Empirical likelihood for nonparametric regression models with spatial autoregressive errors

Yinghua Li, Yongsong Qin & Yuan Li

Journal of the Korean Statistical Society

ISSN 1226-3192
Volume 50
Number 2

J. Korean Stat. Soc. (2021) 50:447-478
DOI 10.1007/s42952-020-00088-z

Springer

Journal of the Korean Statistical Society (2021) 50:447–478
https://doi.org/10.1007/s42952-020-00088-z

Online ISSN 2005-2863
Print ISSN 1226-3192

RESEARCH ARTICLE

Empirical likelihood for nonparametric regression models with spatial autoregressive errors

Yinghua Li[1,2] · Yongsong Qin[2] · Yuan Li[1]

Received: 6 December 2019 / Accepted: 22 September 2020 / Published online: 10 October 2020

Abstract
In this paper, we propose to use the empirical likelihood (EL) method to construct confidence regions for nonparametric regression models with spatial autoregressive errors. It is shown that the EL statistics for the related parameters asymptotically have chi-squared distributions, which are used to construct confidence regions for the parameters. Results from simulation study and real data analysis are also presented.

Keywords Nonparametric regression · Spatial autoregressive error · Empirical likelihood · Confidence region

AMS Subject Classification Primary 62G05 · secondary 62E20

1 Introduction

We firstly outline the introduction of the empirical likelihood (EL) method and its applications in dealing with some types of dependent data (other than spatial data). The EL method as a nonparametric method is an important approach in constructing confidence intervals, introduced by Owen (1988, 1990, 1991, 2001), which can be robust under various distributional assumptions but may still have good properties analogous to the parametric likelihood method. This method has only been used to deal with independent observations for a considerable time after it was introduced. To deal with dependent data, Kitamura (1997) first proposed the blockwise EL (BEL) method to construct confidence intervals for parameters with mixing samples. Chen and Wong (2009) developed BEL method to obtain confidence

✉ Yongsong Qin
ysqin@gxnu.edu.cn

1 School of Economics and Statistics, Guangzhou University, Guangzhou 510006, Guangdong, China

2 School of Mathematics and Statistics, Guangxi Normal University, Guilin 541004, Guangxi, China

intervals for quantiles with mixing samples. For time series, Mykland (1995) made the connection between the dual likelihood and the EL and applied the EL approach to models with martingale structures. Chuang and Chan (2002) introduced the EL method to the autoregressive (AR) models where the disturbances form a martingale difference sequence. Chan and Ling (2006) developed the EL for regular generalized autoregressive conditional heteroskedasticity (GARCH) models.

In this part, we summarize the progress of the generalized method of moments (GMM), quasi maximum likelihood (QML) method and nonparametric method (other than EL method) used in parametric/nonparametric spatial models. Spatial data arise in a variety of fields, including econometrics, epidemiology, environmental science, image analysis, oceanography and many others. Many spatial econometric models were inspired by questions arising in regional science and economic geography, where the units of observations are geographically determined and the structure of the dependence among these units is related to location and distance. Among spatial econometric models, the mixed regressive, spatial autoregressive (MRSAR) model introduced by Cliff and Ord (1973) is one of the most important models. The development in testing and estimation of MRSAR models based on the GMM and QML methods can be found in Anselin (1988), Cressie (1993), Anselin and Bera (1998), Lee (2004, (2007) and Jin and Lee (2018), among others. MRSAR models are among parametric spatial models. In many applications, data may not be fitted well by parametric models. There also exist some developments on the statistical inference for nonparametric spatial regression models to address this problem. For example, for nonparametric spatial regression models with spatial autoregressive errors, Wang et al. (2016) studied the asymptotic normality of the estimators of the models. Lu et al. (2014) considered the estimation of semi-parametric varying-coefficient quantile regression with spatial data under mixing condition. Su and Jin (2010) proposed a QML estimator for partially linear spatial autoregressive models. Hallin et al. (2001) obtained the asymptotic normality of kernel density estimation in a spatial linear process. Hallin et al. (2004a, b) proposed local linear method to estimate the spatial regression function and established the asymptotic normality of the estimators of the regression function and its derivatives under mild regularity assumptions, respectively.

We now state the progress of the EL and BEL methods used in parametric/nonparametric spatial models. Jin and Lee (2019) studied the generalized EL (GEL) estimation and tests of parametric spatial models. Qin (2019) independently investigated the EL method for MRSAR models. Nordman (2008), Nordman and Daniel (2008) and Bandyopadhyay et al. (2015) used the BEL method to nonparametric spatial models.

It is worthwhile to look at another class of models related to the spatial models mentioned above: spatial conditionally autoregressive (CAR) models (e.g., Besag et al. 1991; Banerjee et al. 2014). The spatial CAR state that the conditional mean of observation at a (spatial) site i given all other observations is a linear function of all observations (except for the observation at site i), where $i = 1, 2, \ldots, n$ and n is the total sites. On the other hand, the spatial autoregressive (SAR) models mentioned in this paper emphasize that the mean of observation at a (spatial) site i is a linear function of the means of all observations (except for the observation at site i). In some cases, CAR models and SAR models agree. But in most cases, they are different. For more details, refer to Ord (1975).

In this paper, we propose to use the EL method to construct confidence region for nonparametric spatial regression models, where we do not need to use the BEL method to avoid the choice problem of block size in BEL method. It is shown that the EL statistics for the related parameters asymptotically have chi-squared distributions, which are used to construct confidence regions for the parameters. To the best of our knowledge, the EL for nonparametric spatial regression models has not appeared in the literature yet. This work focuses on inference when we have continuous response values. The EL method may be extended to the binary or count data sets as in the spatial generalized linear mixed models (e.g., Diggle and Tawn 1998), which is left for our future study. Our asymptotic results are based on a growing observation region (increasing-domain). Zhang (2004) found that under fixed domain, covariance parameters are not consistently estimated in the model-based geostatistics. No general results are available for nonparametric spatial regression models under fixed domain.

The article is organized as follows. Section 2 gives the main results. Results from a simulation study are reported in Sect. 3. Section 4 presents the analysis of real data. Some concluding remarks are given in Sect. 5. All the technical details are presented in Sect. 6.

2 Main results

We consider the following nonparametric regression model (1) and (2) with spatial autoregressive errors:

$$y_i = m(x_i) + R_i, \quad 1 \le i \le n, \tag{1}$$

$$R = \rho W_n R + \epsilon_{(n)}, \tag{2}$$

where n is the number of spatial units, $m(\cdot)$ is a unknown smooth function, $\{y_i\}$ are scalar responses and $\{x_i\} \in [0,1]^r (r \ge 1)$ are fixed design points, $R = (R_1, \dots, R_n)^T$ is an $n \times 1$ vector of errors (disturbances), ρ is the scalar autoregressive parameter with $|\rho| < 1$, W_n is an $n \times n$ spatial weighting matrix of constants, $\epsilon_{(n)}$ is an $n \times 1$ vector of innovations which satisfies

$$E\epsilon_{(n)} = 0, Var(\epsilon_{(n)}) = \sigma^2 I_n.$$

Remark 1 The spatial weighting matrix W_n has a zero diagonal and is usually taken as row-normalized matrix (i.e. the L_2 norm of its every row is unity). In addition, the elements of W_n are all non-negative. W_n may be symmetric, but needs not be so. By the spectral radius theorem for the non-negative row-normalized matrix, the conventional space $(-1, 1)$ of ρ is a subset of the interval $(1/\lambda_{min}(W_n), 1/\lambda_{max}(W_n))$, where $\lambda_{min}(W_n)$ and $\lambda_{max}(W_n)$ denote the minimum and maximum eigenvalues of W_n, respectively. For more details, refer to Ord (1975).

For given $x \in (0,1)^r$, a commonly used type of estimator of $m(x)$ in model (1) is

$$\hat{m}_n(x) = \sum_{i=1}^{n} \widetilde{W}_{ni}(x) y_i, \tag{3}$$

where $\widetilde{W}_{ni}(u), i = 1, \dots, n$ are non-negative weighted functions. Take $0 < h = h_n \to 0$ and a non-negative kernel function $K(u), u \in R^r$. Let $K_h(u) = K(u/h)$. In this article, $\widetilde{W}_{ni}(u)$ is chosen as the Nadaraya–Watson weight

$$\widetilde{W}_{ni}(u) = \frac{K_h(u - x_i)}{\sum_{j=1}^{n} K_h(u - x_j)}.$$

From Eq. (1), we know that $R_i = y_i - m(x_i)$ and thus its estimator is

$$\hat{R}_i = y_i - \hat{m}_n(x_i). \tag{4}$$

Denote $\hat{R} = (\hat{R}_1, \dots, \hat{R}_n)^T$ and $\hat{\epsilon}_{(n)} = \hat{R} - \rho W_n \hat{R}$. Based on model (2), we adopt the quasi-maximum likelihood method (QMLE) to estimate ρ and σ^2. Let $A_n(\rho) = I_n - \rho W_n$ and suppose that $A_n(\rho)$ is nonsingular, then the log-likelihood function is

$$L(\rho, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 + \log|A_n(\rho)| - \frac{1}{2\sigma^2}\hat{\epsilon}_{(n)}^T \hat{\epsilon}_{(n)}, \tag{5}$$

where $\hat{\epsilon}_{(n)} = A_n(\rho)\hat{R}$. For the log-likelihood function (5), given ρ, the QMLE of σ^2 is

$$\hat{\sigma}_n^2(\rho) = \frac{1}{n}\hat{\epsilon}_{(n)}^T \hat{\epsilon}_{(n)} = \frac{1}{n}\hat{R}^T A_n^T(\rho) A_n(\rho) \hat{R}. \tag{6}$$

The concentrated log-likelihood function of ρ is

$$L(\rho) = -\frac{n}{2}(\log(2\pi) + 1) - \frac{n}{2}\log\hat{\sigma}_n^2(\rho) + \log|A_n(\rho)|. \tag{7}$$

The QMLE $\hat{\rho}_n$ of ρ maximizes the concentrated likelihood (7) and the QMLE of σ^2 is $\hat{\sigma}_n^2(\hat{\rho}_n)$. We note that the QMLEs $\hat{\rho}_n$ and $\hat{\sigma}_n^2(\hat{\rho}_n)$ of ρ and σ^2 are also the solutions of the estimating Eqs. (8) and (9) below.

We are now in the position to present the EL statistics for $\theta \hat{=} (\rho, \sigma^2)^T$ and $\vartheta = m(x)$ for given $x \in (0,1)^r$, respectively.

