} \leq a_r b_r < \infty$ for some constants $a_r, b_r > 0$. \square

LEMMA 2. *Suppose that the elements $a_n^{i,j}$ of \mathbf{A}_n are $\mathcal{O}(1/m_n)$ uniformly over i, j . If \mathbf{B}_n is uniformly bounded in matrix norm $||| \cdot |||_\infty$ (respectively, $||| \cdot |||_1$), then the elements of $\mathbf{A}_n \mathbf{B}_n$ (respectively, $\mathbf{B}_n \mathbf{A}_n$) are $\mathcal{O}(1/m_n)$ uniformly over i, j , in which case $\text{tr}(\mathbf{A}_n \mathbf{B}_n) = \text{tr}(\mathbf{B}_n \mathbf{A}_n) = \mathcal{O}(N_n/m_n)$.*

PROOF. Let $\mathbf{C}_n = [c_n^{i,j}]_{i,j=1}^{N_n} = \mathbf{A}_n \mathbf{B}_n$. Then $c_n^{i,j} = \sum_{l=1}^{N_n} a_n^{i,l} b_n^{l,j} = \mathcal{O}(1/m_n) \sum_{l=1}^{N_n} b_n^{l,j} = \mathcal{O}(1/m_n)$. \square

LEMMA 3. *Suppose that \mathbf{A}_n is uniformly bounded in ℓ_∞ , $\text{tr}(\mathbf{A}_n) = \mathcal{O}(1)$, $\text{tr}(\mathbf{A}_n \mathbf{A}_n') = \mathcal{O}(1)$ and $\text{tr}(\mathbf{A}_n^2) = \mathcal{O}(1)$. Let $\{a_n\}$ denote a sequence of constants such that $a_n = o(1)$. Then we have $a_n \boldsymbol{\nu}_n' \mathbf{A}_n \boldsymbol{\nu}_n = o_p(1)$.*

PROOF. We have $E(\boldsymbol{\nu}_n' \mathbf{A}_n \boldsymbol{\nu}_n) = o_p(1)$ and $\text{Var}(\boldsymbol{\nu}_n' \mathbf{A}_n \boldsymbol{\nu}_n) = o_p(1)$. By the Chebyshev's inequality, $a_n \boldsymbol{\nu}_n' \mathbf{A}_n \boldsymbol{\nu}_n = a_n E(\boldsymbol{\nu}_n' \mathbf{A}_n \boldsymbol{\nu}_n) + o_p(1) = o_p(1)$. \square

LEMMA 4. *Suppose that \mathbf{A}_n is uniformly bounded in matrix norm $||| \cdot |||_\infty$ and \mathbf{Z}_n is an $N_n \times k$ matrix uniformly bounded in ℓ_∞ . Then $N_n^{-1/2} \mathbf{Z}_n' \mathbf{A}_n \boldsymbol{\nu}_n = \mathcal{O}_p(1)$. Further, if the limit of $\mathbf{Z}_n' \mathbf{A}_n \mathbf{A}_n' \mathbf{Z}_n / N_n$ exists and is positive definite, then*

$$N_n^{-1/2} \mathbf{Z}_n' \mathbf{A}_n \boldsymbol{\nu}_n \xrightarrow{D} N(\mathbf{0}, \sigma^2 \lim_{n \rightarrow \infty} \mathbf{Z}_n' \mathbf{A}_n \mathbf{A}_n' \mathbf{Z}_n / N_n).$$

PROOF. Let $\mathbf{u} = \mathbf{Z}'_n \mathbf{A}_n \boldsymbol{\nu}_n$ with $u_l = \sum_{j=1}^{N_n} (\sum_{i=1}^{N_n} z_n^{i,l} a_n^{i,j}) \nu_{n,j}$. Then