2.1 EL for θ

In order to derive the EL statistic of ρ and σ^2, we observe that

$$\partial L(\rho,\sigma^2)/\partial\rho = \frac{1}{\sigma^2}\{\hat{\epsilon}_{(n)}^T W_n A_n^{-1}(\rho)\hat{\epsilon}_{(n)} - \sigma^2 tr(W_n A_n^{-1}(\rho))\}$$
$$= \frac{1}{\sigma^2}\{\hat{\epsilon}_{(n)}^T \tilde{G}_n \hat{\epsilon}_{(n)} - \sigma^2 tr(\tilde{G}_n)\},$$
$$\partial L(\rho,\sigma^2)/\partial\sigma^2 = \frac{1}{2\sigma^4}(\hat{\epsilon}_{(n)}^T\hat{\epsilon}_{(n)} - n\sigma^2),$$

where $G_n = W_n A_n^{-1}(\rho)$ and $\tilde{G}_n = \frac{1}{2}(G_n + G_n^T)$. Letting above derivatives be 0, we obtain the following estimating equations:

$$\hat{\epsilon}_{(n)}^T \tilde{G}_n \hat{\epsilon}_{(n)} - \sigma^2 tr(\tilde{G}_n) = 0, \tag{8}$$

$$\hat{\epsilon}_{(n)}^T \hat{\epsilon}_{(n)} - n\sigma^2 = 0. \tag{9}$$

We use $\tilde{g}_{ij}$ to denote the (i,j) element the matrix $\tilde{G}_n$, and adapt the convention that any sum with an upper index of less than one is zero. To deal with the quadratic form in (8), we follow Kelejian and Prucha (2001) to introduce a martingale difference array. Define the σ-fields: $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_i = \sigma(\epsilon_1, \epsilon_2, \dots, \epsilon_i), 1 \le i \le n$. Let

$$\tilde{Y}_{in} = \tilde{g}_{ii}(\hat{\epsilon}_i^2 - \sigma^2) + 2\hat{\epsilon}_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\hat{\epsilon}_j, \tag{10}$$

where $\hat{\epsilon}_i$ is the i-th component of $\hat{\epsilon}_{(n)} = A_n(\rho)\hat{R}$. Then $\mathcal{F}_{i-1} \subseteq \mathcal{F}_i$, and if $\{\hat{\epsilon}_i\}$ are replaced by $\{\epsilon_i\}$, $\{\tilde{Y}_{in}, \mathcal{F}_i, 1 \le i \le n\}$ form a martingale difference array. In other words, $\{\tilde{Y}_{in}, \mathcal{F}_i, 1 \le i \le n\}$ asymptotically form a martingale difference array, and

$$\hat{\epsilon}_{(n)}^T \tilde{G}_n \hat{\epsilon}_{(n)} - \sigma^2 tr(\tilde{G}_n) = \sum_{i=1}^n \tilde{Y}_{in}. \tag{11}$$

Based on (8)–(11), we propose the following EL ratio statistic for θ:

$$L_{1n}(\theta) = \sup_{p_i, 1\le i\le n} \prod_{i=1}^n (np_i),$$

where $\{p_i\}$ satisfy

$$\sum_{i=1}^n p_i \left\{ \tilde{g}_{ii}(\hat{\epsilon}_i^2 - \sigma^2) + 2\hat{\epsilon}_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\hat{\epsilon}_j \right\} = 0,$$
$$\sum_{i=1}^n p_i(\hat{\epsilon}_i^2 - \sigma^2) = 0,$$

Let

$$\omega_i(\theta) = \begin{pmatrix} \tilde{g}_{ii}(\hat{\epsilon}_i^2 - \sigma^2) + 2\hat{\epsilon}_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\hat{\epsilon}_j \\ \hat{\epsilon}_i^2 - \sigma^2 \end{pmatrix}.$$

Following Owen (1990), one can show that

$$\ell_{1n}(\theta) \hat{=} -2\log L_{1n}(\theta) = 2\sum_{i=1}^{n} \log\{1 + \lambda_1^T(\theta)\omega_i(\theta)\}, \tag{12}$$

where $\lambda_1(\theta) \in R^2$ is the solution of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\omega_i(\theta)}{1 + \lambda_1^T(\theta)\omega_i(\theta)} = 0. \tag{13}$$

2.2 EL for $\vartheta = m(x)$ for given $x \in (0,1)^r$

Rewrite Eqs. (1) and (2) as

$$y_i = m(x_i) + \rho \sum_{j=1}^{n} w_{ij} R_j + \epsilon_i, \tag{14}$$

where $W_n = (w_1, \dots, w_n)^T$, $w_i = (w_{ij})_{n\times 1}$ is a vector. w_{ij} places different emphases on different sites, site j 'close' to site i means that R_j exerts more influence on R_i than those far away and then w_{ij} is large. Conventionally, we assume that $w_{ii} = 0, 1 \le i \le n$.

Similar to Lei and Qin (2011), let

$$Z_{in}(\vartheta) = K_h(x - x_i)(\tilde{y}_i - \vartheta), \tag{15}$$

where $\tilde{y}_i = y_i - \hat{\rho}_n \sum_{j=1}^{n} w_{ij}\hat{R}_j$. We define the log-empirical likelihood ratio for ϑ as

$$\ell_{2n}(\vartheta) \hat{=} -2\max \sum_{i=1}^{n} \log(np_i),$$

where the maximum is taken over all sets of nonnegative numbers $p_1, \dots, p_n$ summing to 1 and such that $\sum_{i=1}^{n} p_i Z_{in}(\vartheta) = 0$. By the Lagrange multiplier method, that

$$\ell_{2n}(\vartheta) = 2\sum_{i=1}^{n} \log(1 + \lambda_2(\vartheta) Z_{in}(\vartheta)), \tag{16}$$

where $\lambda_2(\vartheta)$ is the solution of the equation

$$\frac{1}{n}\sum_{i=1}^{n} \frac{Z_{in}(\vartheta)}{1 + \lambda_2(\vartheta) Z_{in}(\vartheta)} = 0. \tag{17}$$

We use ρ_0 to denote the true value of ρ. Let $\mu_j = E\epsilon_1^j, j = 3, 4$. Use $Vec(diagA)$ to denote the vector formed by the diagonal elements of a matrix A and $\|a\|$ to denote the L_2-norm of a vector a. To obtain the asymptotical distributions of $\ell_{1n}(\theta)$ and $\ell_{2n}(\vartheta)$, we need following assumptions.

A1. $m(\cdot)$ satisfies the Lipschitz condition on $[0,1]^r$, that is, for any $x_1, x_2 \in [0,1]^r$, there exists a constant C_0 such that

$$|m(x_1) - m(x_2)| \le C_0||x_1 - x_2||.$$

A2. $\{\epsilon_i, 1 \le i \le n\}$ are independent and identically distributed random variables with mean 0, variance $\sigma^2 > 0$ and $E|\epsilon_1|^{4+\eta_1} < \infty$ for some $\eta_1 > 0$.

A3. The non-negative kernel function K is bounded with bounded support which satisfies $\sum_{i=1}^n K_h(x - x_i) = O(nh^r)$. Weight functions $\{\widetilde{W}_{ni}(x), 1 \le i \le n\}$ are non-negative, and as $n \to \infty$, there exist constants $M > 0, 0 < \alpha_1 < 1/r$, such that

(i) $\sum_{i=1}^n \widetilde{W}_{ni}(x) = 1$;
(ii) $\sum_{i=1}^n \widetilde{W}_{ni}(x) I(||x - x_i|| > Mn^{-\alpha_1}) = 0$;
(iii) $\max_{1\le i\le n} \widetilde{W}_{ni}(x) = O(n^{-(1-\alpha_1 r)})$;

A4. Let W_n, $A_n^{-1}(\rho)$ and $\{x_i\}$ be as described above. They satisfy the following conditions:

(i) The row and column sums of W_n and $A_n^{-1}(\rho_0)$ are uniformly bounded in absolute value;
(ii) $A_n(\rho)$ are uniformly bounded in either row or column sums, uniformly ρ in a compact space Λ. The true ρ_0 is in the interior of Λ.
(iii) $\{x_i\}$ are uniformly bounded.

A5. (i) The elements w_{ij} of W_n are at most of order γ_n^{-1}, uniformly in all i, j, where the rate sequence γ_n is a divergent, the ratio $\gamma_n^{1+\eta}/n \to 0$ for some $\eta > 0$, and satisfies

$$\lim_{n\to\infty}\left(\frac{\gamma_n}{n}\ln|\sigma^2 A_n^{-1}(\rho_0)(A_n^{-1}(\rho_0))^T| - \frac{\gamma_n}{n}\ln|\sigma_n^{*2}(\rho)A_n^{-1}(\rho)(A_n^{-1}(\rho))^T|\right) \ne 0,$$

whenever $\rho \ne \rho_0$, where $\sigma_n^{*2}(\rho) = \sigma^2/\{ntr[(A_n^{-1}(\rho_0))^T A_n^T(\rho)A_n(\rho)A_n^{-1}(\rho_0)]\}$.

(ii) As $n \to \infty$, $n^{1-2\alpha_1}h^r \to 0$, $n^{\alpha_1 r}h^r \to 0$ and $nh^r/\gamma_n \to 0$.

A6. There is a constants $c_j > 0, j = 1, 2$, such that $0 < c_1 \le \lambda_{min}\left(n^{-1}\Sigma\right) \le \lambda_{max}\left(n^{-1}\Sigma\right) \le c_2 < \infty$, where $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the minimum and maximum eigenvalues of a matrix A, respectively,

$$\Sigma = \Sigma^T = Cov\left\{\sum_{i=1}^n \omega_i(\theta)\right\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$
$$\Sigma_{11} = 2\sigma^4 tr(\tilde{G}_n^2) + (\mu_4 - 3\sigma^4)||Vec(diag\tilde{G}_n)||^2, \tag{18}$$
$$\Sigma_{12} = 2\sigma^4 tr(\tilde{G}_n) + (\mu_4 - 3\sigma^4)tr(\tilde{G}_n),$$
$$\Sigma_{22} = 2n\sigma^4 + (\mu_4 - 3\sigma^4)n.$$

Remark 2 Conditions A1-A3 are commonly used conditions for nonparametric regression models (e.g. Lei and Qin 2011). Conditions A4 and A6 are common assumptions for SAR models. For example, A4 and A6 are used in Assumptions 4, 5 and 6 in Lee (2004), the analog of $0 < c_1 \leq \lambda_{min}\left(n^{-1}\Sigma\right)$ (e.g. $n^{-1}\sigma_Q^2 \geq c$ for some constant $c > 0$ in the proof of Lemma 2 in this article) is employed in the assumption of Theorem 1 in Kelejian and Prucha (2001). From Condition A4, one can see that $\lambda_{max}\left(n^{-1}\Sigma\right) \leq c_2 < \infty$. For convenience, we list this consequence of A4 and A6 as a condition here. Condition A5(i) is used to show the asymptotic normality of $\hat{\rho}_n$, which is also used in Wang et al. (2016). Condition A5(ii) is used to derive the asymptotic distribution of $\ell_{2n}(\vartheta)$.

We now state the main results.

Theorem 1 *Suppose that Assumptions* (A1)–(A6) *are satisfied. Then under model* (2), *as* $n \to \infty$,

$$\ell_{1n}(\theta) \xrightarrow{d} \chi_2^2,$$

where χ_2^2 *is a chi-squared distributed random variable with two degrees of freedom.*

Let $\chi_\alpha^2(k)$ satisfy $P(\chi_k^2 \leq \chi_\alpha^2(k)) = \alpha$ for $0 < \alpha < 1$. It follows from Theorem 1 that an EL based confidence region for θ with asymptotically correct coverage probability α can be constructed as

$$\{\theta : \ell_{1n}(\theta) \leq \chi_\alpha^2(2)\}.$$

The following is the the asymptotical distribution result of the log-empirical likelihood ratio on $\vartheta = m(x)$.

Theorem 2 *Suppose that Assumptions* (A1)–(A6) *are satisfied. Then as* $n \to \infty$,

$$\ell_{2n}(\vartheta) \xrightarrow{d} \chi_1^2,$$

where χ_1^2 *is a chi-squared distributed random variable with one degree of freedom.*

From this result, the EL based confidence interval for ϑ with asymptotically correct coverage probability α can be constructed as

$$\{\vartheta : \ell_{2n}(\vartheta) \leq \chi_\alpha^2(1)\}.$$

3 Simulations

In this section, we carry out some simulations to show how our EL method performs. The models are considered as

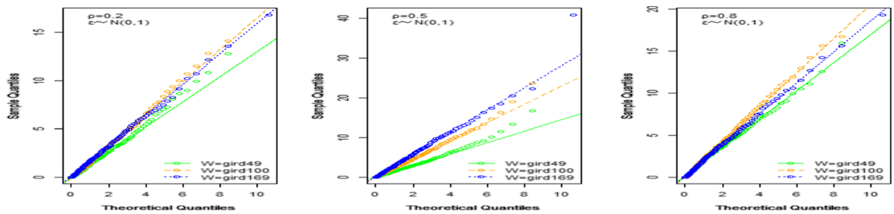
$$\begin{aligned} y_i &= sin(x_i) + R_i, \quad 1 \leq i \leq n, \\ R &= \rho W_n R + \epsilon_{(n)}, \text{ for } \rho = 0.2, 0.5, 0.8, \end{aligned}$$

where $x_i = \frac{i}{n+1}$, $1 \le i \le n$, and $\epsilon_i' s$ are taken from $N(0, 1)$, $t(5)$ and $\chi_4^2 - 4$, respectively. For the contiguity weight matrix $W_n = (W_{ij})$, we choose $W_{ij} = 1$ if spatial units i and j are neighbours by queen contiguity rule (namely, they share common border or vertex), $W_{ij} = 0$ otherwise (Anselin 1988, p. 18). A transformation is often used in applications to convert the matrix W_n to the unity of row-sums. We used the standardized version of W_n in our simulations, namely W_{ij} was replaced by $W_{ij} / \sum_{j=1}^n W_{ij}$. In the estimation of the nonparametric regression function, we let $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$, use Nadaraya–Watson weight as

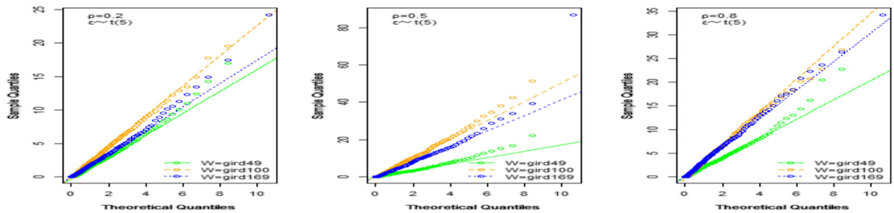
$$\widetilde{W}_{ni}(u) = \frac{K_h(u - x_i)}{\sum_{j=1}^n K_h(u - x_j)}$$

and select the h by the cross-validation (CV) method stated in Wang et al. (2016).

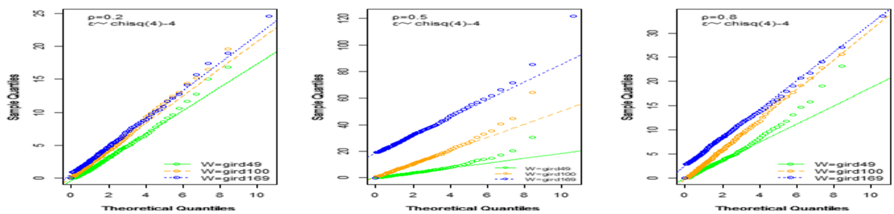
We consider three ideal cases of spatial units: $n = m \times m$ regular grid with $m = 7, 10, 13$, denoting W_n as $grid_{49}, grid_{100}$ and $grid_{169}$, respectively. We generated 1, 000 samples and compared the sample quantiles of $\ell_{1n}(\theta)$ with the quantiles of the χ_2^2 presented in Fig. 1. At the same time, for $x = 0.2, 0.5$ and 0.8, we compared the sample quantiles of $\ell_{2n}(\vartheta)$ with the quantiles of the χ_1^2 shown in Figs. 2, 3 and 4. We note that the x-axis and y-axis have different scales in Figs. 1, 2, 3 and 4,



$\epsilon \sim N(0, 1), \rho = 0.2, 0.5, 0.8$ and $\sigma^2 = 1$



$\epsilon \sim t(5), \rho = 0.2, 0.5, 0.8$ and $\sigma^2 = 5/3$



$\epsilon \sim \chi_4^2 - 4, \rho = 0.2, 0.5, 0.8$ and $\sigma^2 = 8$

Fig. 1 Q–Q plots of $\ell_{1n}(\theta)$ and χ_2^2

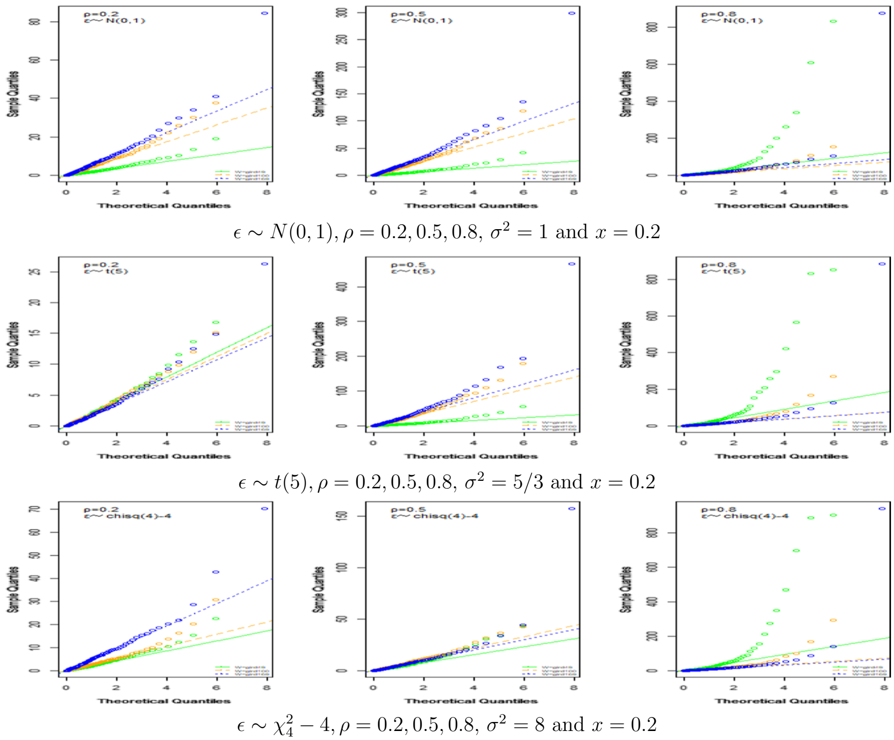
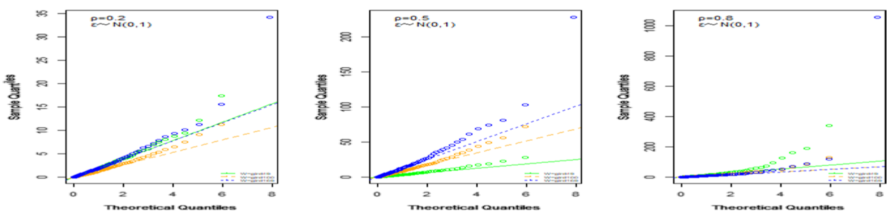


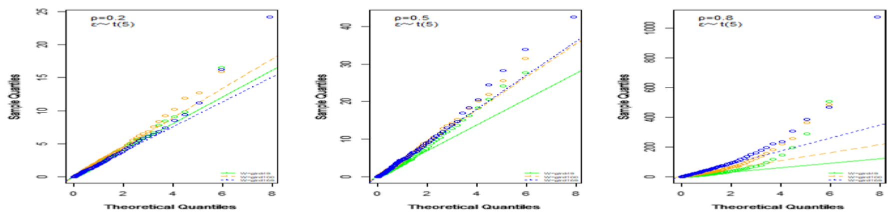
Fig. 2 Q–Q plots of $\ell_{2n}(\vartheta)$ and χ_1^2 with $x = 0.2$

which come from the R system. In this way, the empirical quantiles of the EL statistics and its theoretical quantiles are not approximately located on a 45 degree straight line. They can be modified to have the same scale. However, if they are modified to have the same scale, the figures can not show complete quantiles and look imperfect. Based on the simulation results, when $\epsilon_{(n)}$ is normally distributed, all sample distributions and the theoretic distributions agree well. We can also see, when $\epsilon_{(n)}$ is not normally distributed, that all the sample distributions fit the theoretic distributions well as $|\rho|$ is less than 0.8. All the sample distributions also fit the theoretic distributions well as $|\rho|$ is large when the sample sizes are large.

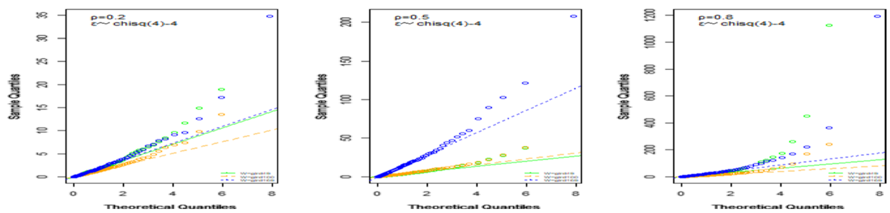
In addition, using the same models and the simulated samples as above and taking the nominal level $\alpha = 0.95$, we conducted a small simulation study to compare the finite sample performances of the confidence intervals for the spatial correlation coefficient ρ based on the EL method proposed in this paper and the normal approximation (NA) method in Wang et al. (2016), where $n = m \times m$ with $m = 7, 10, 13, 16, 20, 30$ and 40, respectively. The coverage probabilities (CP) and the average lengths (AL) of the confidence intervals for ρ in 1000 simulations were shown in Tables 1, 2 and 3. From the simulation results, we can see that the CP of the EL based confidence intervals converge to the nominal level $\alpha = 0.95$



$\epsilon \sim N(0,1), \rho = 0.2, 0.5, 0.8, \sigma^2 = 1$ and $x = 0.5$



$\epsilon \sim t(5), \rho = 0.2, 0.5, 0.8, \sigma^2 = 5/3$ and $x = 0.5$



$\epsilon \sim \chi^2_4 - 4, \rho = 0.2, 0.5, 0.8, \sigma^2 = 8$ and $x = 0.5$

Fig. 3 Q–Q plots of $\ell_{2n}(\vartheta)$ and χ^2_1 with $x = 0.5$

as n is large enough, whether the error term $\epsilon_{(n)}$ is normally distributed or not. On the other hand, the CP of the NA based confidence intervals also go closer to α as n is large enough, but the performance of the EL method is better compared with the NA method in terms of CP and AL.

4 Real data analysis

In order to illustrate the proposed procedures in Sect. 2, the real data analysis for two examples are presented here.

Example 1 The data come from 49 contiguous Planning Neighborhoods in Columbus, Ohio (e.g. Table 12.1 in Anselin 1988, p. 189). The data set contains crime variable (y) (the combined total of residential burglaries and vehicle thefts per thousand households in the neighborhood), income (x) (in thousand dollars). We considered fitting the data via the following model:

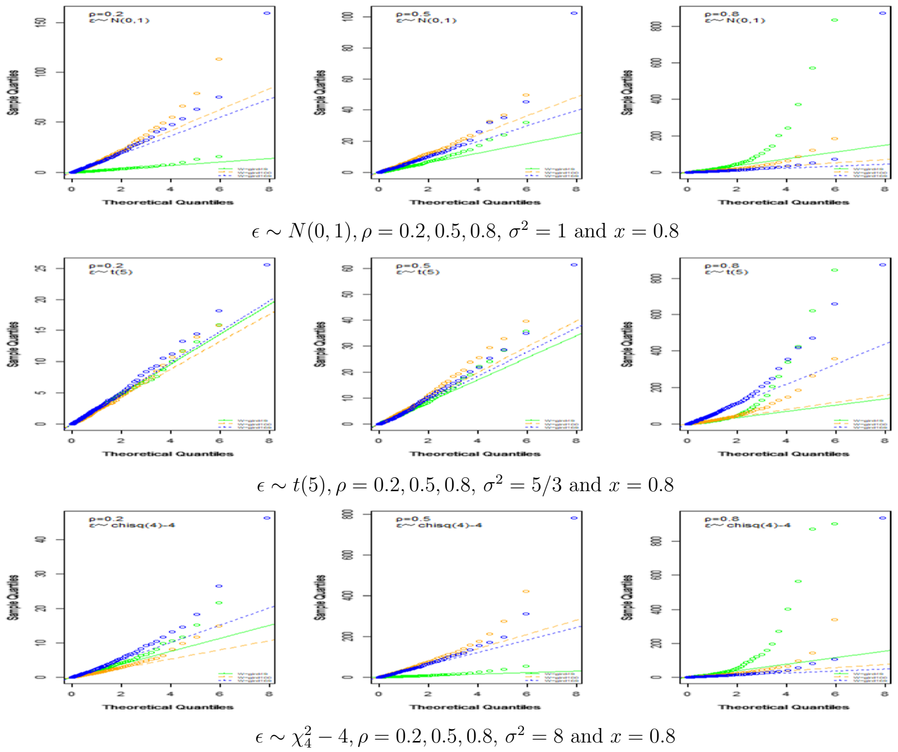


Fig. 4 Q–Q plots of $\ell_{2n}(\vartheta)$ and χ^2_1 with $x = 0.8$

$$
\begin{aligned}
Y_n &= m(X_n) + R, \\
R &= \rho W_n R + \epsilon_{(n)},
\end{aligned}
$$

where $n = 49, Y_n = (y_1, y_2, \dots, y_n)^T, X_n = (x_1, x_2, \dots, x_n)^T, 1 \le i \le n,\ E\epsilon_{(n)} = 0, Var(\epsilon_{(n)}) = \sigma^2 I_n$, the spatial weighting matrix W_n and $\{\tilde{W}_{ni}, 1 \le i \le n\}$ were selected by the method in Sect. 4.