$$\begin{aligned}
& P(|N_n^{-1/2} u_l| > C_\epsilon) \\
&= P\left(N_n^{-1/2} \left| \sum_{j=1}^{N_n} \left(\sum_{i=1}^{N_n} z_n^{i,l} a_n^{i,j} \right) \nu_{n,j} \right| > C_\epsilon\right) = P\left(N_n^{-1} \left\{ \sum_{j=1}^{N_n} \left(\sum_{i=1}^{N_n} z_n^{i,l} a_n^{i,j} \right) \nu_{n,j} \right\}^2 > C_\epsilon^2\right) \\
&\leq P\left(\left\{ N_n^{-1} \sum_{j=1}^{N_n} \left(\sum_{i=1}^{N_n} z_n^{i,l} a_n^{i,j} \right)^2 \right\} \left(\sum_{j=1}^{N_n} \nu_{n,j}^2 \right) > C_\epsilon^2\right) \leq P\left(z^2 a^2 \left(\sum_{j=1}^{N_n} \nu_{n,j}^2 \right) > C_\epsilon^2\right) \\
&= P(\chi_n^2 > C),
\end{aligned}$$

where χ_n^2 is a χ^2 variable on N_n degrees of freedom, z is the upper bound of the elements of \mathbf{Z}_n and a is the upper bound of matrix norm $\|\cdot\|_\infty$ of \mathbf{A}_n . Thus $N_n^{-1/2} \mathbf{Z}'_n \mathbf{A}_n \boldsymbol{\nu}_n = \mathcal{O}_p(1)$. Further, $N_n^{-1/2} \mathbf{u} = N_n^{-1/2} \mathbf{Z}'_n \mathbf{A}_n \boldsymbol{\nu}_n \sim N(0, N_n^{-1} \sigma^2 (\mathbf{Z}'_n \mathbf{A}_n \mathbf{A}'_n \mathbf{Z}_n))$. It is obvious that if the limit of $\mathbf{Z}'_n \mathbf{A}_n \mathbf{A}'_n \mathbf{Z}_n / N_n$ exists and is positive definite, then

$$N_n^{-1/2} \mathbf{Z}'_n \mathbf{A}_n \boldsymbol{\nu}_n \xrightarrow{D} N(0, \sigma^2 \lim_{n \rightarrow \infty} \mathbf{Z}'_n \mathbf{A}_n \mathbf{A}'_n \mathbf{Z}_n / N_n).$$

□

LEMMA 5. Suppose that \mathbf{A}_n is uniformly bounded either in matrix norm $\|\cdot\|_\infty$ or $\|\cdot\|_1$ and the elements $a_n^{i,j}$ are $\mathcal{O}(1/m_n)$ uniformly over i, j . Then $E(\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n) = \mathcal{O}(N_n/m_n)$ and $\text{Var}(\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n) = \mathcal{O}(N_n/m_n)$. If $\lim_{n \rightarrow \infty} (m_n/N_n) = 0$, then

$$m_n N_n^{-1} \{\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n - E(\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n)\} = o_p(1).$$

PROOF. First, we note that $E(\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n) = \sigma^2 \text{tr}(\mathbf{A}_n) = \mathcal{O}(N_n/m_n)$ and $\text{Var}(\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n) = \sigma^4 \{\text{tr}(\mathbf{A}_n \mathbf{A}'_n) + \text{tr}(\mathbf{A}_n^2)\}$. Since $\text{tr}(\mathbf{A}_n \mathbf{A}'_n) = \text{tr}(\mathbf{A}_n^2) = \mathcal{O}(N_n/m_n)$, we have $\text{Var}(\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n) = \mathcal{O}(N_n/m_n)$. By the Chebyshev's inequality, when $\lim_{n \rightarrow \infty} m_n N_n^{-1} = 0$, we have

$$m_n N_n^{-1} \{\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n - E(\boldsymbol{\nu}'_n \mathbf{A}_n \boldsymbol{\nu}_n)\} = \mathcal{O}_p(1) (m_n/N_n)^{1/2} = o_p(1).$$

□

LEMMA 6. Suppose that the matrix \mathbf{A}_n is a non-negative $N_n \times N_n$ matrix with elements $a_n^{i,j} = \nu_n^{i,j} / \sum_{j=1}^{N_n} \nu_n^{i,j}$ and $\nu_n^{i,j} \geq 0$ for all i, j . If $\sum_{j=1}^{N_n} \nu_n^{i,j}$ for all i, j are uniformly bounded away from zero at the order of m_n and $\sum_{i=1}^{N_n} \nu_n^{i,j}$ for

all i, j are $\mathcal{O}(m_n)$, then \mathbf{A}_n is uniformly bounded in matrix norms $||| \cdot |||_\infty$ and $||| \cdot |||_1$.