We first used the quasi-maximum likelihood method to estimate ρ (i.e. assume the $\epsilon_{(n)}$ is normally distributed) and local linear fitting method to estimate $m(x)$, respectively, then separately employed the EL method in Section 2 and the NA method in Wang et al. (2016) to obtain the confidence intervals for parameters ρ and $m(x)$ at three selected points $x = 0.3, 0.5$ and 0.8 with confidence level 0.95, which were shown in Table 4.

From Table 4, we may give the following conclusions. The estimator of the spatial parameter is $\rho = 0.6227$ with its confidence interval not containing 0 which indicates there exists a substantial spatial relationship among the disturbances. The estimators of nonparametric regression function $m(x)$ at three selected points $x = 0.3, 0.5$ and 0.8 are 38.6698, 28.0025 and 21.3531, respectively. The results also illustrate that the lengthes of the EL based intervals are uniformly shorter than those

Table 1 Coverage probabilities (CP) and average lengthes (AL) of the EL and NA confidence intervals for ρ with $\epsilon \sim N(0,1)$

	W_n	CP		AL	
		EL	NA	EL	NA
$\rho = 0.2$	$grid_{49}$	0.885	0.905	0.9358	0.9361
	$grid_{100}$	0.894	0.905	0.6719	0.7573
	$grid_{169}$	0.921	0.915	0.5231	0.5398
	$grid_{256}$	0.921	0.934	0.4287	0.4686
	$grid_{400}$	0.938	0.928	0.3456	0.3832
	$grid_{900}$	0.932	0.930	0.2293	0.2337
	$grid_{1600}$	0.940	0.934	0.1635	0.1762
$\rho = 0.5$	$grid_{49}$	0.846	0.757	0.6653	0.6708
	$grid_{100}$	0.891	0.827	0.4327	0.4940
	$grid_{169}$	0.907	0.857	0.3454	0.3909
	$grid_{256}$	0.912	0.885	0.2717	0.3233
	$grid_{400}$	0.917	0.894	0.2392	0.2628
	$grid_{900}$	0.938	0.916	0.1790	0.1841
	$grid_{1600}$	0.932	0.922	0.1257	0.1370
$\rho = 0.8$	$grid_{49}$	0.756	0.616	0.2998	0.3626
	$grid_{100}$	0.797	0.710	0.2623	0.2725
	$grid_{169}$	0.838	0.750	0.2072	0.2190
	$grid_{256}$	0.866	0.778	0.1582	0.1827
	$grid_{400}$	0.886	0.826	0.1326	0.1500
	$grid_{900}$	0.886	0.868	0.1033	0.1079
	$grid_{1600}$	0.918	0.914	0.0717	0.0818

of the NA based intervals, which may be explained as that the EL based method performs better than the NA based method in this case.

Example 2 The data come from 288 prefecture-level cities in China, collected from National Bureau of Statistics of China and Anjuke. The data set contains the logarithm of housing price (y) and the logarithm of income per household (x) in the year of 2016. The model and estimation method are the same as Example 1. The only difference is that the sample size is $n = 288$ in this example while the sample size is $n = 49$ in Example 1. The results of the real data analysis for this example were illustrated in Table 5. From these results, one can obtain similar conclusions as Example 1.

Table 2 Coverage probabilities (CP) and average lengthes (AL) of the EL and NA confidence intervals for ρ with $\epsilon \sim t(5)$

	n	CP		AL	
		EL	NA	EL	NA
$\rho = 0.2$	$grid_{49}$	0.853	0.846	0.9370	0.9770
	$grid_{100}$	0.896	0.884	0.6712	0.6909
	$grid_{169}$	0.908	0.902	0.5231	0.5372
	$grid_{256}$	0.912	0.905	0.4277	0.4609
	$grid_{400}$	0.915	0.908	0.3377	0.3843
	$grid_{900}$	0.924	0.918	0.2200	0.2318
	$grid_{1600}$	0.934	0.933	0.1623	0.1754
$\rho = 0.5$	$grid_{49}$	0.803	0.752	0.6134	0.7933
	$grid_{100}$	0.865	0.814	0.4504	0.5478
	$grid_{169}$	0.880	0.839	0.3553	0.4212
	$grid_{256}$	0.883	0.855	0.2508	0.3438
	$grid_{400}$	0.896	0.877	0.2413	0.2743
	$grid_{900}$	0.916	0.906	0.1790	0.1851
	$grid_{1600}$	0.928	0.922	0.1200	0.1379
$\rho = 0.8$	$grid_{49}$	0.699	0.640	0.2888	0.3366
	$grid_{100}$	0.797	0.690	0.2482	0.2520
	$grid_{169}$	0.808	0.712	0.2287	0.2597
	$grid_{256}$	0.831	0.763	0.1721	0.2079
	$grid_{400}$	0.876	0.818	0.1256	0.1494
	$grid_{900}$	0.905	0.825	0.1000	0.1088
	$grid_{1600}$	0.930	0.855	0.0537	0.0788

5 Concluding remarks

For nonparametric regression models with spatial autoregressive errors, we have studied the construction of EL confidence regions for the parameters and nonparametric regression function in these models. A spatial nonparametric model is an alternative choice to fit spatial data when parametric models can not fit the data well. Wang et al. (2016) studied the QMLE of parameters and nonparametric regression function as well as the construction of NA based confidence regions for parameters and nonparametric regression function in a spatial nonparametric model. Our simulation results show that the EL confidence regions perform better than the NA confidence regions when the model errors are not normally distributed. In other words, the EL method provides a competitive choice to construct confidence regions for a spatial nonparametric model.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (11671102, 11731015, 12061017), the Natural Science Foundation of Guangxi (2017GXNSFAA198349) and the Program on the High Level Innovation Team and Outstanding Scholars in Universities of Guangxi Province. The authors are thankful to the referees for constructive suggestions.

Table 3 Coverage probabilities (CP) and average lengthes (AL) of the EL and NA confidence intervals for ρ with $\epsilon \sim \chi_4^2 - 4$

	n	CP		AL	
		EL	NA	EL	NA
$\rho = 0.2$	$grid_{49}$	0.854	0.855	0.9236	0.9447
	$grid_{100}$	0.883	0.884	0.6748	0.6993
	$grid_{169}$	0.896	0.892	0.4485	0.5212
	$grid_{256}$	0.922	0.905	0.4166	0.4280
	$grid_{400}$	0.922	0.915	0.3453	0.3983
	$grid_{900}$	0.936	0.928	0.2280	0.2337
	$grid_{1600}$	0.938	0.932	0.1700	0.1747
$\rho = 0.5$	$grid_{49}$	0.804	0.758	1.0210	1.0965
	$grid_{100}$	0.838	0.794	0.8194	0.8440
	$grid_{169}$	0.881	0.843	0.3652	0.4230
	$grid_{256}$	0.887	0.857	0.2757	0.3428
	$grid_{400}$	0.895	0.867	0.2547	0.2753
	$grid_{900}$	0.902	0.890	0.1729	0.1823
	$grid_{1600}$	0.914	0.898	0.1320	0.1376
$\rho = 0.8$	$grid_{49}$	0.725	0.641	0.2891	0.3516
	$grid_{100}$	0.786	0.700	0.2519	0.2597
	$grid_{169}$	0.833	0.725	0.2328	0.2609
	$grid_{256}$	0.855	0.779	0.1782	0.2097
	$grid_{400}$	0.879	0.827	0.1404	0.1664
	$grid_{900}$	0.884	0.858	0.1020	0.1078
	$grid_{1600}$	0.912	0.892	0.0620	0.0788

Table 4 Analysis results for the neighborhood crime data (with ALs shown in brackets)

Variable	Estimation	Confidence interval (EL)	Confidence interval (NA)
ρ	0.6227	[0.3877, 0.8007] (0.4130)	[0.3857, 0.8597] (0.4740)
$m(0.3)$	38.6698	[34.5354, 41.1134] (6.5780)	[35.3138, 42.0257] (6.7119)
$m(0.5)$	28.0025	[27.3952, 33.8954] (6.5002)	[21.4444, 30.9929] (9.5485)
$m(0.8)$	21.3531	[18.7309, 22.2467] (3.5158)	[18.7906, 23.9156] (5.1250)

Table 5 Analysis results for the housing price data (with ALs shown in brackets)

Variable	Estimation	Confidence interval (EL)	Confidence interval (NA)
ρ	0.4201	[0.4011, 0.5555] (0.1544)	[0.2826, 0.5577] (0.2751)
$m(0.3)$	8.4242	[8.3768, 8.4832] (0.1064)	[8.2692, 8.5791] (0.3099)
$m(0.5)$	8.6687	[8.5384, 8.7345] (0.1961)	[8.5830, 8.7855] (0.2025)
$m(0.8)$	9.1760	[9.1590, 9.3005](0.1415)	[9.0877, 9.2643] (0.1766)

Appendix

To prove the main results, we need some lemmas.

Lemma 1 *Let $\eta_1, \eta_2, \ldots, \eta_n$ be a sequence of stationary random variables, with $E|\eta_1|^s < \infty$ for some constants $s > 0$ and $C > 0$. Then*

$$\max_{1 \le i \le n} |\eta_i| = o(n^{1/s}), \quad a.s.$$

Proof It is straightforward. □

Lemma 2 *Suppose that Assumptions (A1)–(A6) are satisfied. Then as $n \to \infty$,*

$$Z_n = \max_{1 \le i \le n} ||\omega_i(\theta)|| = o_p(n^{2/(4+\eta_1)}) \ \ a.s., \tag{19}$$

$$\Sigma^{-1/2} \sum_{i=1}^{n} \omega_i(\theta) \xrightarrow{d} N(0, I_2), \tag{20}$$

$$n^{-1} \sum_{i=1}^{n} \omega_i(\theta)\omega_i^T(\theta) = n^{-1}\Sigma + o_p(1), \tag{21}$$

$$\sum_{i=1}^{n} ||\omega_i(\theta)||^3 = O_p(n), \tag{22}$$

where Σ is given in (18).

Proof For convenience, denote

$$m(X_n) = (m(x_1), \ldots, m(x_n))^T, \hat{m}_n(X_n) = (\hat{m}_n(x_1), \ldots, \hat{m}_n(x_n))^T.$$

Note that

$$\begin{aligned} Z_n &\le \max_{1 \le i \le n} \left| \tilde{g}_{ii}(\hat{\epsilon}_i^2 - \sigma^2) + 2\hat{\epsilon}_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\hat{\epsilon}_j \right| + \max_{1 \le i \le n} |\hat{\epsilon}_i^2 - \sigma^2| \\ &\le \max_{1 \le i \le n} |\tilde{g}_{ii}(\hat{\epsilon}_i^2 - \sigma^2)| + \max_{1 \le i \le n} \left| 2\hat{\epsilon}_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\hat{\epsilon}_j \right| + \max_{1 \le i \le n} |\hat{\epsilon}_i^2 - \sigma^2|, \end{aligned}$$

and

$$\begin{aligned}
\hat{\epsilon}_{(n)} &= A_n(\rho)\hat{R} = A_n(\rho)\{Y_n - \hat{m}_n(X_n)\} \\
&= A_n(\rho)\{Y_n - m(X_n) + m(X_n) - \hat{m}_n(X_n)\} \\
&= A_n(\rho)\{R + m(X_n) - \hat{m}_n(X_n)\}, \\
&= \epsilon_{(n)} + A_n(\rho)\{m(X_n) - \hat{m}_n(X_n)\}, \\
&= \epsilon_{(n)} + A_n(\rho)\begin{pmatrix} m(x_1) - \hat{m}_n(x_1) \\ \vdots \\ m(x_n) - \hat{m}_n(x_n) \end{pmatrix}
\end{aligned} \tag{23}$$

and

$$\begin{aligned}
m(x_j) - \hat{m}_n(x_j) &= m(x_j) - \sum_{i=1}^{n} \widetilde{W}_{ni}(x_j) y_i \\
&= \sum_{i=1}^{n} \widetilde{W}_{ni}(x_j)\{m(x_j) - m(x_i)\} - \sum_{i=1}^{n} \widetilde{W}_{ni}(x_j) R_i \\
&= O(n^{-\alpha_1}) - (\widetilde{W}_{n1}(x_j), \dots, \widetilde{W}_{nn}(x_j))R \\
&= O(n^{-\alpha_1}) - \sqrt{Var((\widetilde{W}_{n1}(x_j), \dots, \widetilde{W}_{nn}(x_j))R)}\, O_p(1) \\
&= O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2}).
\end{aligned} \tag{24}$$

Thus,

$$\hat{\epsilon}_{(n)} = \epsilon_{(n)} + O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2}).$$

By Conditions A2 and A4(i) and Lemma 1, we have

$$\begin{aligned}
&\max_{1\le i\le n} |\tilde{g}_{ii}(\hat{\epsilon}_i^2 - \sigma^2)| = \max_{1\le i\le n} |\tilde{g}_{ii}| o_p(n^{2/(4+\eta_1)}) = o_p(n^{2/(4+\eta_1)}), \\
&\max_{1\le i\le n} \left| \hat{\epsilon}_i \sum_{j=1}^{i-1} \tilde{g}_{ij} \hat{\epsilon}_j \right| = (\max_{1\le i\le n} |\hat{\epsilon}_i|)^2 \cdot \max_{1\le i\le n} \left| \sum_{j=1}^{i-1} \tilde{g}_{ij} \right| = o_p(n^{2/(4+\eta_1)}), \\
&\max_{1\le i\le n} |\hat{\epsilon}_i^2 - \sigma^2| = o_p(n^{2/(4+\eta_1)}),
\end{aligned}$$

Thus $Z_n = o_p(n^{2/(4+\eta_1)})$. (19) is proved.

For any given $l = (l_1, l_2)^T \in R^2$ with $||l|| = 1$, where $l_1, l_2 \in R$. Then

$$\begin{aligned}\sum_{i=1}^{n} l^T\omega_i(\theta) &= \sum_{i=1}^{n}\left\{l_1\left\{\tilde{g}_{ii}(\hat{\epsilon}_i^2-\sigma^2)+2\hat{\epsilon}_i\sum_{j=1}^{i-1}\tilde{g}_{ij}\hat{\epsilon}_j\right\}+l_2(\hat{\epsilon}_i^2-\sigma^2)\right\}\\
&= \sum_{i=1}^{n}(l_1\tilde{g}_{ii}+l_2)(\hat{\epsilon}_i^2-\sigma^2)+2\sum_{i=1}^{n}\sum_{j=1}^{i-1}l_1\tilde{g}_{ij}\hat{\epsilon}_i\hat{\epsilon}_j\\
&= \hat{\epsilon}_{(n)}^T U_n\hat{\epsilon}_{(n)}-\sigma^2 tr(U_n)\\
&= [\epsilon_{(n)}+O_p(n^{-\alpha_1}+n^{-(1-\alpha_1 r)/2})]^T U_n[\epsilon_{(n)}\\
&+ O_p(n^{-\alpha_1}+n^{-(1-\alpha_1 r)/2})]-\sigma^2 tr(U_n),\\
&= \epsilon_{(n)}^T U_n\epsilon_{(n)}-\sigma^2 tr(U_n)+2O_p(n^{-\alpha_1}+n^{-(1-\alpha_1 r)/2})U_n\epsilon_{(n)}\\
&+ O_p(n^{-2\alpha_1}+n^{-(1-\alpha_1 r)}),\end{aligned}$$

where $U_n=(u_{ij})_{n\times n}, u_{ii}=l_1\tilde{g}_{ii}+l_2, u_{ij}=l_1\tilde{g}_{ij}(i\neq j)$. Note that

$$U_n\epsilon_{(n)}=\begin{pmatrix}\sum_j u_{1j}\epsilon_j\\ \vdots\\ \sum_j u_{nj}\epsilon_j\end{pmatrix},$$

and

$$\sum_j u_{ij}\epsilon_j=\sqrt{Var\left(\sum_j u_{ij}\epsilon_j\right)}O_p(1)=\sqrt{\sum_j u_{ij}^2}O_p(1)=O_p(1), i=1,2,\cdots,n.$$

Then

$$\begin{aligned}\sum_{i=1}^{n} l^T\omega_i(\theta) &= \epsilon_{(n)}^T U_n\epsilon_{(n)}-\sigma^2 tr(U_n)+o_p(1)\\
&= \sum_{i=1}^{n}\left\{u_{ii}(\epsilon_i^2-\sigma^2)+2\sum_{j=1}^{i-1}u_{ij}\epsilon_i\epsilon_j\right\}+o_p(1)\\
&= Q_n+o_p(1),\end{aligned}$$

where $Q_n=\sum_{i=1}^{n}\{u_{ii}(\epsilon_i^2-\sigma^2)+2\sum_{j=1}^{i-1}u_{ij}\epsilon_i\epsilon_j\}$.

We now derive the variance of Q_n. It can be shown that

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij}^2 &= \sum_{i=1}^{n}\left\{(l_1\tilde{g}_{ii}+l_2)^2+\sum_{i\neq j}(l_1\tilde{g}_{ij})^2\right\}\\
&= \sum_{i=1}^{n}\left\{(l_1\tilde{g}_{ii})^2+2l_1l_2\tilde{g}_{ii}+l_2^2+\sum_{i\neq j}(l_1\tilde{g}_{ij})^2\right\}\\
&= 2l_1l_2\sum_{i=1}^{n}\tilde{g}_{ii}+nl_2^2+\sum_{i=1}^{n}\sum_{j=1}^{n}(l_1\tilde{g}_{ij})^2\\
&= 2l_1l_2tr(\tilde{G}_n)+nl_2^2+l_1^2tr(\tilde{G}_n^2),\\
\sum_{i=1}^{n} u_{ii}^2 &= \sum_{i=1}^{n}(l_1\tilde{g}_{ii}+l_2)^2\\
&= l_1^2\sum_{i=1}^{n}\tilde{g}_{ii}^2+2l_1l_2tr(\tilde{G}_n)+nl_2^2\\
&= l_1^2||Vec(diag\tilde{G}_n)||^2+2l_1l_2tr(\tilde{G}_n)+nl_2^2,
\end{aligned}
$$

It follows that the variance of Q_n is

$$
\begin{aligned}
\sigma_Q^2 &= 2\sum_{i=1}^{n}\sum_{j=1}^{n}u_{ij}^2\sigma^4+\sum_{i=1}^{n}\left\{u_{ii}^2(\mu_4-3\sigma^4)\right\}\\
&= 2\sigma^4\{l_1^2tr(\tilde{G}_n^2)+2l_1l_2tr(\tilde{G}_n)+nl_2^2\}\\
&+(\mu_4-3\sigma^4)\{l_1^2||Vec(diag\tilde{G}_n)||^2+2l_1l_2tr(\tilde{G}_n)+nl_2^2\}\\
&= l^T\Sigma l,
\end{aligned}
$$

where Σ is given in (18). From Condition A6, one can see that $n^{-1}\sigma_Q^2 \geq c_1 > 0$. From Theorem 1 in Kelejian and Prucha (2001), we have

$$
\frac{Q_n-E(Q_n)}{\sigma_Q}\xrightarrow{d}N(0,1).
$$

Noting that $E(Q_n)=0$, we thus have (43).

Next we will prove (21), i. e.

$$
n^{-1}\sum_{i=1}^{n}(l^T\omega_i(\theta))^2 = n^{-1}\sigma_Q^2+o_p(1). \tag{25}
$$

Note that

$$
\begin{aligned}
l^T\omega_i(\theta) &= u_{ii}(\hat{\epsilon}_i^2 - \sigma^2) + 2\sum_{j=1}^{i-1} u_{ij}\hat{\epsilon}_i\hat{\epsilon}_j \\
&= u_{ii}(\epsilon_i^2 - \sigma^2) + 2\sum_{j=1}^{i-1} u_{ij}\epsilon_i\epsilon_j + u_{ii}\epsilon_i O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2}) \\
&+ u_{ii}O_p(n^{-2\alpha_1} + n^{-(1-\alpha_1 r)}) \\
&+ \sum_{j=1}^{i-1} u_{ij}\epsilon_i O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2}) + \sum_{j=1}^{i-1} u_{ij}\epsilon_j O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2}) \\
&+ \sum_{j=1}^{i-1} u_{ij}O_p(n^{-2\alpha_1} + n^{-(1-\alpha_1 r)}) \\
&= Y_{in} + o_p(1),
\end{aligned}
$$

where

$$
Y_{in} = u_{ii}(\epsilon_i^2 - \sigma^2) + 2\sum_{j=1}^{i-1} u_{ij}\epsilon_i\epsilon_j.
$$

Thus

$$
n^{-1}\sum_{i=1}^{n}(l^T\omega_i(\theta))^2 = n^{-1}\sum_{i=1}^{n} Y_{in}^2 + 2n^{-1}\sum_{i=1}^{n} Y_{in}\cdot o_p(1) + o_p(1),
$$

and

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n} Y_{in} &= n^{-1}\sum_{i=1}^{n}\{u_{ii}(\epsilon_i^2 - \sigma^2)\} + 2n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{i-1} u_{ij}\epsilon_i\epsilon_j \\
&= \sqrt{n^{-2}\sum_{i=1}^{n} u_{ii}^2}\,O_p(1) + \sqrt{n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{i-1} u_{ij}^2}\,O_p(1) \\
&= o_p(1).
\end{aligned}
$$

It follows that

$$
n^{-1}\sum_{i=1}^{n}(l^T\omega_i(\theta))^2 = n^{-1}\sum_{i=1}^{n} Y_{in}^2 + o_p(1).
$$

Let $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_i = \sigma(\epsilon_1, \epsilon_2, \dots, \epsilon_i), 1 \le i \le n$. Then $\{Y_{in}, \mathcal{F}_i, 1 \le i \le n\}$ form a martingale difference array. Note that