PROOF. Observe that $|||\mathbf{A}_n|||_\infty = 1$. Furthermore, $\sum_{j=1}^{N_n} \nu_n^{i,j} \geq c_1 m_n$ and $\sum_{i=1}^{N_n} \nu_n^{i,j} \leq c_2 m_n$ for some constants $c_1, c_2 \geq 0$. Thus $\sum_{i=1}^{N_n} \left(\nu_n^{i,j} / \sum_{j=1}^{N_n} \nu_n^{i,j} \right) \leq \sum_{i=1}^{N_n} \nu_n^{i,j} / (c_1 m_n) \leq c_2 / c_1$, which implies that $|||\mathbf{A}_n|||_1$ is uniformly bounded. \square

LEMMA 7. Suppose that $||| \sum_{k=1}^q a_k \mathbf{A}_{n,k} ||| \leq c$ uniformly for a constant $0 < c < 1$, where $\mathbf{A}_{n,k}$, $k = 1, \dots, q$ are $N_n \times N_n$ matrices and $||| \cdot |||$ is a matrix norm. Then $|||\mathbf{S}_n^{-1}|||$ is uniformly bounded, where $\mathbf{S}_n = \mathbf{I}_n - \sum_{k=1}^q a_k \mathbf{A}_{n,k}$.

PROOF. Let $\mathbf{A}_n = \sum_{k=1}^q a_k \mathbf{A}_{n,k} / ||| \sum_{k=1}^q a_k \mathbf{A}_{n,k} |||$ and $\lambda_n = ||| \sum_{k=1}^q a_k \mathbf{A}_{n,k} |||$. Then $|||\mathbf{S}_n^{-1}||| \leq \sum_{i=0}^{\infty} |||\lambda_n \mathbf{A}_n|||^i = \sum_{i=0}^{\infty} \lambda_n^i = 1/(1 - \lambda_n) \leq 1/(1 - c) < \infty$ (Corollary 5.6.16, Horn and Johnson (1985)). \square

Acknowledgments. We thank the anonymous reviewers and the editor for their constructive comments that helped improve the paper. Funding has been provided for this research from a USDA Cooperative State Research, Education and Extension Service (CSREES) Hatch project.

References.

- ANSELIN, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.
- BANDYOPADHYAY, S. and MAITY, A. (2011). Analysis of Sabine river flow data using semiparametric spline modeling. *Journal of Hydrology* In press.
- BIVAND, S., PEBESMA, E. and GOMEZ-RUBIO, V. (2008). *Applied Spatial Data Analysis with R*. Wiley, New York.
- CRESSIE, N. (1993). *Statistics for Spatial Data*, revised ed. Wiley, New York.
- DU, J., ZHANG, H. and MANDREKAR, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Annals of Statistics* **37** 3330-3361.
- GAETAN, C. and GUYON, X. (2010). *Spatial Statistics and Modelling*. Springer, New York.
- HORN, R. A. and JOHNSON, C. R. (1985). *Matrix Analysis*. Cambridge University Press.
- KELEJIAN, H. H. and PRUCHA, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* **104** 217-257.
- KELLY, J. L. (1955). *General Topology*. Van Nostrand.
- LAHIRI, S. N. (1996). On inconsistency of estimators based on spatial data under infill asymptotics. *Sankhya Series A* **58** 403-417.