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n} Y_{in}^2 - n^{-1}\sigma_Q^2 &= n^{-1}\sum_{i=1}^{n}(Y_{in}^2 - EY_{in}^2) \\
&= n^{-1}\sum_{i=1}^{n}\{Y_{in}^2 - E(Y_{in}^2|\mathcal{F}_{i-1}) + E(Y_{in}^2|\mathcal{F}_{i-1}) - EY_{in}^2\} \\
&= n^{-1}S_{n1} + n^{-1}S_{n2},
\end{aligned}
\tag{26}
$$

where $S_{n1} = \sum_{i=1}^{n}\{Y_{in}^2 - E(Y_{in}^2|\mathcal{F}_{i-1})\}$, $S_{n2} = \sum_{i=1}^{n}\{E(Y_{in}^2|\mathcal{F}_{i-1}) - EY_{in}^2\}$. Next we will prove

$$
n^{-1}S_{n1} = o_p(1), \tag{27}
$$

and

$$
n^{-1}S_{n2} = o_p(1). \tag{28}
$$

It suffices to prove $n^{-2}E(S_{n1}^2) \to 0$ and $n^{-2}E(S_{n2}^2) \to 0$ respectively. Obviously,

$$
Y_{in}^2 = u_{ii}^2(\epsilon_i^2 - \sigma^2)^2 + B_i^2\epsilon_i^2 + 2u_{ii}B_i(\epsilon_i^2 - \sigma^2)\epsilon_i,
$$

where $B_i = 2\sum_{j=1}^{i-1} u_{ij}\epsilon_j$. Thus

$$
E(Y_{in}^2|\mathcal{F}_{i-1}) = u_{ii}^2E(\epsilon_i^2 - \sigma^2)^2 + B_i^2\sigma^2 + 2u_{ii}B_i\mu_3.
$$

It follows that

$$
\begin{aligned}
n^{-2}E(S_{n1}^2) &= n^{-2}\sum_{i=1}^{n} E\{Y_{in}^2 - E(Y_{in}^2|\mathcal{F}_{i-1})\}^2 \\
&= n^{-2}\sum_{i=1}^{n} E[u_{ii}^2\{(\epsilon_i^2 - \sigma^2)^2 - E(\epsilon_i^2 - \sigma^2)^2\} + B_i^2(\epsilon_i^2 - \sigma^2) \\
&\quad + 2u_{ii}B_i(\epsilon_i^3 - \sigma^2\epsilon_i - \mu_3)]^2 \\
&\le Cn^{-2}\sum_{i=1}^{n} E[u_{ii}^4\{(\epsilon_i^2 - \sigma^2)^2 - E(\epsilon_i^2 - \sigma^2)^2\}^2] \\
&\quad + Cn^{-2}\sum_{i=1}^{n} E\{B_i^4(\epsilon_i^2 - \sigma^2)^2\} \\
&\quad + Cn^{-2}\sum_{i=1}^{n} E\{u_{ii}^2B_i^2(\epsilon_i^3 - \sigma^2\epsilon_i - \mu_3)^2\}.
\end{aligned}
\tag{29}
$$

By Condition A2, we have

$$
n^{-2}\sum_{i=1}^{n} E[u_{ii}^4\{(\epsilon_i^2 - \sigma^2)^2 - E(\epsilon_i^2 - \sigma^2)^2\}^2] \le Cn^{-1} \to 0, \tag{30}
$$

and

$$n^{-2}\sum_{i=1}^{n} E\{B_i^4(\epsilon_i^2-\sigma^2)^2\} \le Cn^{-2}\sum_{i=1}^{n} E\left(\sum_{j=1}^{i-1} u_{ij}\epsilon_j\right)^4 \\ \le Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{i-1} u_{ij}^4\mu_4 + Cn^{-2}\sum_{i=1}^{n}\left(\sum_{j=1}^{i-1} u_{ij}^2\sigma^2\right)^2 \le Cn^{-1} \to 0. \tag{31}$$

Similarly, one can show that

$$n^{-2}\sum_{i=1}^{n} E\{u_{ii}^2 B_i^2(\epsilon_i^3-\sigma^2\epsilon_i-\mu_3)^2\} \to 0. \tag{32}$$

From (29)–(32), we have $n^{-2}E(S_{n1}^2) \to 0$. Furthermore,

$$E(Y_{in}^2) = E\{E(Y_{in}^2|\mathcal{F}_{i-1})\} = u_{ii}^2E(\epsilon_i^2-\sigma^2)^2 + \sigma^2E(B_i^2) + 2u_{ii}\mu_3E(B_i) \\ = u_{ii}^2E(\epsilon_i^2-\sigma^2)^2 + 4\sigma^2\left(\sum_{j=1}^{i-1} u_{ij}^2\sigma^2\right).$$

Thus,

$$n^{-2}E(S_{n2}^2) = n^{-2}E\left[\sum_{i=1}^{n}\{E(Y_{in}^2|\mathcal{F}_{i-1}) - EY_{in}^2\}\right]^2 \\ = n^{-2}E\left[\sum_{i=1}^{n}\left\{B_i^2\sigma^2 - 4\sigma^2\sum_{j=1}^{i-1}u_{ij}^2\sigma^2 + 2u_{ii}\mu_3B_i\right\}\right]^2 \\ = n^{-2}\sum_{i=1}^{n} E\left[\sigma^2\left\{\left(2\sum_{j=1}^{i-1}u_{ij}\epsilon_j\right)^2 - 4\sum_{j=1}^{i-1}u_{ij}^2\sigma^2\right\}\right. \\ \left. +2u_{ii}\mu_3\left(2\sum_{j=1}^{i-1}u_{ij}\epsilon_j\right)\right]^2 \\ \le Cn^{-2}\sum_{i=1}^{n} E\left[\sigma^2\left\{\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_j\right)^2 - \sum_{j=1}^{i-1}u_{ij}^2\sigma^2\right\}\right]^2 \\ + Cn^{-2}\sum_{i=1}^{n} E\left\{2u_{ii}\mu_3\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_j\right)\right\}^2. \tag{33}$$

Note that

$$n^{-2}\sum_{i=1}^{n} E\left[\sigma^2\left\{\left(\sum_{j=1}^{i-1} u_{ij}\epsilon_j\right)^2 - \sum_{j=1}^{i-1} u_{ij}^2\sigma^2\right\}\right]^2 \le n^{-2}\sigma^4 \sum_{i=1}^{n} E\left(\sum_{j=1}^{i-1} u_{ij}\epsilon_j\right)^4$$
$$\le Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{i-1} u_{ij}^4\mu_4 + Cn^{-2}\sum_{i=1}^{n}\left(\sum_{j=1}^{i-1} u_{ij}^2\sigma^2\right)^2 \le Cn^{-1} \to 0, \quad (34)$$

and

$$n^{-2}\sum_{i=1}^{n} E\left\{2u_{ii}\mu_3\left(\sum_{j=1}^{i-1} u_{ij}\epsilon_j\right)\right\}^2 = 4\mu_3^2\sigma^2 n^{-2}\sum_{i=1}^{n} u_{ii}^2 \sum_{j=1}^{i-1} u_{ij}^2 \le Cn^{-1} \to 0, \quad (35)$$

where we have used Conditions A2 and A4(i). From (33)–(35), we have $n^{-2}ES_{n2}^2 \to 0$. The proof of (25) is thus complete.

Finally, we will prove (22). Note that

$$||\omega_i(\theta)||^3 \le |\tilde{g}_{ii}(\hat{\epsilon}_i^2 - \sigma^2) + 2\hat{\epsilon}_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\hat{\epsilon}_j|^3 + |\hat{\epsilon}_i^2 - \sigma^2|^3$$
$$= |\tilde{g}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j|^3 + |\epsilon_i^2 - \sigma^2|^3 + o_p(1), \quad (36)$$

where we use $\hat{\epsilon}_i = \epsilon_i + O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2})$. By Conditions A2 and A4(i),

$$\sum_{i=1}^{n} E\left|\tilde{g}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j\right|^3$$
$$\le C\sum_{i=1}^{n} E|\tilde{g}_{ii}(\epsilon_i^2 - \sigma^2)|^3 + C\sum_{i=1}^{n} E\left|2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j\right|^3$$
$$\le C\sum_{i=1}^{n} E|\tilde{g}_{ii}(\epsilon_i^2 - \sigma^2)|^3 + C\sum_{i=1}^{n} E|\epsilon_i|^3 \sum_{j=1}^{i-1} E|\tilde{g}_{ij}\epsilon_j|^3$$
$$+ C\sum_{i=1}^{n} E|\epsilon_i|^3\left\{\sum_{j=1}^{i-1} E(\tilde{g}_{ij}\epsilon_j)^2\right\}^{3/2} = O(n), \quad (37)$$

$$\sum_{i=1}^{n} E|\epsilon_i^2 - \sigma^2|^3 = O(n). \quad (38)$$

From (36)–(38),we have

$$\sum_{i=1}^{n} E||\omega_i(\theta)||^3 = O(n). \tag{39}$$

Further, using (39) and Markov inequality, we obtain $\sum_{i=1}^{n} ||\omega_i(\theta)||^3 = O_p(n)$. Thus (22) is proved. □

Lemma 3 *Suppose that Assumptions* (A1)–(A6) *are satisfied, then*

$$\sqrt{\frac{n}{\gamma_n}}(\hat{\rho}_n - \rho_0) \xrightarrow{d} N(0, \sigma^2_{\rho_0}), \tag{40}$$

where

$$\sigma^2_{\rho_0} = \lim_{n\to\infty} \left\{ \frac{\gamma_n}{n} \left[tr(C_n C_n^S) + (\mu_4 - 3\sigma_0^4)||Vec(diagC_n)||^2 \right] \right\}^{-1},$$
$$C_n = G_{n0} - \frac{tr(G_{n0})}{n} I_n, C_n^S = C_n + C_n^T, G_{n0} = W_n A_n^{-1}(\rho_0).$$

Proof By the Taylor expansion of $\partial L(\hat{\rho}_n)/\partial \rho$ at $\rho = \rho_0$, the asymptotic distribution of $\hat{\rho}_n$ will follow from

$$\sqrt{\frac{n}{\gamma_n}}(\hat{\rho}_n - \rho_0) = -\left(\frac{\gamma_n}{n} \frac{\partial^2 L(\widetilde{\rho})}{\partial \rho^2} \right)^{-1} \sqrt{\frac{\gamma_n}{n}} \frac{\partial L(\rho_0)}{\partial \rho}, \tag{41}$$

where $\widetilde{\rho}$ is between $\hat{\rho}_n$ and ρ_0.

Firstly, the first- and second-order derivatives of the concentrated log-likelihood of ρ are

$$\frac{\partial L(\rho)}{\partial \rho} = \frac{1}{\hat{\sigma}_n^2(\rho)} \hat{R}^T W_n^T A_n(\rho) \hat{R} - tr(W_n A_n^{-1}(\rho)), \tag{42}$$

$$\begin{aligned} \frac{\partial^2 L(\rho)}{\partial \rho^2} = \frac{2}{n\hat{\sigma}_n^4(\rho)} (\hat{R}^T W_n^T A_n(\rho) \hat{R})^2 - \frac{1}{\hat{\sigma}_n^2(\rho)} \hat{R}^T W_n^T W_n \hat{R} \\ - tr([W_n A_n^{-1}(\rho)]^2), \end{aligned} \tag{43}$$

where $\hat{R} = (\hat{R}_1, \ldots, \hat{R}_n)^T$ and $\hat{\sigma}_n^2(\rho) = (1/n)\hat{R}^T A_n^T(\rho) A_n(\rho) \hat{R}$. Note that

$$\begin{aligned} \frac{\gamma_n}{n} \hat{R}^T W_n^T W_n \hat{R} &= \frac{\gamma_n}{n} (R^T + \hat{R}^T - R^T) W_n^T W_n (R + \hat{R} - R) \\ &= \frac{\gamma_n}{n} R^T W_n^T W_n R + \frac{\gamma_n}{n} (\hat{R}^T - R^T) W_n^T W_n R + \frac{\gamma_n}{n} R^T W_n^T W_n (\hat{R} - R) \\ &\quad + \frac{\gamma_n}{n} (\hat{R}^T - R^T) W_n^T W_n (\hat{R} - R), \end{aligned}$$

and

$$\hat{R} - R = \begin{pmatrix} m(x_1) - \hat{m}_n(x_1) \\ \cdots \\ m(x_n) - \hat{m}_n(x_n) \end{pmatrix}.$$