- LAHIRI, S. N. (2003). Central limit theorems for weighted sums under some stochastic and fixed spatial sampling designs. *Sankhya Series A* **65** 356–388.
- LAHIRI, S. N., LEE, Y. and CRESSIE, N. (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning and Inference* **103** 65–85.
- LEE, L.-F. (2004). Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models (with supplements). *Econometrica* **72** 1899–1925.
- LINDGREN, F., LINDSTRÖ, J. and RUE, H. (2010). An explicit link between Gaussian fields and Gaussian Markov random fields: the SPDE approach. Technical Report, Norwegian University of Science and Technology, Norway.
- LOH, W.-L. (2005). Fixed-domain asymptotics for a subclass of Matern-type Gaussian random fields. *Annals of Statistics* **33** 2344–2394.
- MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146.
- MATSUDA, Y. and YAJIMA, Y. (2009). Fourier analysis of irregularly spaced data on \mathbb{R}^d . *Journal of the Royal Statistical Society Series B* **71** 191–217.
- NEWBY, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica* **59** 1161–1167.
- ROBINSON, P. M. (2010). The efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics* In press.
- ROBINSON, P. M. and THAWORNKAIWONG, S. (2010). Statistical inference on regression with spatial dependence. Under submission.
- RUE, H. and TJELMELAND, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics* **29** 31–49.
- SCHABENBERGER, O. and GOTWAY, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman and Hall, Boca Raton.
- WALLER, L. and GOTWAY, C. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley, Hoboken.
- WHITE, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.
- WHITTLE, P. (1954). On Stationary processes in the plane. *Biometrika* **41** 431–449.
- YING, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Annals of Statistics* **21** 1567–1590.
- ZHANG, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association* **99** 250–261.
- ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936.

TABLE 1

Means and standard deviations (SD) in parentheses of maximum likelihood estimates (MLE) of the model parameters on a 4×4 lattice with varying sub-lattice sizes 1×1 , 2×2 , and 4×4 within each cell of the lattice, based on 100 simulated data.

Parameter	Truth	MLE	Sub-lattice size		
			1×1	2×2	4×4
β_0	2.0	Mean	1.9951	2.0030	1.9923
		SD	(0.3156)	(0.1568)	(0.0755)
β_1	2.0	Mean	2.0025	2.0048	2.0026
		SD	(0.3766)	(0.1729)	(0.0781)
θ_1	0.2	Mean	0.0691	0.0746	0.0705
		SD	(0.3002)	(0.2717)	(0.2701)
σ^2	1.0	Mean	0.7875	0.9442	0.9983
		SD	(0.3083)	(0.1828)	(0.0938)

TABLE 2

Means and standard deviations (SD) in parentheses of maximum likelihood estimates (MLE) of the model parameters on a 8×8 lattice with varying sub-lattice sizes 1×1 , 2×2 , and 4×4 within each cell of the lattice, based on 100 simulated data.

Parameter	Truth	MLE	Sub-lattice size		
			1×1	2×2	4×4
β_0	2.0	Mean	1.9938	2.006	1.9902
		SD	(0.1568)	(0.0851)	(0.0350)
β_1	2.0	Mean	1.9754	2.0053	1.9988
		SD	(0.1734)	(0.0923)	(0.0437)
θ_1	0.2	Mean	0.1426	0.1635	0.1318
		SD	(0.1462)	(0.1465)	(0.1402)
σ^2	1.0	Mean	0.9331	0.9895	1.004
		SD	(0.1680)	(0.0851)	(0.0423)

ZHU, J., HUANG, H.-C. and REYES, P. (2010). On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society Series B* **72** 389–402.

DEPARTMENT OF STATISTICS

UNIVERSITY OF KENTUCKY

LEXINGTON, KY 40506

USA

E-MAIL: yanbing.zheng@uky.edu

DEPARTMENT OF STATISTICS AND DEPARTMENT OF ENTOMOLOGY

UNIVERSITY OF WISCONSIN

MADISON, WI 53706

USA

E-MAIL: jzhu@stat.wisc.edu

TABLE 3

Means and standard deviations (SD) in parentheses of maximum likelihood estimates (MLE) of the model parameters on a 16×16 lattice with varying sub-lattice sizes 1×1 , 2×2 , and 4×4 within each cell of the lattice, based on 100 simulated data.

Parameter	Truth	MLE	Sub-lattice size		
			1×1	2×2	4×4
β_0	2.0	Mean	2.0061	1.9968	2.0013
		SD	(0.0661)	(0.0357)	(0.0179)
β_1	2.0	Mean	2.0089	2.0055	1.9997
		SD	(0.0810)	(0.0498)	(0.0210)
θ_1	0.2	Mean	0.2092	0.1913	0.1880
		SD	(0.0749)	(0.0721)	(0.0732)
σ^2	1.0	Mean	0.9680	0.9992	1.0054
		SD	(0.0955)	(0.0431)	(0.0211)