By (24), we have $m(x_j) - \hat{m}_n(x_j) = O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2})$, then

$$\hat{R} - R = O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2}).$$

Thus,

$$\frac{\gamma_n}{n}\hat{R}^T W_n^T W_n \hat{R} = \frac{\gamma_n}{n} R^T W_n^T W_n R + o_p(1) = \frac{\gamma_n}{n}\epsilon_{(n)}^T G_{n0}^T G_{n0}\epsilon_{(n)} + o_p(1). \tag{44}$$

Similarly,

$$\frac{\gamma_n}{n}\hat{R}^T W_n^T A_n(\rho)\hat{R} = \frac{\gamma_n}{n}\epsilon_{(n)}^T G_{n0}^T \epsilon_{(n)} + (\rho_0 - \rho)\frac{\gamma_n}{n}\epsilon_{(n)}^T G_{n0}^T G_{n0}\epsilon_{(n)} + o_p(1). \tag{45}$$

When $\lim_{n\to\infty}\gamma_n = \infty$, $\frac{1}{n}\hat{R}^T W_n^T A_n(\rho)\hat{R} = o_p(1)$ and $\hat{\sigma}_n^2(\rho) = \sigma_0^2 + o_p(1)$ uniformly on Λ. It follows that

$$\frac{\gamma_n}{n}\frac{\partial^2 L(\rho)}{\partial \rho^2} = -\frac{1}{\sigma_0^2}\frac{\gamma_n}{n}\epsilon_{(n)}^T G_{n0}^T G_{n0}\epsilon_{(n)} - \frac{\gamma_n}{n} tr([W_n A^{-1}(\rho)]^2) + o_p(1),$$

uniformly on Λ. Under A4(ii), $(\gamma_n/n)tr([W_n A_n^{-1}(\rho)]^3) = O(1)$ uniformly on Λ. Then by the Taylor expansion,

$$\begin{aligned} \frac{\gamma_n}{n}\left(\frac{\partial^2 L(\widetilde{\rho})}{\partial \rho^2} - \frac{\partial^2 L(\rho_0)}{\partial \rho^2}\right) &= -\frac{\gamma_n}{n}\{tr([W_n A_n^{-1}(\widetilde{\rho})]^2) - tr(G_{n0}^2)\} + o_p(1) \\ &= -2\frac{\gamma_n}{n} tr([W_n A_n^{-1}(\widetilde{\rho})]^3)(\widetilde{\rho} - \rho_0) + o_p(1) \\ &= o_p(1) \end{aligned} \tag{46}$$

for any $\widetilde{\rho}$ which converges in probability to ρ_0. Furthermore,

$$\begin{aligned} &\frac{\gamma_n}{n}\left[\frac{\partial^2 L(\rho_0)}{\partial \rho^2} - E\left(\frac{\partial^2 L(\rho_0)}{\partial \rho^2}\right)\right] \\ &= -\frac{1}{\sigma_0^2}\frac{\gamma_n}{n}[\epsilon_{(n)}^T G_{n0}^T G_{n0}\epsilon_{(n)} - \sigma_0^2 tr(G_{n0}^T G_{n0})] + o_p(1) = o_p(1). \end{aligned} \tag{47}$$

On the other hand, combine (42) with (45), we have

$$\sqrt{\frac{\gamma_n}{n}}\frac{\partial L(\rho_0)}{\partial \rho} = \frac{1}{\hat{\sigma}_n^2(\rho_0)}\sqrt{\frac{\gamma_n}{n}} q_n + o_p(1),$$

where $q_n = \epsilon_{(n)}^T C_n^T \epsilon_{(n)}$ and $C_n = G_{n0} - [\frac{tr(G_{n0})}{n}]I_n$. As $E(q_n) = \sigma_0^2 tr(C_n) = 0$ and $\sigma_{q_n}^2 = Var(q_n) = \sigma_0^4[tr(C_n^T C_n) + tr(C_n^2)] + (\mu_4 - 3\sigma_0^4)||Vec(diagC_n)||^2$. Then by the central limit theorem for quadratic functions, we have $(q_n - E(q_n))/\sigma_{q_n} \xrightarrow{d} N(0,1)$. It follows that

$$\sqrt{\frac{\gamma_n}{n}}\frac{\partial L(\rho_0)}{\partial \rho} = \frac{\sqrt{\gamma_n/n}\sigma_{qn}}{\hat{\sigma}_n^2(\rho_0)} \cdot \frac{q_n - E(q_n)}{\sigma_{q_n}} + o_p(1) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\frac{\gamma_n}{n}\frac{\sigma_{q_n}^2}{\sigma_0^4}\right). \tag{48}$$

By (41), (46), (47) and (48), the proof of Lemma 3 is complete. □

Lemma 4 *Suppose that Assumptions* (A1)–(A6) *are satisfied. Then as* $n \to \infty$,

$$\widetilde{Z}_n = \max_{1\le i\le n}|Z_{in}(\vartheta)| = o_p(n^{1/(4+\eta_1)}), \tag{49}$$

$$\sigma_n^{-1}\sum_{i=1}^n Z_{in}(\vartheta) \xrightarrow{d} N(0,1), \tag{50}$$

$$\sigma_n^{-2}\sum_{i=1}^n Z_{in}^2(\vartheta) = 1 + o_p(1), \tag{51}$$

$$\sum_{i=1}^n |Z_{in}(\vartheta)|^3 = O_p(nh^r + n^{1-3\alpha_1} + n^{1-3(1-\alpha_1 r)/2} + n\gamma_n^{-3/2}), \tag{52}$$

where $\sigma_n^2 = \sigma^2 V_n^T V_n$, $V_n = (K_h(x-x_1), \dots, K_h(x-x_n))^T$.

Proof Note that

$$\begin{aligned}
Z_{in}(\vartheta) &= K_h(x-x_i)(\tilde{y}_i - \vartheta)\\
&= K_h(x-x_i)\left\{ y_i - \hat{\rho}_n\sum_{j=1}^n w_{ij}\hat{R}_j - m(x)\right\}\\
&= K_h(x-x_i)\left\{ m(x_i) + \rho\sum_{j=1}^n w_{ij}R_j + \epsilon_i - \hat{\rho}_n\sum_{j=1}^n w_{ij}\hat{R}_j - m(x)\right\}\\
&= K_h(x-x_i)\{m(x_i) - m(x)\}\\
&\quad + K_h(x-x_i)\epsilon_i - K_h(x-x_i)\left(\hat{\rho}_n\sum_{j=1}^n w_{ij}\hat{R}_j - \rho\sum_{j=1}^n w_{ij}R_j\right),
\end{aligned}$$

and

$$
\begin{aligned}
&K_h(x-x_i)\left(\hat{\rho}_n\sum_{j=1}^{n}w_{ij}\hat{R}_j-\rho\sum_{j=1}^{n}w_{ij}R_j\right)\\
&\quad=K_h(x-x_i)\left\{\hat{\rho}_n\sum_{j=1}^{n}w_{ij}(\hat{R}_j-R_j)+(\hat{\rho}_n-\rho)\sum_{j=1}^{n}w_{ij}R_j\right\}\\
&\quad=K_h(x-x_i)\left\{\hat{\rho}_n\sum_{j=1}^{n}w_{ij}(m(x_j)-\hat{m}_n(x_j))+(\hat{\rho}_n-\rho)\sum_{j=1}^{n}w_{ij}R_j\right\}\\
&\quad=K_h(x-x_i)\hat{\rho}_n\sum_{j=1}^{n}w_{ij}\{m(x_j)-\hat{m}_n(x_j)\}\\
&\qquad+K_h(x-x_i)(\hat{\rho}_n-\rho)\sum_{j=1}^{n}w_{ij}\beta_j^T\epsilon_{(n)},
\end{aligned}
\tag{53}
$$

where $\beta_j, 1\le j\le n$ satisfy $A_n^{-1}(\rho)=\begin{pmatrix}\beta_1^T\\ \vdots\\ \beta_n^T\end{pmatrix}$. It can be shown that

$$
\begin{aligned}
m(x_j)-\hat{m}_n(x_j)&=m(x_j)-\sum_{i=1}^{n}\widetilde{W}_{ni}(x_j)y_i\\
&=\sum_{i=1}^{n}\widetilde{W}_{ni}(x_j)\{m(x_j)-m(x_i)\}-\sum_{i=1}^{n}\widetilde{W}_{ni}(x_j)R_i\\
&=O(n^{-\alpha_1})-(\widetilde{W}_{n1}(x_j),\dots,\widetilde{W}_{nn}(x_j))R\\
&=O(n^{-\alpha_1})-\sqrt{Var((\widetilde{W}_{n1}(x_j),\dots,\widetilde{W}_{nn}(x_j))R)}O_p(1)\\
&=O_p(n^{-\alpha_1}+n^{-(1-\alpha_1 r)/2}),
\end{aligned}
\tag{54}
$$

and

$$
\begin{aligned}
\sum_{j=1}^{n}w_{ij}\beta_j^T\epsilon_{(n)}&=||\sum_{j=1}^{n}w_{ij}\beta_j^T||O_p(1)=\sqrt{\sum_{k=1}^{n}\left[\sum_{j=1}^{n}w_{ij}\beta_{jk}\right]^2}O_p(1)\\
&=\sqrt{\gamma_n^{-2}\sum_{k=1}^{n}\left[\sum_{j=1}^{n}|\beta_{jk}|\right]^2}O_p(1)=O_p(n^{1/2}\gamma_n^{-1}),
\end{aligned}
$$

where $\beta_j=(\beta_{j1},\cdots,\beta_{jn}), j=1,2,\dots,n$. By Lemma 3, we have $\hat{\rho}_n-\rho=O_p(n^{-1/2}\gamma_n^{1/2})$. Using Conditions A1-A4 and Lemma 1, we can obtain

$$
\begin{aligned}
\widetilde{Z}_n \leq & \max_{1\leq i\leq n} |K_h(x-x_i)\{m(x_i)-m(x)\}| + \max_{1\leq i\leq n} |K_h(x-x_i)\epsilon_i| \\
& + \max_{1\leq i\leq n} |K_h(x-x_i)\hat{\rho}_n \sum_{j=1}^{n} w_{ij}\{m(x_j)-\hat{m}_n(x_j)\}| \\
& + \max_{1\leq i\leq n} |K_h(x-x_i)(\hat{\rho}_n-\rho) \sum_{j=1}^{n} w_{ij}\beta_j^T \epsilon_{(n)}| \\
& = o_p(n^{1/(4+\eta_1)}).
\end{aligned}
$$

(49) is thus proved. Next we will prove (50). Based on (53) and (53), we have

$$
\begin{aligned}
\sum_{i=1}^{n} Z_{in}(\vartheta) = & \sum_{i=1}^{n} K_h(x-x_i)\epsilon_i + \sum_{i=1}^{n} K_h(x-x_i)\{m(x_i)-m(x)\} \\
& + \hat{\rho}_n \sum_{i=1}^{n} K_h(x-x_i) \sum_{j=1}^{n} w_{ij}\{m(x_j)-\hat{m}_n(x_j)\} \\
& + (\hat{\rho}_n-\rho) \sum_{i=1}^{n}\sum_{j=1}^{n} K_h(x-x_i) w_{ij}\beta_j^T \epsilon_{(n)}.
\end{aligned} \tag{55}
$$

Further, by conditions A1, A3, A4, it follows that

$$
\sum_{i=1}^{n} K_h(x-x_i)\{m(x_i)-m(x)\} = O(nh^{r+1}), \tag{56}
$$

$$
\begin{aligned}
& \hat{\rho}_n \sum_{i=1}^{n} K_h(x-x_i) \sum_{j=1}^{n} w_{ij}\{m(x_j)-\hat{m}_n(x_j)\} \\
& = O(n^{1-\alpha_1}h^r + n^{(1+\alpha_1 r)/2}h^r),
\end{aligned} \tag{57}
$$

and

$$
\sum_{i=1}^{n}\sum_{j=1}^{n} K_h(x-x_i) w_{ij}\beta_j^T \epsilon_{(n)} = \sum_{i=1}^{n} K_h(x-x_i) \sum_{j=1}^{n} w_{ij}\beta_j^T \epsilon_{(n)} = O_p(n^{3/2}h^r\gamma_n^{-1}).
$$

As $\hat{\rho}_n - \rho = O_p(n^{-1/2}\gamma_n^{1/2})$, it follows that

$$
(\hat{\rho}_n-\rho) \sum_{i=1}^{n}\sum_{j=1}^{n} K_h(x-x_i) w_{ij}\beta_j^T \epsilon_{(n)} = O_p(\gamma_n^{-1/2} nh^r). \tag{58}
$$

From (55)–(58), we obtain

$$
\sum_{i=1}^{n} Z_{in}(\vartheta) = \sum_{i=1}^{n} K_h(x-x_i)\epsilon_i + O_p(n^{1-\alpha_1}h^r + n^{(1+\alpha_1 r)/2}h^r + \gamma_n^{-1/2} nh^r). \tag{59}
$$

(59) can be rewritten as

$$\sum_{i=1}^{n} Z_{in}(\vartheta) = V_n^T \epsilon_{(n)} + O_p(n^{1-\alpha_1} h^r + n^{(1+\alpha_1 r)/2} h^r + \gamma_n^{-1/2} n h^r). \tag{60}$$

Noting that $V_n^T V_n \sim nh^r$ by Lei and Qin (2011) and using the central limit theorem for independent sums, we obtain

$$\sum_{i=1}^{n} Z_{in}(\vartheta) \xrightarrow{d} N(0, \sigma^2 V_n^T V_n), \tag{61}$$

as $\sigma^{-1}(n^{1-\alpha_1} h^r + n^{(1+\alpha_1 r)/2} h^r + \gamma_n^{-1/2} n h^r)(V_n^T V_n)^{-1/2} \to 0$ by Condition A5(ii). The proof of (50) is complete.

Next we will prove (51). Note that

$$\begin{aligned}
\sigma_n^{-2} \sum_{i=1}^{n} Z_{in}^2(\vartheta) &= \sigma_n^{-2} \sum_{i=1}^{n} \Big\{ K_h(x-x_i)\epsilon_i + K_h(x-x_i)\{m(x_i) - m(x)\} \\
&\quad + O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2} + \gamma_n^{-1/2}) \Big\}^2 \\
&= \sigma_n^{-2} \sum_{i=1}^{n} \Big\{ K_h(x-x_i)\epsilon_i + K_h(x-x_i)\{m(x_i) - m(x)\} \Big\}^2 \\
&\quad + \sigma_n^{-2} n \cdot O_p(n^{-2\alpha_1} + n^{-(1-\alpha_1 r)} + \gamma_n^{-1}) \\
&\quad + 2\sigma_n^{-2} \sum_{i=1}^{n} \Big\{ K_h(x-x_i)\epsilon_i + K_h(x-x_i)\{m(x_i) - m(x)\} \Big\} \\
&\quad \times O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2} + \gamma_n^{-1/2}) \\
&= \sum_{j=1}^{3} T_{nj} + o_p(1),
\end{aligned} \tag{62}$$

where

$$T_{n1} = \sigma_n^{-2} \sum_{i=1}^{n} \Big\{ K_h(x-x_i)\epsilon_i \Big\}^2, \quad T_{n2} = \sigma_n^{-2} \sum_{i=1}^{n} \Big\{ K_h(x-x_i)\{m(x_i) - m(x)\} \Big\}^2,$$

$$T_{n3} = 2\sigma_n^{-2} \sum_{i=1}^{n} \Big\{ K_h^2(x-x_i)\epsilon_i \Big\} \Big\{ m(x_i) - m(x) \Big\}.$$

(51) is valid if we can show that

$$\sum_{j=1}^{3} T_{nj} = 1 + o_p(1). \tag{63}$$

Note that

$$ET_{n1} = \sigma_n^{-2}\sigma^2 \sum_{i=1}^{n} K_h^2(x - x_i) = 1.$$

Next we will show that

$$E|T_{n1} - ET_{n1}|^{(4+\eta_1)/2} \to 0.$$

By the moment inequality, we have

$$\begin{aligned}
E|T_{n1} - ET_{n1}|^{(4+\eta_1)/2} &= \sigma_n^{-(4+\eta_1)} E|\sum_{i=1}^{n} K_h^2(x - x_i)(\epsilon_i^2 - E\epsilon_i^2)|^{(4+\eta_1)/2} \\
&\leq C\sigma_n^{-(4+\eta_1)} \left\{ \sum_{i=1}^{n} EK_h^{4+\eta_1}(x - x_i)|\epsilon_i^2 - E\epsilon_i^2|^{(4+\eta_1)/2} \right. \\
&\quad \left. + \left(\sum_{i=1}^{n} EK_h^4(x - x_i)(\epsilon_i^2 - E\epsilon_i^2)^2 \right)^{(4+\eta_1)/4} \right\} \\
&\leq C\sigma_n^{-(4+\eta_1)} \{nh^r + (nh^r)^{(4+\eta_1)/4}\} \to 0.
\end{aligned}$$

It follows that

$$T_{n1} = 1 + o_p(1).$$

On the other hand,

$$\begin{aligned}
T_{n2} &= \sigma_n^{-2} \sum_{i=1}^{n} \left\{ K_h(x - x_i)\{m(x_i) - m(x)\} \right\}^2 \\
&\leq C\sigma_n^{-2} \sum_{i=1}^{n} \left\{ K_h(x - x_i)\{m(x_i) - m(x)\} I(||x_i - x|| \geq Mh) \right\}^2 \\
&\quad + C\sigma_n^{-2} \sum_{i=1}^{n} \left\{ K_h(x - x_i)\{m(x_i) - m(x)\} I(||x_i - x|| < Mh) \right\}^2 \\
&\leq C\sigma_n^{-2} nh^{r+2} \to 0,
\end{aligned} \tag{64}$$

$$\begin{aligned}
ET_{n3}^2 &= 4\sigma_n^{-4}\sigma^2 \sum_{i=1}^{n} K_h^4(x - x_i)\{m(x_i) - m(x)\}^2 \\
&\leq 4\sigma_n^{-4}\sigma^2 h^2 \sum_{i=1}^{n} K_h^4(x - x_i) \leq C\sigma_n^{-4} nh^{r+2} \to 0.
\end{aligned} \tag{65}$$

(51) is thus verified. Finally, we prove (52). By C_r inequality,

$$\begin{aligned}\sum_{i=1}^{n} E|Z_{in}(\vartheta)|^3 &= \sum_{i=1}^{n} E|K_h(x-x_i)\epsilon_i + K_h(x-x_i)\{m(x_i)-m(x)\} \\ &+ O_p(n^{-\alpha_1} + n^{-(1-\alpha_1 r)/2} + \gamma_n^{-1/2})|^3 \\ &\le C\sum_{i=1}^{n} E|K_h(x-x_i)\epsilon_i|^3 + C\sum_{i=1}^{n} |K_h(x-x_i)\{m(x_i)-m(x)\}|^3 \\ &+ Cn(n^{-3\alpha_1} + n^{-3(1-\alpha_1 r)/2} + \gamma_n^{-3/2}) \\ &\le Cnh^r + Cnh^{r+3} + Cn^{1-3\alpha_1} + Cn^{1-3(1-\alpha_1 r)/2} + Cn\gamma_n^{-3/2} \\ &= O(nh^r + n^{1-3\alpha_1} + n^{1-3(1-\alpha_1 r)/2} + n\gamma_n^{-3/2}).\end{aligned} \tag{66}$$

Proofs of Theorems 1 and 2 Using Lemmas 2 and 4 respectively, similar to the proof of Theorem 1 in Qin and Li (2011), one can prove Theorems 1 and 2. □

References

Anselin, L. (1988). *Spatial econometrics: methods and models*. Amsterdam: Kluwer Academic Publishers.

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. A. Giles (Eds.), *Handbook of applied economics statistics*. New York: Marcel Dekker.

Bandyopadhyay, S., Lahiri, N., & Nordman, D. J. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *Annals of Statistics*, *43*, 519–545.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data* (2nd ed.). New York: CRC Press.

Besag, J., York, J., & Mollin, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*, 1–20.

Chan, N. H., & Ling, S. Q. (2006). Empirical likelihood for GARCH models. *Econometric Theory*, *22*, 403–428.

Chen, S. X., & Wong, C. M. (2009). Smoothed block empirical likelihood for quantiles of weakly dependent processes. *Statistica Sinica*, *19*, 71–81.

Chuang, C. S., & Chan, N. H. (2002). Empirical likelihood for autoregressive models, with applications to unstable time series. *Statistica Sinica*, *12*, 387–407.

Cliff, A. D., & Ord, J. K. (1973). *Spatial autocorrelation*. London: Pion Ltd.

Cressien, N. (1993). *Statistics for spatial data*. New York: Wiley.

Diggle, P. J., & Tawn, J. A. (1998). Model-based geostatistics. *Journal of Applied Statistics*, *47*, 299–350.

Hallin, M., Lu, Z., & Tran, L. T. (2001). Density estimation for spatial linear processes. *Bernoulli*, *7*, 657–668.

Hallin, M., Lu, Z., & Tran, L. T. (2004a). Density estimation for spatial processes: the L_1 theory. *Journal of Multivariate Analysis*, *88*, 61–75.

Hallin, M., Lu, Z., & Tran, L. T. (2004b). Local linear spatial regression. *Annals of Statistics*, *32*, 2469–2500.

Jin, F., & Lee, L. F. (2018). Outer-product-of-gradients tests for spatial autoregressive models. *Regional Science and Urban Economics*, *72*, 35–57.

Jin, F., & Lee, L. F. (2019). GEL estimation and tests of spatial autoregressive models. *Journal of Econometrics*, *208*, 585–612.

Kelejian, H. H., & Prucha, I. R. (2001). On the asymptotic distribution of the Moran *I* test statistic with applications. *Journal of Econometrics*, *104*, 219–257.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics*, *25*, 2084–2102.
Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, *72*, 1899–1925.
Lee, L. F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, *137*, 489–514.
Lei, Q., & Qin, Y. (2011). Empirical likelihood for non-parametric regression models with missing responses: multiple design case. *Acta Mathematicae Applicatae Sinica, English Series*, *27*, 1–12.
Lu, Z., Tang, Q., & Cheng, L. (2014). Estimating spatial quantile regression with functional coefficients: a robust semiparametric framework. *Bernoulli*, *20*, 164–189.
Mykland, P. A. (1995). Dual likelihood. *Annals of Statistics*, *23*, 396–421.
Nordman, D. J. (2008). An empirical likelihood method for spatial regression. *Metrika*, *68*, 351–363.
Nordman, D. J., & Daniel, P. C. (2008). Point and interval estimation of variogram models using spatial empirical likelihood. *Journal of the American Statistical Association*, *103*, 350–361.
Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*, 120–126.
Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237–249.
Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*, *18*, 90–120.
Owen, A. (1991). Empirical likelihood for linear models. *Annals of Statistics*, *19*, 1725–1747.
Owen, A. (2001). *Empirical likelihood*. New York: Chapman & Hall.
Qin, Y., & Li, Y. (2011). Empirical likelihood for linear models under negatively associated errors. *Journal of Multivariate Analysis*, *102*, 153–163.
Qin, Y. (2019). Empirical likelihood for spatial autoregressive models with spatial autoregressive disturbances. *Sankhya A*. https://doi.org/10.1007/s13171-019-00166-3.
Su, L., & Jin, S. (2010). Profile quasi-maximum likelihood estimation of spatial autoregressive models. *Journal of Econometrics*, *157*, 18–33.
Wang, H. X., Lin, J. G., & Wang, J. D. (2016). Nonparametric spatial regression with spatial autoregressive error structure. *Statistics*, *50*, 60–75.
Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based seostatistics. *Journal of the American Statistical Association*, *99*, 250–261.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.